UNIVERSITY OF LATVIA, FACULTY OF BIOLOGY

DEPARTMENT OF MICROBIOLOGY AND BIOTECHNOLOGY

NILS ROSTOKS

STRUCTURE AND REGULATION OF THE RAFFINOSE CATABOLISM OPERON PROMOTER

Author M.Sc. N. Rostoks

Scientific supervisor Prof. Dr. habil. biol. I. Muižnieks

RĪGA, 1999

Doctoral thesis "Structure and regulation of the raffinose catabolism operon promoter" was prepared at the Department of Microbiology and Biotechnology, Faculty of Biology, University of Latvia from 1995 till 1998

Habilitation and Promotion Council of the Council of Science of Latvia
Council 12. Biology (Molecular biology, biochemistry, biotechnology, microbiology, virology)

Chairman of the Council
Prof. Dr. habil. biol. Elmārs Grēns

.....

Opponents

Dr. Rainer Ulrich (Institute of Medical Virology, Charité, Humboldt University, Berlin, Germany)

Dr. habil. biol. Kestutis Sasnauskas (Institute of Biotechnology, Vilnius, Lithuania)

Dr. habil. biol. Alexander Rapoport (Institute of Microbiology and Biotechnology, University of Latvia, Rīga, Latvia)

Defence of the doctoral thesis will take place at the session of the Habilitation and Promotion Council, 1999 June 21, 14:00, in the Biomedical Research and Study Centre, Rīga, Rātsupītes 1.

Doctoral thesis is available at the Library of the University of Latvia,
Kalpaka Blvd. 4.

CONTENTS

ABBREVIATIONS

AR1, AR2	– activating region 1 and 2 of CRP
bp	– base pair(s)
cAMP	– cyclic adenosine monophosphate
CAP	– catabolite activator protein (the same as CRP)
CRP	– cAMP receptor protein
ds	– double stranded
EMSA	– electrophoretic mobility shift assay
HTH	– helix – turn – helix motif in DNA binding proteins
NTP(s)	– ribonucleotide triphosphate(s)
OD_{600}	– optical density at the wave length 600 nm
PAAG	– polyacrylamide gel
rafP	– raffinose operon promoter / operator region
RafR	– *Escherichia coli* plasmid-borne raffinose operon repressor protein
RNAP	– *Escherichia coli* RNAP holoenzyme

Rostoks, N. 1999. Rafinozes katabolisma operona promotera struktūra un regulācija. Latvijas Universitāte, Rīga, 58 lpp.

KOPSAVILKUMS

Escherichia coli enteropatogenitātes plazmīdu rafinozes operons kodē proteīnus, kas nepieciešami trisaharīda rafinozes aktīvam transportam un sašķelšanai monosaharīdos. To veido *rafA* (α-galaktozidāzes), *rafB* (permeāzes) un *rafD* (invertāzes) gēni, kuru transkripciju regulē *rafP* promoters. Ja barotnē nav šķelšanai piemērotu cukuru, promotera aktivitāti bloķē *raf* represors (RafR). Vienīgais zināmais *raf* operona induktors ir disaharīds melibioze, ko veido invertāze, atšķeļot fruktozi no rafinozes molekulas.

Promocijas darbā pētīta *E. coli raf* operona promotera *rafP* DNS un DNS – proteīnu kompleksu struktūra, topoloģija un loma transkripcijas regulācijā.

Darbā izmantotās EMSA, DNāzes I protekcijas (fūtprintinga) un proteīnu-DNS kompleksu topoloģijas pētīšanas metodes ir raksturotas pirmajā publikācijā. Sekvences atkarīgās un proteīnu inducētās DNS topoloģijas noskaidrošanai būtiski svarīga bija *rafP* fragmentu klonēšana tiešo atkārtojumu (tandēma dimēru) veidā, kā arī mutaģenēzes metodes attāluma mainīšanai starp dažādiem proteīnus piesaistošajiem DNS rajoniem.

Otrajā publikācijā apkopoti pētījumi, kas pierāda *E. coli* cAMP receptora proteīna (CRP) piedalīšanos *rafP* regulācijā. Izmantojot ģenētiskas un bioķīmiskas metodes, atrasta CRP proteīna saistīšanās vieta *rafP* sekvencē un izmērīts promotera aktivācijas līmenis. Izpētītas CRP mijiedarbības ar *rafP* DNS un RafR. CRP un RafR spēj vienlaikus saistīt *rafP* DNS, veidojot kompaktu nukleoproteīnu struktūru. kurā CRP piedalās kā *raf* operona korepresors, kas nodrošina gan represijas, gan aktivācijas efektivitāti.

Šūnām augot dabiskajos substrātos, vairāku oglekļa avotu maisījuma klātbūtnē, ne RafR, ne CRP nevarētu saistīt *rafP*. Promotera aktivitāte šādos apstākļos (bazālais ekspresijas līmenis) raksturota trešajā publikācijā, kas sagatavota žurnālam FEMS *Microbiology Letters*. Pirmo reizi parādīts, ka *rafP* bazālā ekspresijas līmeņa nodrošināšanai līdztekus –35 secībai nepieciešams arī A/T bagāts rajons no –59 līdz –39 nukleotīdam, kura funkcijas ir analogas baktēriju konstitutīvo promoteru UP

elementam. Iegūtie rezultāti norāda uz promotera DNS topoloģijas lomu transkripcijas regulācijā.

Plazmīdu replikonu topoloģijas ietekme uz gēnu ekspresijas un rekombinācijas procesu aktivitāti parādīta arī divos ziņojumos Baltijas valstu ģenētikas kongresos, kuru tēzes pievienotas promocijas darbam.

Darbā raksturotas arī *rafP* mijiedarbības ar RNS polimerāzi (RNSP), kā arī CRP un RafR, transkripcijas iniciācijas procesā. CRP ne tikai stimulē RNSP – *rafP* slēgtā kompleksa veidošanos, kas atbilst zināmajam transkripcijas regulācijas mehānismam I. klases CRP aktivētajos promoteros, bet arī modulē slēgtā kompleksa izomerizēšanu par atvērto kompleksu. Represora un RNSP vienlaicīga saistīšanās pie promotera ir līdz šim maz raksturots fenomens. Lai gan RafR inhibē RNSP – *rafP* atvērtā kompleksa veidošanos, tomēr proteīnu vienlaicīga saistīšanās var nodrošināt strauju *raf* operona ekspresiju, ja barotnē parādās melibioze vai rafinoze.

Darba rezultāti ir apkopoti 3 rakstos, no kuriem viens ir sagatavošanā un ziņoti 2 starptautiskās konferencēs kā arī vairākās Latvijas Universitātes Zinātniskajās konferencēs.

Rostoks, N. 1999. Structure and regulation of the raffinose catabolism operon promoter, University of Latvia, Riga, 58 pp.

ABSTRACT

The plasmid-borne raffinose catabolism operon of *Escherichia coli* encodes proteins necessary for inducible uptake of the trisaccharide raffinose and its breakdown to monosaccharides. The operon includes *rafA* (α-galactosidase, α-Gal), *rafB* (permease) and *rafD* (invertase) genes, which are transcribed from a common promoter *rafP*. In the absence of inducer, *raf* repressor, RafR, inhibits the promoter activity upon binding to two operator sites within *rafP*. The only known inducer of the *raf* operon is disaccharide melibiose, which is produced from raffinose by invertase.

This doctoral thesis deals with the structure and topology of the *E. coli* raffinose operon promoter DNA and *rafP* – protein complexes, as well as their role in regulation of the transcription.

The EMSA and DNase I footprinting techniques, as well as the approaches to characterize the topology of protein – DNA complexes, are described in the Publication 1. The cloning of tandem dimers of the *rafP* DNA fragments and site-directed mutagenesis were critically important for characterization of sequence dependent and protein – induced DNA bending and topology.

Publication 2 summarizes the studies on the role of cAMP receptor protein, CRP, in the regulation of *rafP* activity. CRP dependent activation and localization of the CRP site within the *rafP* are determined by combination of genetic and biochemical approaches. CRP interactions with *rafP* and RafR are characterized as well. CRP and RafR can bind to *rafP* simultaneously. Thus a compact nucleoprotein complex is formed, where CRP serves as a co-repressor and ensures both efficient repression and rapid activation of the *rafP*.

The growth of cells in natural environments in presence of mixture of different carbon sources should prevent both CRP and RafR from binding to the *rafP*. The activity of *rafP* under such conditions (basal expression level) has been characterized in the Publication 3, which is prepared for FEMS *Microbiology Letters*. For the first time it has been shown that along with the –35 element, an A/T-rich DNA sequence,

which is positioned between –59 and –39 in respect to mRNA start point, is indispensable for the basal level of *rafP* expression. This sequence functionally resembles the UP element of some constitutively transcribed bacterial promoters. The results show the importance of DNA topology in regulation of transcription.

The influence of the replicon topology on gene expression and recombination events has also been reported at the 1st and 2nd Genetical Congresses of the Baltic States and the corresponding abstracts are appended to this thesis.

The interactions of the RNA polymerase, RNAP, with the *rafP* and the role of CRP and RafR in the transcription initiation are characterized as well. CRP stimulates the formation of the RNAP – *rafP* closed complex in compliance with the model of the transcription activation at the class I CRP-dependent promoters. In addition, CRP modulates the formation of the open promoter complex. The simultaneous binding of RNAP and repressor to the promoter is insufficiently documented phenomenon. Although RafR inhibits the formation of the RNAP – promoter open complex, simultaneous binding may ensure the mechanism of a rapid expression of *raf* operon, if raffinose or melibiose become available in the growth environment.

The results of the work are reflected in three papers and reported at two international workshops, as well as at the Scientific conferences of the University of Latvia.

1. INTRODUCTION

The experimental work presented in this doctoral thesis focuses on DNA elements, protein-DNA interactions and on spatial structure of protein-DNA complexes, which are involved in the regulation of promoter of bacterial *raf* operon, *rafP*.

Expression of sugar catabolism genes in bacteria has served as a paradigm for transcription regulation mechanisms. The transcription initiation even in such a relatively simple organism as *E. coli* is a complex multistage process, which involves RNAP and a number of auxiliary proteins. Induction or repression of certain mRNA synthesis, growth, cell division and sporulation are regulated by intracellular signals, which are generated in response to the changes in environmental conditions. The role of proteins and protein binding DNA sequences in the regulation of transcription is well documented. The significance of the three-dimensional structure of the transcription initiation complex is not so well understood.

The initiation of transcription and its regulation by protein factors and promoter topology in bacterial cells will be described in the following chapters. The structure and regulation of the plasmid-borne *raf* operon will be discussed as well.

1.1. Transcription initiation in Escherichia coli

The transcription process in *E. coli* is initiated by the binding of RNAP to the promoter region of DNA molecule. *E. coli* RNAP core enzyme, which is capable of transcription elongation, has a following subunit structure $\alpha_2\beta\beta'$ (reviewed by Ishihama, 1981). The core enzyme is further joined by one of the 7 different σ (sigma) subunits (Lonetto & Gross, 1996) to form the holoenzyme capable of transcription initiation.

The transcription initiation process is extensively reviewed by Record *et al.*, 1996. Briefly, initiation event can be divided into 4 classes of events according to their key intermediates as depicted in Figure 1:

Class I. Reversible initial specific binding. RNAP holoenzyme binds to the dsDNA of the promoter to form a "closed" complex. The promoter DNA remains completely double stranded.

Class II. Reversible conformational change. The key event of this step is the reversible DNA melting at the start site of transcription, which is driven by the free energy of RNAP binding.

Class III. Reversible binding of initiating NTPs. The NTP, which is complementary to the first nucleotide of template strand, binds at the position +1 of the open complex thus forming the first of several possible initiation complexes. Any of these may be advanced by binding of the next NTP and formation of phosphodiesther bound or, alternatively, may be reversed by either cleavage of a terminal NTP or by release of the short RNA chain, which is called abortive transcript.

Class IV. Transition to elongation (promoter clearance/escape). Once the length of the nascent RNA chain in the initiating complex has reached 7 – 12 nucleotides, the σ subunit is released and RNAP is irreversibly switched to elongation phase.

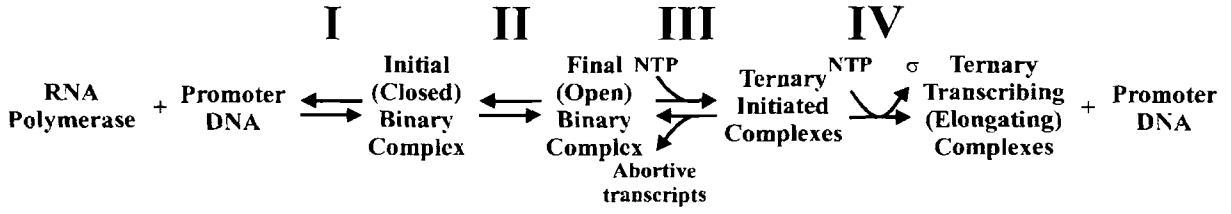


Figure 1. A scheme of the transcription initiation process. Adapted from Record *et al.*, 1996.

The transition from one step to another, except for the step IV, is always reversible process, which may be characterized by dissociation constant as in the classic Michaelis-Menten kinetics. Each of the above steps can be further divided according to the specific intermediates of the process. Both the RNAP and the promoter DNA undergo structural changes during the transcription initiation (reviewed by deHaseth *et al.*, 1998).

The most significant **structural changes in the RNAP** molecule occur during the holoenzyme building, when σ^{70} subunit undergoes a structural change, leading to specific DNA binding propensity (Dombroski *et al.*, 1993; Callaci *et al.*, 1999). The σ^{70} subunit alone is not able to bind the promoter due to intramolecular protein – protein interactions with inhibitory N-terminal domain. Further, during the late stage

of closed complex, while the promoter DNA remains double stranded, the RNAP "jaws", which are formed by β and β' subunits, close around the DNA double helix. This structure of "jaws" is thought to contain the active site of the polymerase and to be open in the holoenzyme (Darst *et al.*, 1989) and to be initially closed in the core enzyme (Polyakov *et al.*, 1995; Darst *et al.*, 1998).

Conformational changes in the promoter DNA take place during the open complex formation. They involve the unwinding of the DNA double helix by half a turn and strand separation from the –10 region till the start site. The energy for formation of an open complex is provided by the RNAP binding. RNAP is known to introduce distortions in the promoter DNA, such as unwinding of double helix (Su & McClure, 1994) and DNA bending (Heumann *et al.*, 1988), which are believed to facilitate the strand opening. Once the open complex has formed, it is stabilized by the specific binding of RNAP to the single stranded DNA of the non-template strand (Marr & Roberts, 1997; deHaseth & Helmann, 1995). The formation of a functional open complex from –12 to +2 also requires Mg^{2+} ions, as well as the binding of NTPs, which contribute to stability of open complex even without hydrolysis and synthesis of phosphodiesther bonds (Zaychikov *et al.*, 1997).

Ternary complexes of promoter, polymerase and NTPs are ready for covalent linking of the first nucleotides. However, the initiation of the RNA strand synthesis does not ensure the successful elongation process. A significant fraction of open complexes repeatedly generate abortive transcripts, RNA oligonucleotides up to 10 n in length. RNAP molecules, which form the open complexes, have a certain probability either of becoming irreversibly trapped in abortive synthesis or of proceeding with the transcription in elongation mode (Kubori & Shimamoto, 1996).

Every step in transcription initiation is characterized by a specific reaction rate and is an eventual subject of regulation. The regulation of transcription initiation is exerted either by the sequence and structure of promoter DNA or by protein factors.

1.2. DNA sequence elements involved in the regulation of transcription initiation

The transcription process is initiated at the specific DNA sequences – promoters. Four sequence elements in RNAP σ^{70} promoters are known (Figure 2): i)

hexanucleotide centered at position –10 in respect to the transcription start site; ii) hexanucleotide centered at position –35; iii) spacer region between –10 and –35 elements; iv) A/T rich region between –40 and –60 called UP element or "third element". Not all of the named sequence elements are necessarily present in every promoter.

The initial RNAP binding to the promoter involves the specific contacts of σ subunit with both the –35 and –10 regions, although only upstream part of –10 element seems to be important for this binding (Dombroski, 1997; Dombroski *et al.*, 1992). Current kinetic investigations support this mechanism rather than the initial binding only at the –35 region, where the contacts with –10 region are necessary for open complex building.

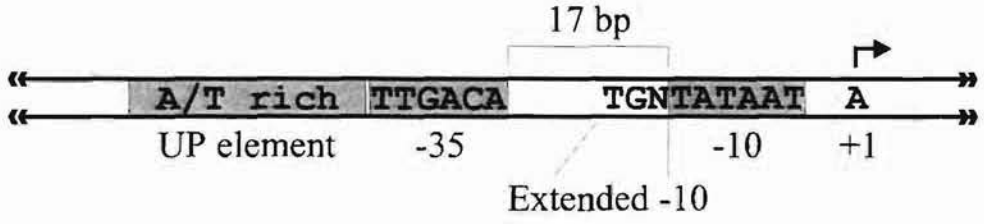


Figure 2. Summary of the RNAP σ^{70} promoter elements (adapted from deHaseth *et al.*, 1998).

Comparative studies have revealed the consensus sequences of –10 and –35 elements (Figure 2). Generally, the promoters that match the consensus sequence better are more proficient both *in vivo* and *in vitro*. Different affinity of RNAP for the promoters with different promoter element sequences sets the basal expression level at widely varying levels.

A consensus spacer of 17 bp in length between –10 and –35 elements has been established. Tentative model explaining the conserved spacer length has been proposed (Ayers *et al.*, 1989). The transient alignment of the –10 and –35 regions upon RNAP binding and torsional strain of the spacer DNA is postulated. The free energy of the strain facilitates the formation of promoter open complex.

A subclass of *E. coli* promoters exists without a recognizable –35 element. Such promoters have been found to have an "extended" –10 region (Figure 2) (Kumar *et al.*, 1993). The recognition of an "extended" promoter by RNAP involves different

region of the σ subunit (Barne *et al.*, 1997). In absence of any additional factors, their activity is comparable to the promoters with standard –35 elements.

An additional DNA element, which may significantly increase the promoter activity, has been identified in several promoters quite recently (Ross *et al.*, 1993). This promoter module is often called an UP element or "third element". It is A/T rich sequence located between positions –60 and –40 in respect to transcription start. SELEX procedure has been used to deduce the UP element consensus sequence (Estrem *et al.*, 1998). UP element functions by contacting the C-terminal domain of the RNAP α subunit and stabilizing the binding of RNAP to the promoter (Rao *et al.*, 1994; Ross *et al.*, 1998). The same RNAP α subunit protein surface region is responsible for contacts with the transcription activating proteins, such as CRP, and with the DNA of UP element (Gaal *et al.*, 1996; Murakami *et al.*, 1996).

The topology of promoter DNA may also contribute to the regulation of promoter activity. It has been shown that curved DNA inserts can activate a promoter by mimicking the CRP induced DNA bending *in vivo* (Bracco *et al.*, 1989) and *in vitro* (Gartenberg & Crothers, 1991). The transcription activation by the curved DNA is attributed to the facilitated wrapping of the promoter DNA around RNAP (Gartenberg & Crothers, 1991) or to the preferential binding of the RNAP α subunit to the A/T rich bent DNA sequences (Aiyar *et al.*, 1998). The creation of transcriptionally competent promoter architecture and the translation of environmental signals into modulated promoter response are reviewed in Pérez-Marin & deLorenzo (1997).

1.3. Regulation of transcription initiation by protein factors

The transcription initiation at the majority of the *E. coli* promoters is regulated by different protein factors. If gene expression is switched on or off in response to appropriate signals, the regulation may be provisionally characterized as positive or negative control. Proteins additionally regulate even the few inherently strong promoters, e.g., *rrnB P1* transcribing the rRNA. The mechanisms of action of the transcription factors are extremely diverse and may affect any step in the process of transcription initiation. The same protein factors may participate both in transcription activation and repression depending on the physiological conditions in the cell.

1.3.1. Positive control

The typical transcription activator binds to its specific site at or near promoter DNA and contacts the transcription machinery. Bacterial transcription activators may facilitate any step in initiation process, although the most common is the stabilization of initial RNAP binding to promoter and stimulation of closed complex isomerization to an open complex.

The *E. coli* cAMP receptor protein, CRP, also referred to as catabolite gene activator protein, CAP, is one of the best studied transcription activators (reviewed by Kolb *et al.*, 1993; Busby & Ebright, 1997). In response to increased cAMP level, which serves as a co-activator, CRP activates transcription initiation at more than 30 *E. coli* promoters by binding to its cognate sites located at various distances upstream from core promoter sequences. CRP is a specific DNA binding protein consisting of two identical subunits, whose C-terminal domains contain HTH DNA binding motifs (Schultz *et al.*, 1991). HTH motif of each subunit of CRP binds to one half site of the palindromic DNA sequence (Gunasekera *et al.*, 1992). The X-ray analysis of CRP – consensus site DNA co-crystals has revealed two major kinks of ca. 40° each contributing to overall ca. 90° DNA bending induced by CRP and DNA wrapping around CRP (Schultz *et al.*, 1991).

Promoters activated by CRP are classified according to the distance between the center of CRP site and the transcription start site. Promoters with CRP site centered at position –61.5 are termed **Class I CRP-dependent promoters**. At Class I CRP-dependent promoters, e.g., *lacP1* of the lactose catabolism operon, CRP binds adjacent to RNAP and exerts its activating effect through contacting the C-terminal domain of the RNAP α subunit (Figure 3A). The surface exposed loop on CRP molecule (activating region 1 or AR1) formed by amino acids 156-162 is responsible for contacts with RNAP, which result in improved RNAP binding to the promoter and more stable closed complex. Although CRP is a homodimer, only the AR1 of the promoter - proximal subunit contacts RNAP (Zhou *et al.*, 1993). CRP also activates promoters from positions upstream the –61.5, e.g., -71.5, -82.5 and -92.5 (Gaston *et al.*, 1990; Ushida & Aiba, 1990). The flexibility of the interdomain linker of the RNAP α subunit is likely to enable the C-terminal domain of the α subunit to contact CRP located at different distances (Jeon *et al.*, 1997). The action of CRP bound to Class I promoter is an example of transcription activation by recruitment of RNAP to

the promoter. This strategy is successfully used by a number of prokaryotic and eukaryotic transcription factors (Dove *et al.*, 1997; reviewed by Ptashne & Gann, 1997).

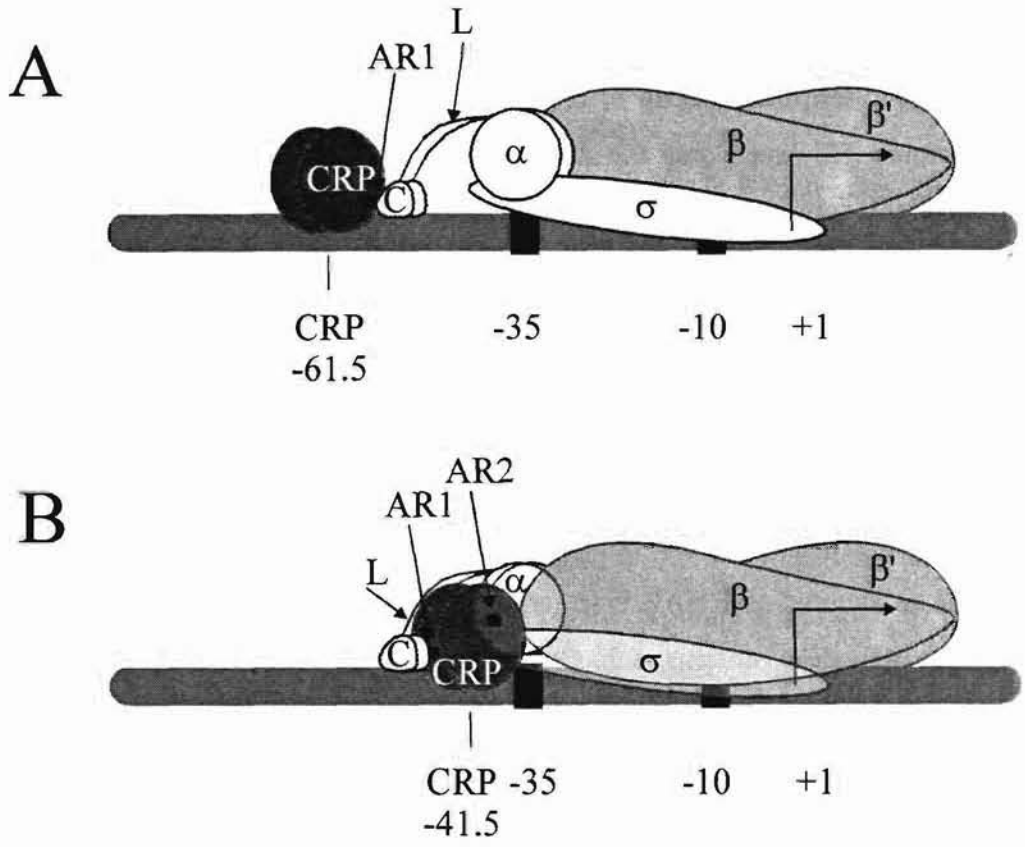


Figure 3. A scheme of CRP-mediated transcription activation at A, Class I and B, Class II CRP-dependent promoters. "α" denotes the α subunit of RNAP, which is connected to its C-terminal domain "C" via flexible interdomain linker "L". "σ, β and β'" denote the respective subunits of RNAP. At the Class I promoters transcription activation is mediated by protein – protein interactions between AR1 of the *downstream* subunit of CRP and the C-terminal domain of the RNAP α subunit. At the Class II promoters transcription activation is achieved by 2 different protein – protein contacts: i) interactions between AR1 of the *upstream* subunit of CRP and the C-terminal domain of the RNAP α subunit; ii) interactions between AR2 of the *downstream* CRP subunit and the N-terminal domain of the RNAP α subunit (transparent in the drawing). Adapted from Busby & Ebright, 1997.

The promoters, where CRP binds to the sites centered at position –41.5 as in the case of *gal* promoter *P1*, are named **Class II CRP-dependent promoters**. In contrast to Class I promoters, RNAP contacts DNA both upstream and downstream of CRP bound at position –41.5 (Attey *et al.*, 1994; Belyaeva *et al.*, 1996) (Figure 3B). Two contacts of the CRP with RNAP α subunit are responsible for activation, each contact influencing a different step in transcription initiation (Savery *et al.*, 1998). The contact of the C-terminal domain of α subunit with AR1 improves the formation of closed complex. However, only the AR1 on the upstream CRP subunit is contacted and a different subset of amino acid residues is essential (Zhou *et al.*, 1994). Additional positively charged activating region on the CRP surface opposite the DNA binding domain has been identified comprising amino acids 19, 21 and 101 and is denoted as AR2. Only the AR2 of the downstream CRP subunit is involved in the contacts with negatively charged region on the N-terminal domain of RNAP α subunit (Niu *et al.*, 1996). This contact is implicated in stimulation of closed complex isomerization into open promoter complex (Rhodius *et al.*, 1997).

The α subunit of RNAP is a target not only for CRP, but also for other transcription activating proteins, e.g., OxyR (Tao *et al.*, 1995) and FIS (Bokal *et al.*, 1997) (reviewed by Ishihama, 1993), although the different amino acid subset on α subunit is contacted in each case (Ebright, 1993; Bokal *et al.*, 1995).

The contacts with transcription activators are not limited only to the RNAP α subunit. For instance, α subunit contacts with UP element DNA are critical in initial binding of RNAP to *ada* and *aidB* promoters (adaptive response genes). For further steps in transcription initiation the contacts between transcription activator Ada and C-terminal region of the σ^{70} subunit of RNAP are required (Landini *et al.*, 1998). Another example of σ^{70} being a target for transcriptional activator is the interaction of λcI protein with C-terminal region of σ^{70} at λP_{RM}, where the isomerization to open complex is stimulated (Kuldell & Hochschild, 1994; Li *et al.*, 1994).

RNAP β' subunit, which is primarily involved in the catalytic activity of the polymerase (Zaychikov *et al.*, 1996), may also be a target for transcriptional activator. The bacteriophage N4 single stranded DNA binding protein (SSB) interacts with the highly conserved C-terminal region of β' subunit of promoter-bound RNAP (Miller *et al.*, 1997). The DNA binding activity of SSB is not required for activation.

The activators may exert their function by contacting different subunits or different domains of a single subunit of RNAP. The steps of transcription initiation influenced by each of these contacts also may be different. The activator may bind to the multiple sites in vicinity of the promoter and act synergistically on transcription activation. Moreover, different activators may bind together to the promoter and activate the transcription by contacting RNAP at several positions, e.g., CRP and FNR (Scott *et al.*, 1995), CRP and λcI (Joung *et al.*, 1994).

The transcription activation provided by promoter UP element may also be further increased by binding of an activator protein upstream from UP element, as exemplified by binding of FIS protein to the *rrnB P1* (Newlands *et al.*, 1992).

1.3.2. Negative control

The term negative control is often applied to global silencing events as DNA methylation, transcription attenuation, etc., which will not be discussed here. Several mechanisms of the transcription repression by the proteins are known (reviewed by Choy & Adhya, 1996). The typical molecular approaches to achieve the protein-mediated repression of promoter activity in bacterial cells will be outlined here.

Steric hindrance seems to be the most straightforward way to bring about repression, i.e., binding of repressor protein to its cognate site(s) overlapping the promoter. This leads to the competition in binding of repressor and RNAP, and since the affinity of repressor for its site is usually higher, the RNAP binding is occluded. Although the most obvious strategy for repression, this is not as widely applicable as could be expected. λcI action at the λP_R promoter is one of the rare experimentally established systems, where the co-operative binding of λcI to adjacent sites blocks the formation of closed promoter complex (Hawley *et al.*, 1985).

Protein – protein interactions contribute to the transcription repression by affecting the steps beyond the initial binding of RNAP. This implies simultaneous binding of repressor and RNAP to the promoter. Direct protein – protein interactions may be responsible for the repression of any of steps of transcription initiation beyond the RNAP binding to promoter.

RNAP caging may occur at some promoters, when RNAP makes additional contacts with DNA upstream the –35 element and downstream the transcription start site. This should eventually lead to DNA wrapping around RNAP (polymerase

caging) (Rees *et al.*, 1993). The transcription activator proteins may participate in building of RNAP caging structures. The repressors may prevent the critical RNAP – DNA contacts.

Multipartite operators and DNA looping. Transcription control by a repressor binding to a single operator site is relatively rare event. Instead, several operator sites are often used by repressor, located even at a great distance from promoter (Collado-Vides *et al.*, 1991; Gralla & Collado-Vides, 1996). These operators may have different affinities for their repressors with correspondingly variable effects on transcription initiation. Overlapping of multiple operators with promoter sequences usually leads to the expectable result – repression by steric hindrance. However, in cases when operator sites are located at a distance from promoter and from each other, different strategy seems to be responsible for repression. The ability of repressor proteins to interact may generate a loop in the DNA, where the RNAP becomes trapped (Ptashne, 1986), or cause the promoter DNA to undergo structural changes, which do not permit RNAP binding.

Repression by antiactivation is a common strategy in eukaryotic systems, although it is known to occur also in bacteria. Repressor may interact with the transcription activating proteins at the intrinsically weak promoters and impair their ability to bind the DNA or interact productively with RNAP. The classic example is the regulon of nucleoside transport and biosynthetic operons, which is negatively controlled by CytR repressor and activated by CRP. CytR action at promoters of *deo*, *cdd* and *tsx* operons consists of preventing CRP from activation (Valentin-Hansen *et al.*, 1996). Since CytR itself has only weak DNA binding ability, the co-operative binding with two CRP molecules is required to occlude RNAP from binding to *deo P2* (Pedersen *et al.*, 1991; Mollegard *et al.*, 1993).

As paradoxical as it may be, the involvement of activator proteins in the repression complexes, seems to be rather rule than exception. CRP, FNR (transcription factor homologous to CRP), OxyR and TyrR may all act as activators or repressors depending on particular promoter and physiological conditions of the cell (Gralla & Collado-Vides, 1996). FNR and OxyR autoregulate the expression of their respective genes by binding to the promoters around the transcription start site. CRP, for instance, acts as a repressor at the galactose operon promoter *gal P2* by binding to the sequence corresponding to the –35 region of the *P2* and switching thus the transcription almost completely to *gal P1* (Weickert & Adhya, 1993).

The most popular and probably the best-studied transcription repressors belong to the family of regulators homologous to Gal and Lac repressors (Weickert & Adhya, 1992). The mechanism of action of these repressors has been extensively studied and detailed models have been proposed (Lewis *et al.*, 1996; Aki *et al.*, 1996), although these models have already been questioned (Perros & Steitz, 1996).

The lactose operon repressor LacI controls the transcription from *lac* promoter *P1* by binding to its operator sites centered at positions +11 (O_1), +400 (O_2) and –82 (O_3) relative to the transcription start site (Reznikoff *et al.*, 1974). The principal operator site is O_1, which also has the highest affinity for LacI. The co-operative binding of LacI to O_1 and O_2 or O_3 provides highly efficient repression by forming a DNA loop (Oehler *et al.*, 1990). The ability to introduce loop in the promoter DNA depends on the capacity of LacI to form tetrameric structures, although the DNA binding form for all the repressors from Lac and Gal family is the dimer. The ability of dimers to interact resides in the specific amino acid sequences at the C-terminus of LacI (Alberti *et al.*, 1991). The different experimental approaches have yielded several explanation of the mechanism of repression by LacI, including inhibition of open promoter complex formation (Straney & Crothers, 1987) and prevention of RNAP from promoter clearance (Lee & Goldfarb, 1991). The recent kinetic studies indicate, however, that LacI affects the initial binding of RNAP and formation of closed complex (Schlax *et al.*, 1995). The current repression model by Lewis *et al.* (1996) does not allow to decide in favor of any of these theories. It does, however, include CRP in repression complex as a tool to facilitate the DNA loop formation between O_1 and O_2, which is supported by co-operative binding of CRP and LacI to their sites at the *lac* promoter (Hudson & Fried, 1990).

GalR is one of the two repressors controlling the *gal* regulon of *E. coli* (Weickert & Adhya, 1993), which encodes enzymes for galactose transport and catabolism. Two overlapping promoters of *gal* regulon *P1* and *P2* are repressed by binding of repressor GalR to two operator sites centered at –60.5 (O_E) and +53.5 (O_I) (Irani *et al.*, 1983). The necessity of DNA looping by GalR for full repression of *P1* and *P2* has been demonstrated (Haber & Adhya, 1988; Choy & Adhya, 1992). The DNA looping and full repression, however, was achievable *in vivo* only, but not *in vitro* (Mandal *et al.*, 1990). Since GalR lacks the leucine minizipper region at the C-terminal domain of protein, which is responsible for LacI tetramerization (Alberti *et al.*, 1991; Weickert & Adhya, 1992), some additional factor must be responsible for

GalR tetramerization. Recently *E. coli* histone-like protein HU was identified as an auxiliary factor necessary for loop formation and complete repression of *gal* promoters (Aki *et al.*, 1996). HU is a basic heterodimeric protein in *E. coli* with the molecular weight around 19 kDa, which is capable of non-specific DNA binding and bending (Schmid, 1990). The occupation of both operator sites by GalR is required for HU binding to *gal* promoters around the transcription start and is necessary for building a higher order nucleoprotein structure responsible for repression of open complex formation (Aki *et al.*, 1996; Aki & Adhya, 1997).

The raffinose operon repressor RafR also belongs to the Lac and Gal repressor family (Weickert & Adhya, 1992).

1.4. Structure of the plasmid-borne raffinose operon of E. coli

The raffinose catabolism operon, which is subject of this study, originated from the natural *E. coli* strain D1021 (Orskov & Orskov, 1973) harboring plasmid pRSD2 (Burkardt *et al.*, 1978). The organization of the operon is shown in the Figure 4. Structural genes *rafA*, *rafB* and *rafD*, which encode α-galactosidase, *raf* permease and invertase, respectively, are co-transcribed from a common promoter *rafP* (Aslanidis *et al.*, 1989).

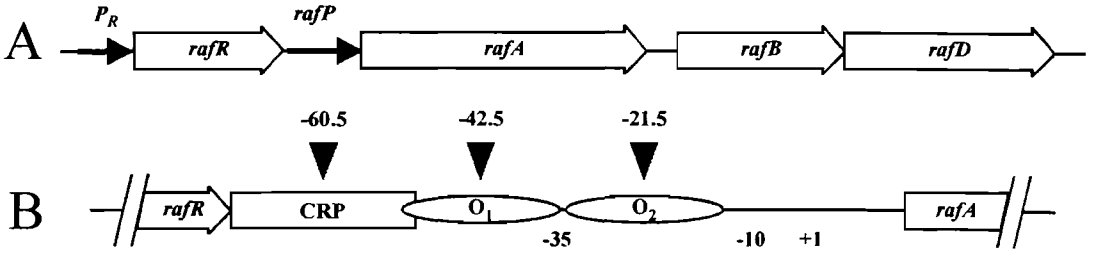


Figure 4. A, Schematic representation of the structure of raffinose catabolism operon and B, of the *rafR – rafA* intergenic region containing the *rafP* and regulatory elements.

The regulation of the operon expression is achieved by the *raf* repressor, RafR. The repressor is encoded by *rafR* gene, which is located immediately 5' to *rafP*. RafR binds to two 18 bp palindromic operator sites O_1 and O_2 centered at the positions –42,5 and –21.5 in respect to *raf* mRNA start point (Aslanidis & Schmitt, 1990).

RafR binds the DNA as a homodimeric molecule. Although the evidence for RafR tetramerization has been presented (Jaenicke *et al.*, 1990), the unassisted *in vivo* formation of RafR tetramers is not plausible due to the lack of the C-terminal leucine minizipper domain.

Binding of RafR to its cognate operator sites is successive and there is no preference in binding affinity for each of the operator sites (Aslanidis *et al.*, 1990). Binding to the operator sites occurs in two consecutive steps, first being RafR binding to any one of the operators and second being the complete occlusion of both operator sites. Since the both operator sites are nearly equivalent in binding affinity, the first RafR dimer is suggested to "oscillate" between two operator sites (Muiznieks & Schmitt, 1994). The binding of the second RafR dimer is hampered ca. 13-fold, which can be explained by steric hindrance (Muiznieks & Schmitt, 1994), because the centers of operator sites are separated by only 2 helical turns and located on the same side of the DNA double helix. The occupation of both operator sites is required for complete ca. 1200-fold repression *in vivo*. The O_1, which is located upstream the –35 element, alone yields only ca. 70-fold repression (6%). The O_2 alone, which is located between –35 and –10 elements, has much more significant impact on promoter activity providing ca. 550-fold repression (45%) (Muiznieks & Schmitt, 1994). The successive binding of RafR to *rafP* obviously provides a tool for step-wise down-regulation of the promoter.

Peculiarity of the *raf* system is its specific inducer melibiose, which results from raffinose breakdown (Aslanidis *et al.*, 1990). The inducer is produced by the invertase, which is encoded by the third gene in operon, *rafD*, and splits the raffinose into melibiose and fructose. This implies that in spite of tight repression ensured by RafR, some basal expression level of the operon should exist. The transport of raffinose into the bacterial cell may be assisted by *lac* permease as well. The invertase has higher affinity for raffinose as the first enzyme in operon – α-Gal (K=4 mM for invertase and K=60 mM for α-Gal) (Schmid & Schmitt, 1976). Translation coupling of *rafB* and *rafD* genes is another adaptation, which may facilitate initial level of raffinose breakdown to achieve subsequent induction (Aslanidis *et al.*, 1989).

Similarly to other sugar catabolism operons, *raf* operon is not only repressed in absence of appropriate substrate or induced by its availability. After the induction has taken place, it can be activated in response to the low glucose level in the growth

media. Genetic studies have indicated the involvement of cAMP-CRP complex in the regulation of raffinose operon (Su *et al.*, 1989).

The previous investigations have shown that the *rafP* of *E. coli raf* operon may be considered as a representative example of bacterial sugar catabolism operon promoters. Although extensively studied, similar promoters from *lac* and *gal* operons still deliver exciting information and unexpected facts about the mechanism of the transcription regulation. Detailed study of *rafP* could be useful in supplying further knowledge on the universal principles of the regulation of gene expression and in disclosing individual molecular adaptations to solve particular tasks of metabolism.

In order to understand the determinants of control of the raffinose operon expression, a study was initiated with an aim to characterize the structure and regulation of transcription initiation of the plasmid-borne raffinose catabolism operon promoter *rafP*.

The following specific tasks were set for this study:

1. Investigation of the interactions of *rafP* with specific transcription regulating proteins RafR and CRP.
2. Studies of RafR and CRP interactions on *rafP* DNA.
3. Investigation of RNAP interactions with *rafP* and the influence of RafR and CRP on these interactions.
4. Determination of the role of different *rafP* modules in establishing the basal level of promoter activity.

2. MATERIALS AND METHODS

2.1. Strains and plasmids

Strains and plasmids used throughout the work are listed in the Table 1.

Table 1. Strains and plasmids

Strain / Plasmid	**Purpose / relevant genotype**	**Reference**
CA8000	α-Gal activity measurements / *crp*$^+$, *cya*$^+$	Sabourin & Beckwith (1975)
CA8306	α-Gal activity measurements / *crp*$^+$, *cya*$^-$	"
DH1	RafR induction experiments with melibiose / *lac*$^+$	Hanahan (1983)
XA3D1	α-Gal activity measurements / *crp*$^-$, *cya*$^+$	Breul *et al.* (1993)
XL1-Blue	General cloning and plasmid DNA preparation	Bullock *et al.* (1987)
pBG2	Overexpression of CRP	Breul *et al.* (1993)
pRU984	Overexpression of RafR	Aslanidis *et al.* (1990)
pRU1301 – pRU1305	Different length deletions of the 5' portion of wt *rafP* sequence in pRU1307	Muiznieks *et al.*, 1999
pRU1307	wt *rafP* and *rafA* gene cloned in pUC8	Muiznieks & Schmitt, 1994
pRU1307B; pRU1307C	Mutagenesis of –35 promoter element in pRU1307 *rafP* sequence	Manuscript 3
pRU1307E, M, N, X	Mutagenesis of pRU1307 *rafP* sequence creating restriction enzyme sites	Muiznieks *et al.*, 1999
pRU1324; pRU1327	Plasmids carrying *rafR-rafP-rafA* gene cluster with mutagenesis-inactivated O_1 and O_2 sites respectively	Muiznieks & Schmitt, 1994
pRU1341 – pRU1344	pRU1307M derivatives with different distances between CRP and O_2	Muiznieks *et al.*, 1999

pRU1340; pRU1350-1; pRU1360-1	Mutagenesis of O_1 region in pRU1307 to disclose the location of the UP element in *rafP*	Manuscript 3
pRU1369; pRU1369+4; pRU1371; pRU1371+5; pRU1372	Oligonucleotide substitutions of the *rafP* upstream region from position –133 till -28	Manuscript 3
pSA508	*In vitro* transcription assay	Choy & Adhya (1993)
pSA508*rafP*	*In vitro* transcription from *rafP*	N. Rostoks, unpublished data

2.2. Bacterial growth media

Bacterial cells were cultured either in liquid 2 x TY or on agarized LB media, which were supplemented with antibiotics and 0.1% melibiose or fructose (Maniatis *et al.*, 1989).

2.3. Chemicals and enzymes

Radiochemicals were from Amersham Pharmacia Biotech Europe GmbH (Freiburg, Germany).

DNA restriction and modification enzymes were from MBI Fermentas (Vilnius, Lithuania), New England Biolabs (Beverly, MA, USA) and from Boehringer Mannheim (Mannheim, Germany), and were used according to manufacturers' recommendations.

E. coli RNAP holoenzyme was from Amersham Pharmacia Biotech AB (Uppsala, Sweden).

2.4. Plasmid construction and mutagenesis

The standard cloning procedures followed Maniatis *et al.* (1989). Site-directed mutagenesis was done by Amersham Sculptor *in vitro* mutagenesis kit (Amersham

Pharmacia Biotech Europe GmbH, Freiburg, Germany) or according to the QuickChange™ Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA).

2.5. α-Gal activity measurements

The determination of enzymatic activity of *raf*A gene product α-Gal was performed as described in Muiznieks & Schmitt (1994). Activity is expressed as the maximum increment in specific activity during the exponential growth of bacterial cultures. One arbitrary unit of enzyme activity corresponds to a change of 0.01 OD_{420} (absorption maximum for *p*-nitrophenol liberated by hydrolysis of *p*-nitrophenyl–α–D-galactopyranoside) per min per 1.0 OD_{600}. The relative plasmid dosage in bacterial cultures was determined according to Stueber & Bujard (1982) and enzyme activity values were normalized accordingly. At least three independent assays were performed for each experiment. Estimated error is ca. 20%.

2.6. DNA sequencing

Dideoxy DNA sequencing was carried out using Thermo Sequenase Radiolabelled Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech Europe GmbH, Freiburg, Germany).

2.7. Protein preparations

Preparation of RafR and CRP has been described in Muiznieks & Rostoks (1998). Briefly, RafR was overexpressed in the form of inclusion bodies in *E. coli* XL1-Blue harboring plasmid pRU984 and was purified as described in Aslanidis *et al.* (1990). RafR inclusion bodies were solubilized in 0.1% SDS or 0.3% N-lauroylsarcosine and 1:100 till 1:1000 dilutions were used directly for EMSA and DNase I footprinting. Protein preparations were typically more than 90% pure as judged by SDS-PAGE. Presence of solubilizing agent in the protein dilutions did not measurably interfere with subsequent protein-DNA binding assay.

CRP was overexpressed in *E.coli* XL1-Blue harboring plasmid pBG2. Both crude cell lysates and affinity purified CRP (Ghosaini *et al.*, 1988) were used for EMSA and DNase I footprinting with no noticeable difference in DNA binding. It

was necessary, however, to use a binding buffer without Mg^{2+} when employing CRP-enriched crude cell lysates due to the presence of Mg dependent endo- and exonucleases and proteases.

Protein concentrations were determined according to Bradford (1976).

2.8. Electrophoretic mobility shift assay

EMSA was carried out by a modified method of Fried & Crothers (1981) as described in Muiznieks & Rostoks (1998). The most significant modification included the composition of the protein - DNA binding buffer. The optimal buffer for both CRP and RafR binding included 10 mM Tris-HCl (pH 8.0), 10 mM KCl, 1 mM EDTA, 1 mM DTT, 50 μg/ml BSA, 3% glycerol and 0.01% Nonidet P-40. When *E. coli* RNAP binding to *rafP* DNA was desired, alone or together with CRP or RafR, the following buffer was used: 20 mM Tris-acetate (pH 8.0), 3 mM Mg acetate, 50 mM KCl, 1 mM DTT, 50 μg/ml BSA and 0.05% Nonidet P-40.

The protein – DNA complexes were formed at 20° C in the total volume of 20 μl for 20 min. and reactions mixtures were immediately loaded onto the gel without the addition of dye markers.

Gel electrophoresis was carried out as described in Muiznieks & Rostoks (1998) except that 0.5 x TAE buffer or 1 x TE buffer (pH 8.0) was used as electrophoresis buffer. These buffers were changed after pre-electrophoresis to prevent buffer exhaustion. Buffers were supplemented with 20 μM cAMP, if binding reactions included CRP.

2.9. DNase I footprinting

DNase I protection assay was performed as described previously (Aslanidis *et al.*, 1990; Muiznieks & Rostoks, 1998).

Protein - DNA complexes were formed as for EMSA in the reaction mixture of 100 μl. The amount of DNase I was determined empirically for every DNA preparation. Footprinting reactions were run on 8% sequencing gel along with the A and G or C and G chemical sequencing reactions, in order to localize the protected

bases. A and G sequencing reactions were prepared by a simplified approach (Muro *et al.*, 1993).

2.10. Hydroxyl radical footprinting

Hydroxyl radical footprinting was performed essentially according to the procedure developed by Tullius & Dombroski (1986), which is described in details by Dixon *et al.* (1991). Protein - DNA binding reactions were formed as described for DNase I footprinting, except that glycerol was excluded from binding reactions, because of its ability to quench hydroxyl radicals.

2.11. Potassium permanganate probing of DNA structure

Potassium permanganate probing of open promoter complexes was performed as described by Gralla *et al.* (1993) and McCarthy *et al.* (1990) under the same conditions as described below for *in vitro* transcription assay.

2.12. In vitro run-off transcription assay

In vitro transcription assay was performed using a technique described by Garges *et al.* (1995) on supercoiled plasmid template carrying the *rafP* and downstream located transcription terminator. For this purpose *rafP* region was PCR-amplified and cloned between the *Eco*RI and *Pst*I sites of the plasmid vector pSA508 (Choy & Adhya, 1993; pSA508 was a kind gift from Dr. H.E. Choy). The reaction conditions were as described in Garges *et al.* (1995).

2.13. Estimation of DNA curvature

Sequence dependent DNA curvature of *rafP* and its derivatives was characterized by the analysis of DNA fragment electrophoretic mobility in native polyacrylamide gels (PAAG) as described in Muiznieks & Rostoks (1998). Intrinsically bent DNA fragments migrate in PAAG electrophoresis slower than straight molecules of identical sequence length. Coefficient K_R is calculated as a ratio of apparent fragment length estimated in PAAG electrophoresis versus its sequence

length. K_R is inversely correlated to the fragment end-to-end length and may be used to estimate the angle of DNA bending (Thompson & Landy, 1988).

2.14. Processing of the images

All the autoradiographs were scanned and the resulting images were inserted in the figures further processed in the Corel Draw 6.0. No filters were applied to scanned images and no other modifications were made, except in some instances the irrelevant parts of the images were cropped (Figures 6, 10, 12, 13).

3. RESULTS AND DISCUSSION

The *rafP* promoter of *E. coli* plasmid-borne raffinose operon is a unique example of compact placement of diverse regulatory sequences within a short, 70 bp DNA sequence.

We have analyzed CRP, RafR and RNAP interactions with the *rafP* (sections 3.1. – 3.7.). DNA sequence elements and topology constraints of the *rafP* upstream region, which are required for basal level expression of the promoter, are described in section 3.8.

3.1. Localization of the CRP site in the rafP

Preliminary results have shown the involvement of CRP in regulation of plasmid-borne *raf* operon in *E. coli* (Su *et al.*, 1989).

The alignment of the *rafP* (Aslanidis & Schmitt, 1990) with the consensus sequence of CRP (Ebright *et al.*, 1984) revealed 64% sequence match upstream of the O_1 site (Figure 5B). The putative CRP site exhibited incomplete dyad symmetry with a canonical right half-site and imperfect left half-site. For functional mapping of the site, a set of deletion derivatives of the *rafP* was constructed (Figure 5A; Table 1) and the promoter activity was tested in different genetic backgrounds (Table 1 in Muiznieks *et al.*, 1999).

These experiments delimited the boundaries of CRP site. Deletion upstream of the position –72 (pRU1304) does not significantly influence the activity of promoter, since the construction retains more than 90% of the wt activity. Deletion of the right half-site of the putative CRP site (pRU1303) reduces the activity of promoter by 80%. Deletion of the whole CRP site (pRU1302) leaves ca. 1% of the wt *rafP* activity. Further decrease in α-Gal accumulation is observed, if the basic promoter elements are eliminated (pRU1301 and pRU1300).

The plasmids with wt *rafP* sequence and their deletion derivatives, where the putative CRP site was eliminated, produced comparable α-Gal amounts in the mutant bacterial strains, which were deficient in CRP (XA3D1) or adenylate cyclase (CA8306). On the average, functional CRP activated the wt *rafP* ca. 30- to 50-fold. Genetic and biochemical evidence suggested the involvement of cAMP-CRP complex

in activation of *rafP* by binding to its sequence between positions –72 and –50 (Muiznieks *et al.*, 1999).

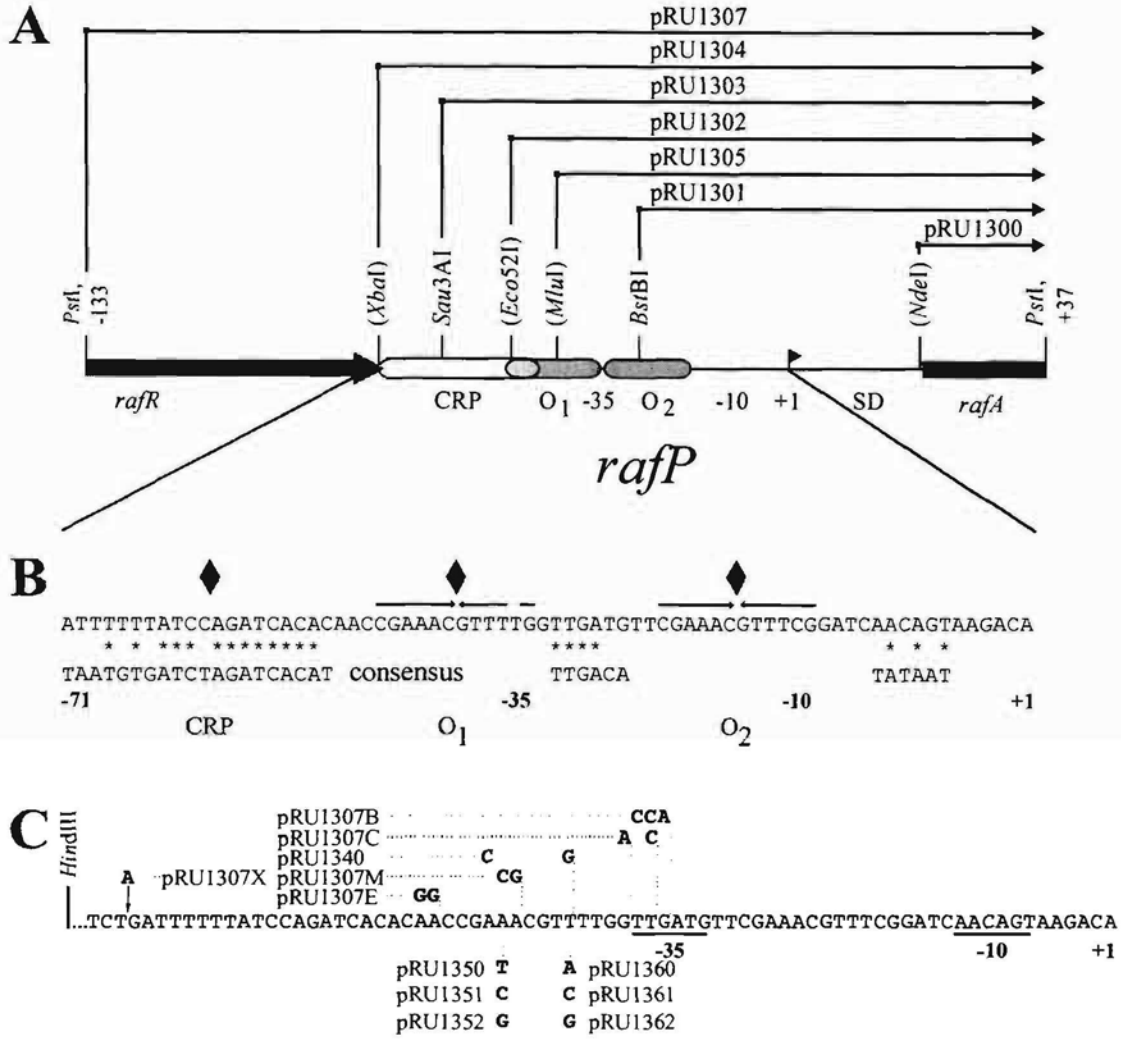


Figure 5. Structure of the wt *rafP* region and the promoter derivatives. A, Map of the 170 bp *Pst*I fragment containing the *rafP* and the fragments of adjacent *rafR* and *rafA* genes. CRP – CRP binding site; O_1 and O_2 – *raf* operator sites; -35 and –10 – *rafP* promoter elements; +1 – transcription start site; SD – Shine – Dalgarno sequence. The *rafP* sequences retained in the plasmids with truncated promoter (pRU1300 – pRU1305) and the plasmid with complete wt *rafP* (pRU1307) are indicated above the map. Description of the plasmids is given in Table 1. Relevant restriction sites are shown, those introduced by site-directed mutagenesis are in parentheses. B, The nucleotide sequence encompassing CRP and both operator sites. Inverted arrows indicate dyad symmetry of RafR binding sites and filled diamonds

show the symmetry axes of the protein binding sites. Consensus sequences of CRP site and basic promoter elements are given below the sequence. Asterisks indicate identical nucleotides in the consensus and *rafP* sequences. C, Nucleotide substitutions and one insertion of A are shown above and below the sequence. The plasmids, which carry respective mutations are denoted (see also Table 1).

The sequence bound by CRP was further defined by DNase I (Figure 2 in Muiznieks *et al.*, 1999) and hydroxyl radical footprinting techniques (Figure 6), which yielded similar results. As expected, the sequences protected by protein are shorter at hydroxyl radical than at DNase I probing due to the ability of the radicals to penetrate DNA - protein complexes more deeply than it is possible for the bulky enzyme. Hydroxyl radical footprinting delineates also the boundaries of the left CRP half site, which is composed of a run of T residues and, therefore, is intrinsically resistant to DNase I attack. Two hypersensitive sites are observed in DNase I footprinting at positions –64 and –55 corresponding to two kinks introduced by CRP binding. CRP site in *rafP* is centered at the position –60.5 in respect to transcription start site. More pronounced protection of the right half site, as well as the downstream extension of the protection, may be explained by higher homology of the *rafP* CRP right half site to the consensus sequence.

DNase I footprinting demonstrated also the simultaneous binding of CRP and two RafR dimers to their cognate sites on *rafP* DNA (Muiznieks *et al.*, 1999). Two RafR dimers cover the region between positions –53 and –7 containing both operators and –35 promoter element. RafR introduces a new DNase I hypersensitive site at the 5' boundary of O_1. The distance between the centers of CRP and O_1 sites is 18 bp and between centers of O_1 and O_2 sites - 21 bp. Therefore, two RafR dimers bind the same face of DNA double helix, whereas the center of the CRP site is rotated by 103° in respect to the centers of RafR binding sequences.

Addition of specific inducer melibiose reverses the operator binding of RafR. The simultaneous binding of CRP and RafR, however, stabilizes RafR against the induction. This suggests that CRP may act as a co-repressor under these conditions. This assumption is corroborated by finding that α-Gal synthesis is repressed more efficiently in *crp*$^+$, *cya*$^+$ cells than in their mutant counterparts (Table 1 in Muiznieks *et al.*, 1999, pRU1324).

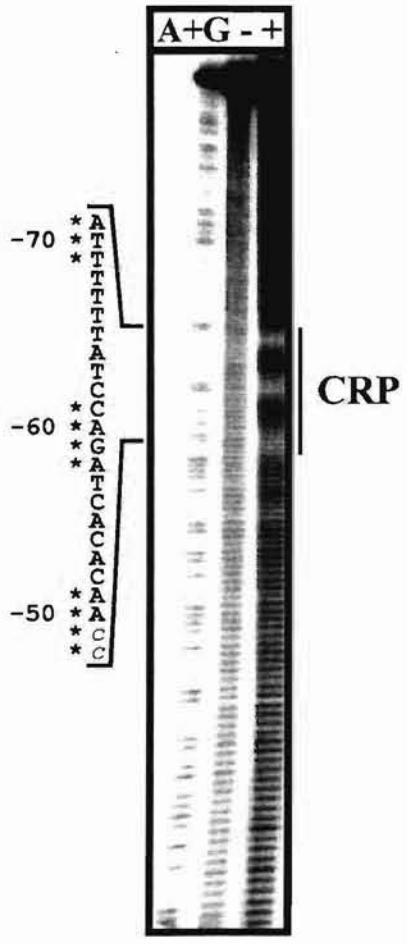


Figure 6. Hydroxyl radical footprinting of CRP – wt *rafP* complex. Lane "A+G" is chemical sequencing reaction; lane "-" denotes the reactivity of the free DNA; lane "+" shows the reactivity of CRP – wt *rafP* DNA. The sequence involved in CRP binding is shown on the left, with the bases specifically protected by CRP marked with asterisks.

3.2. Binding of CRP and RafR to the rafP

EMSA of CRP and RafR complexes with *rafP* DNA demonstrated that CRP, O_1 and O_2 sites can be occupied simultaneously by their cognate proteins (Figure 3 in Muiznieks *et al.*, 1999). Five predictable combinations of protein – DNA complexes are formed with the wt *rafP*, namely CRP- *rafP*, RafR – *rafP*, CRP – RafR – *rafP*, 2xRafR – *rafP* and CRP – 2xRafR - *rafP*. Formation of the last two complexes is less favored, however, since binding of two RafR dimers is mutually obstructive (Muiznieks & Schmitt, 1994). Neither interference, nor co-operativity was observed between CRP and RafR binding both on the wt DNA, as well as on promoter mutants lacking any one of operator sites (Figure 3 in Muiznieks *et al.*, 1999). Obviously, the positioning of CRP site 103° out of the plane respective to operator sites supports mutually independent binding of CRP and RafR, whereas the binding of two RafR dimers at the same face of the double helix is hindered and protein-induced bending is required to accommodate both repressor molecules (Muiznieks & Schmitt, 1994).

3.3. Intrinsic and CRP/RafR-induced bending of the rafP DNA

Sequence-dependent DNA curvature is known to participate in regulation of transcription initiation at various promoters mainly by influencing the formation of transcription-competent nucleoprotein complexes (Pérez-Martin & de Lorenzo, 1997). The affinity of CRP binding also depends on pre-existing curvature and CRP tends to stabilize and enhance DNA bending (Kahn & Crothers, 1992).

The topology of the *rafP* promoter region was assessed by circular permutation assay (Wu & Crothers, 1984) of the 170 bp *Pst*I fragment (Figure 5A). The results are presented in the Figure 4 in Muiznieks *et al.* (1999) and suggest that the center of ca. 50^{o} sequence-dependent DNA curvature is located 40 bp upstream from the center of the CRP site. This structure, however, has no detectable role in regulation of CRP-dependent activity of the *rafP*, since in the plasmids with the deletion of region upstream from CRP site (pRU1304) the production of α-Gal is not affected (Table 1 in Muiznieks *et al.*, 1999).

EMSA of circularly permutated *rafP* fragments complexed with CRP of RafR was employed to determine the extent of the protein-induced DNA bending. The results are shown in the Figure 5 in Muiznieks *et al.* (1999). The protein-induced DNA bending angles were estimated according to Thompson & Landy (1988) and Kim *et al.* (1989).

The angles of the CRP- and RafR-induced bending in the *rafP* DNA are following (Muiznieks *et al.*, 1999):

1. One RafR dimer bends DNA by $95^{o}\pm5^{o}$, which is comparable to the value reported for GalR-induced bent (Zwieb *et al.*, 1989).
2. CRP bends *rafP* DNA by $75^{o}\pm5^{o}$, which is somewhat lower than 96^{o} reported by Kim *et al.* (1989) for CRP – *lacP1* complex and 90^{o} derived from X-ray structure (Schultz *et al.*, 1991). The lower bending angle, nevertheless, is plausible by taking into account the non-canonical left half site of the *raf* CRP site compared to the *lac* counterpart.

3.4. Direction of DNA bending by CRP and RafR

The relative direction of CRP- and RafR-induced DNA bends was determined according to Salvo & Grindley (1987) by changing the spacing between the CRP site and the O_2 (O_1 being non-functional). EMSA of the protein – DNA complexes showed that the mobility of CRP, RafR and *rafP* DNA ternary complexes fluctuated according to the phasing of CRP and O_2 site (Figure 6 in Muiznieks *et al.*, 1999). Placing of the CRP and O_2 sites on the same face of the DNA double helix resulted in smaller mobility shift than placing them on opposite faces. After normalization to exclude the effects of spacer length and nucleotide composition, the mobilities of ternary complexes were plotted against the phasing of the two binding sites. The results imply that the bends imposed by CRP and RafR have opposite directions. Since CRP wraps the promoter DNA around itself, RafR will bend the DNA away from itself, similarly to the other Lac – Gal family repressors (Schumacher *et al.*, 1994; Lewis *et al.*, 1996).

3.5. Interactions of CRP with RNAP

The mechanism of CRP activation of the *rafP* was studied. CRP site at the *rafP* is centered at position –60.5, similar to its position at the *lacP1*. This suggests also the similar activation mechanism by recruitment of RNAP and facilitated formation of closed promoter complex (Ebright, 1993).

Highly purified CRP and *E. coli* RNAP proteins were used do demonstrate the **CRP activation of the *rafP*** in an *in vitro* transcription assay on supercoiled plasmid DNA template. The level of CRP-dependent transcription activation *in vitro* (64-fold) was comparable to 30- to 50-fold increase of α-Gal activity *in vivo* (Figure 7 and Muiznieks *et al.*, 1999). This indicates that no additional cellular components are required for activation to occur.

Further on, the **interactions of the CRP with RNAP** at the *rafP* were studied using EMSA (Figure 8) and DNase I protection assay (Figure 9A).

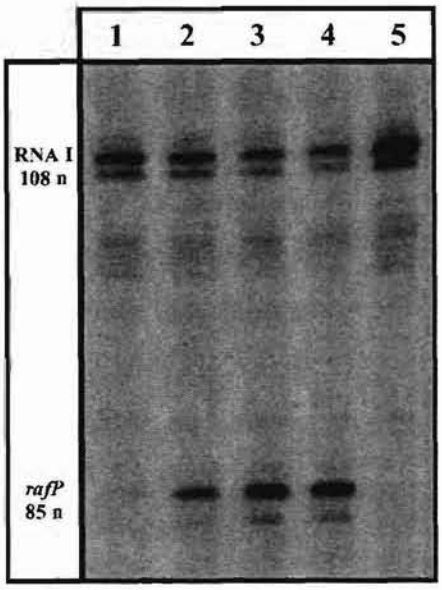


Figure 7. *In vitro* transcription from *rafP* with cAMP titration. DNA concentration is 2.5 nM; RNAP is 16.7 nM; CRP is 20 nM (lanes 1 – 4) and 0 nM (lane 5). Lane 1 – 0.2 μM cAMP; lane – 1.0 μM cAMP; lane – 5.0 μM cAMP; lane 4 – 25.0 μM cAMP; lane 5 – 0 μM cAMP, no CRP. The *rafP*-specific 85 n transcript is marked on the left. For the reference, a 108 n long RNA I transcript was used. RNA I is involved in the repression of plasmid replication and its transcription does not depend upon cAMP-CRP (Tomizawa *et al.*, 1981). The intensity of *rafP*-specific transcripts was normalized against RNA I reference.

Titration of the saturated CRP – *rafP* complexes with RNAP resulted in enhanced polymerase binding to the promoter (Figure 8, lanes 3-6) as compared to titration of naked DNA (Figure 8, lanes 7-10), especially at the lower RNAP concentrations. The titration with RNAP leads also to the appearance of the second, more retarded band, which is much more pronounced in the presence of CRP. This band represents the RNAP – promoter open complex, which has significantly lower mobility due to distorted DNA structure at the transcription start site. If RNAP : DNA molar ratio exceeds 10, the formation of open promoter complex takes place also in absence of CRP (Figure 8, lane 10).

The possibility of non-specific RNAP binding at low RNAP concentrations seems to be excluded, since in the absence of CRP almost no low mobility band can be observed (Figure 8, lanes 3-5 and 7-9). The formation of non-specific RNAP – *rafP* aggregates at high RNAP concentrations can not be ruled out, especially in the absence of CRP (Figure 8, compare lanes 6 and 10), which may indicate that CRP sets the specificity of RNAP binding. The facilitated formation of open promoter complex in the presence of CRP is probably due to the enhancement of the initial binding of RNAP to the *rafP*, although, as will be shown later (Figure 10), the structure of open complex is influenced by CRP as well.

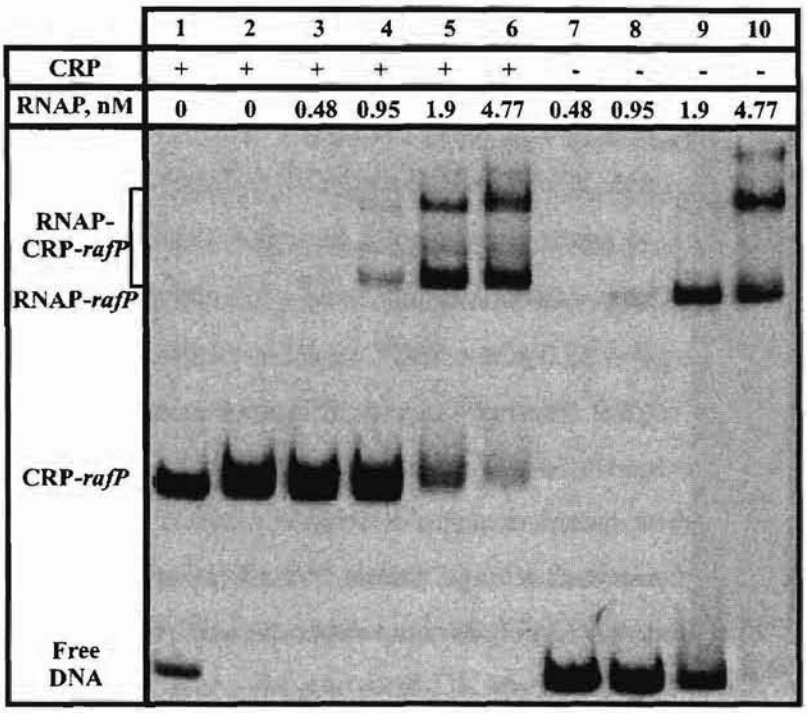


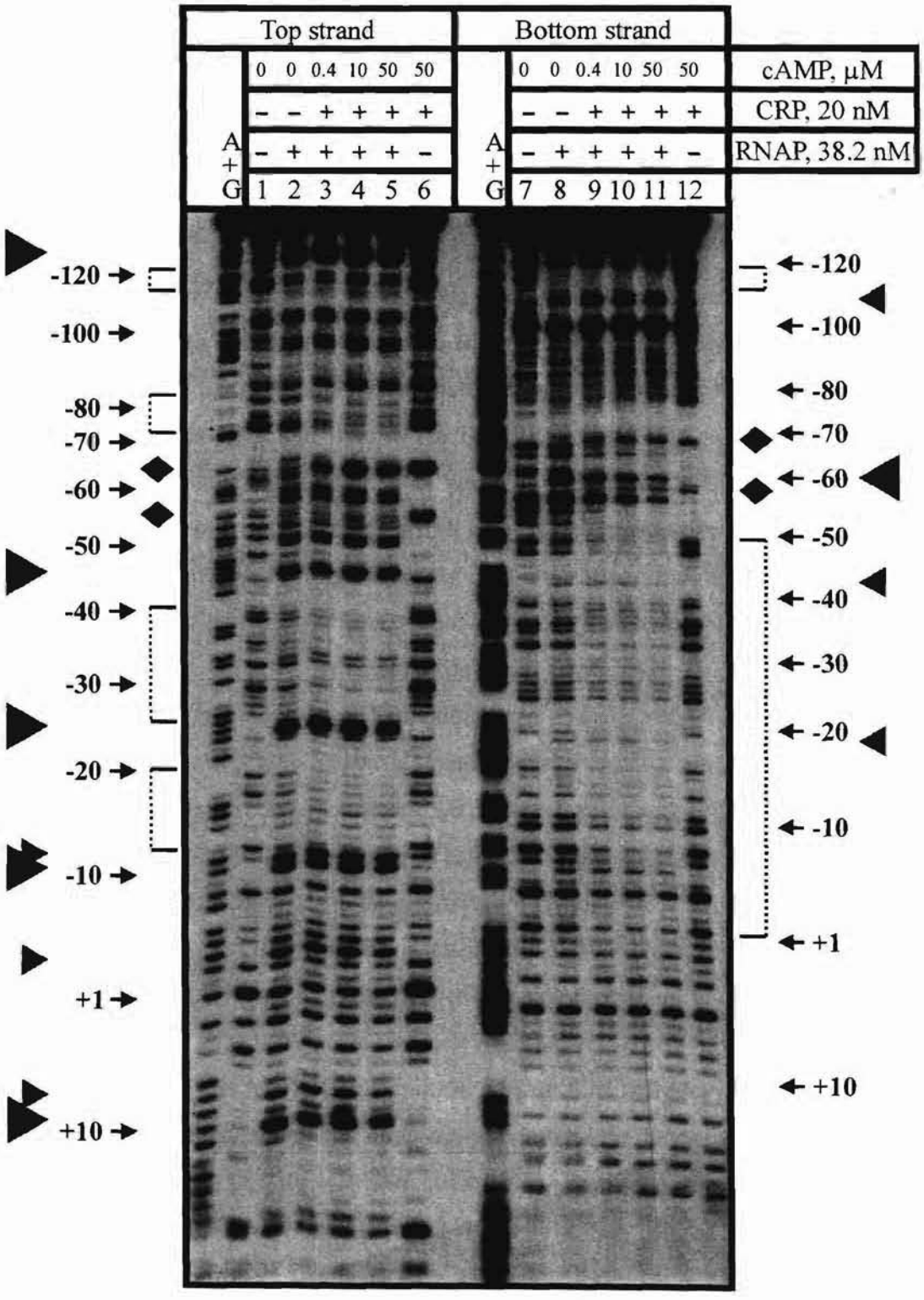
Figure 8. EMSA of RNAP complexes with *rafP* and CRP-*rafP*. DNA concentration is 0.3 nM. RNAP concentration is as indicated above the figure. CRP is 2.2 nM in the lane 1 and 8.8 nM in lanes 2 – 6. Different protein – DNA complexes are indicated on the left.

DNase I probing of the CRP – RNAP – *rafP* ternary complexes was employed to analyze the CRP-dependent modulation of RNAP binding at *rafP* (Figure 9).

The binding of RNAP to *rafP* region in presence of CRP, but without cAMP, seems to be rather weak and there is no clear protection. However, there are several hypersensitive sites induced by RNAP binding, which are not further influenced by CRP. Hypersensitive sites are clustered on the non-template strand from position –10 to +10 and at two positions further upstream, -23 and –44. On the template strand there are only few weak hypersensitive sites at the positions –5, -18, -40 and –58, which become less pronounced, when cAMP is added. If cAMP is present in the reaction, hypersensitive sites at the positions –56 and –66 of template strand and at the positions –54 and –63 of non- template strand can be detected. The protection by CRP of its cognate site is strongly decreased in the presence of RNAP, although the location of the CRP-induced hypersensitive sites is not changed. However, CRP induced hypersensitive sites at the right half site of recognition sequence are less pronounced indicating close interactions between RNAP and CRP. These observations may reflect the differences in DNA bending and nucleoprotein structure formed by individual CRP and RNAP proteins and their complex. Several hypersensitive sites as far upstream as –120 to –130, which exhibit some dependence

on the concentration of cAMP, can be observed in the presence of both RNAP and CRP.

A



B

Figure 9. DNase I probing of RNAP – CRP – *rafP* ternary complexes. A, DNase I footprinting was carried out as described in the "MATERIALS AND METHODS", except that the reactions were supplemented with 1 mM ATP and 1 mM UTP. Structure and extent of RNAP – CRP – *rafP* complexes were examined in the presence of increasing concentrations of cAMP. Both strands of the *rafP* DNA were probed, "Top strand" being non-template and "Bottom strand" being template strand. Concentrations of RNAP, CRP and cAMP are indicated above the figure. DNA concentration is 1.8 nM for top strand and 1.2 nM for bottom strand. The enumeration of the *rafP* sequence in respect to transcription start site for top strand is given in left and for bottom strand - in right margin. The hypersensitive sites introduced in the *rafP* structure by RNAP binding are marked with triangles on both sides of the figure. Their size reflects the relative intensity of the band. Hypersensitive sites introduced by CRP are marked by diamonds. The sites protected by RNAP are bracketed on both sides of the figure. Chemical sequencing reactions specific to purine residues are designated as "A+G". B, Schematic representation of the *rafP* DNA interactions with RNAP and CRP. Sites hypersensitive to DNase I digestion in the presence of RNAP are marked by triangles and those hypersensitive in the presence of CRP – by diamonds. Nucleotides specifically protected by RNAP are marked by filled circles and those protected by CRP are indicated by shaded boxes.

Titration with cAMP facilitates the binding of RNAP to *rafP*, since the clear protection can be observed in several regions both on template and non-template strands. In particular, regions from –16 till –19, from –24 till –28 and from –33 till –38 on non-template strand, and a whole region from about +1 till about –50 on

template strand are protected from DNase I digestion (Figure 9). These results indicate that CRP facilitates the RNAP binding to *rafP* and closed complex formation.

Intriguingly, there is also clear protection around positions –72, -73 and 79 on non-template and around position –70 on template strands, which becomes more pronounced upon increase in cAMP concentration. This may indicate the wrapping of the promoter DNA around RNAP in the presence of CRP. The hypersensitive and protected regions located far upstream between –120 and –130 on both strands (Figure 9) may represent the DNA region, which contacts the back of RNAP.

The formation of the open complex at the *rafP* was investigated using the chemical probing with $KMnO_4$, which preferably reacts with pyrimidine residues, especially timines, in single stranded DNA regions (McCarthy *et al.*, 1990). The RNAP opening of the *rafP* region was investigated in the presence of CRP and increasing concentration of cAMP (Figure 10).

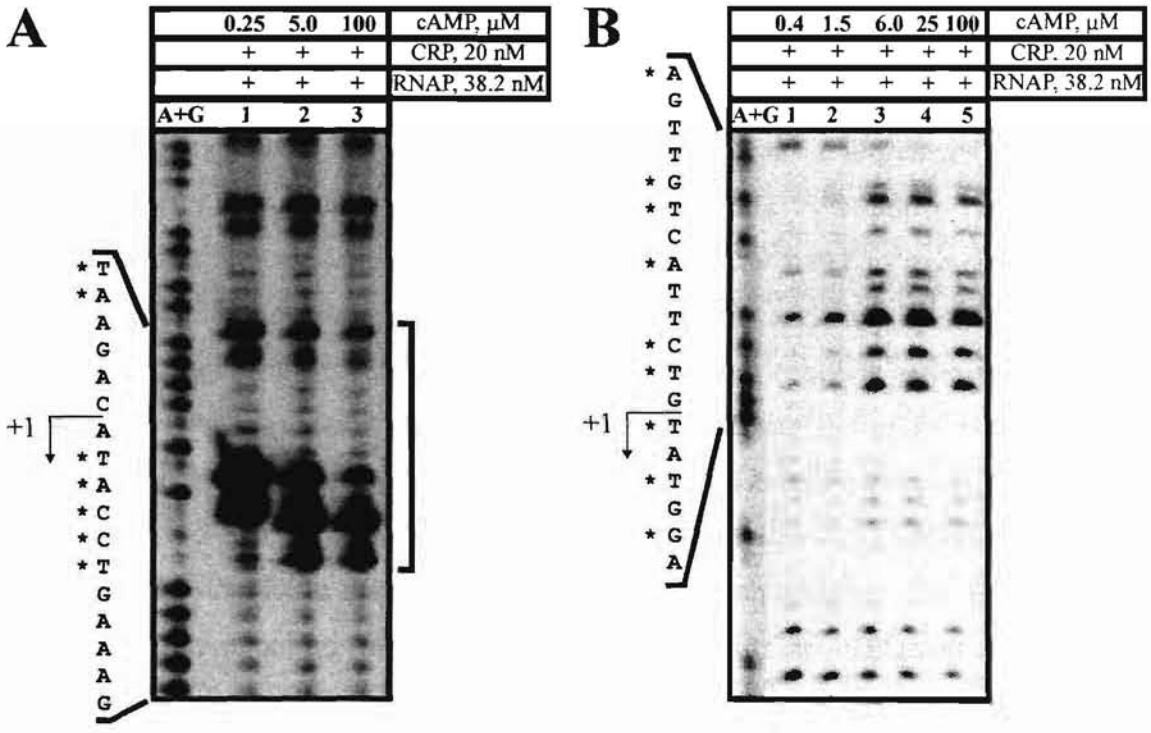


Figure 10. $KMnO_4$ probing of the *rafP* open complex formation on A, top strand and B, bottom strand. RNAP, CRP and cAMP concentrations are indicated above the pictures. DNA concentration is 1.2 nM. DNA sequence of reactive region is shown on left, with hypersensitive nucleotides marked by asterisks.

On the top strand opening of the promoter DNA was detected from the position –6 till +4 already at 0.25 μM concentration of cAMP. Increase in cAMP concentrations leads to DNA melting further downstream, up to the position +6. Meanwhile, the reactivity of the upstream nucleotides decreased significantly.

On the bottom strand week hypersensitive sites were detected at low concentrations of cAMP. Higher than 6 μM cAMP concentrations resulted in major hypersensitive site at +1 position and the additional sites at the positions -9, -8, -6, -3, -2, +3 and +5. $KMnO_4$ hypersensitive site at –13 vanished with increasing cAMP concentrations.

CRP clearly stimulates the RNAP binding at the *rafP* promoter (Figures 8 - 10). Both the $KMnO_4$ hypersensitivity and DNase I protection probing demonstrate that at low cAMP-CRP concentrations RNAP preferably interacts with the top strand of *rafP*. Under those conditions the bottom strand reflects almost no change in DNA structure, with the exception of one weak $KMnO_4$ reactive nucleotide at the position –13 (Figure 10B) and a few weak DNase I hypersensitive sites at positions –5, -18, -40 and –58 (Figure 9B). Increase in cAMP concentration leads to tighter and more extended RNAP interactions with *rafP*. DNA of the bottom strand gets distorted around the transcription start site, while RNAP is tightening the binding to the upstream part of the *rafP*. The interactions between CRP and RNAP may promote also the wrapping of DNA around the polymerase. The changes in DNA topology may permit RNAP to make contacts with the *rafP* at further upstream (-70, -120, Figure 9A) regions and may facilitate the DNA opening at the transcription start site up to position +6 on top strand and up to position +5 on bottom strand (Figure 10). The fact that not only pyrimidines are hypersensitive to $KMnO_4$ modification suggests extremely distorted DNA structure in the open complex.

cAMP concentration-dependent changes in the pattern of $KMnO_4$ reactivity of *rafP* suggest that CRP participates in the formation of open promoter complex. The results shown in Figure 10A imply the downward melting of DNA in the presence of CRP.

CRP apparently stimulates the RNAP binding to the promoter as corroborated by EMSA and DNase I footprinting (Figures 8 and 9). In addition CRP influences the steps beyond the formation of closed complex. The changes introduced in the pattern of $KMnO_4$ reactivity by titration with cAMP suggest involvement of CRP in

modification of open promoter complex. The results presented in the Figure 10A imply the downward melting of DNA in the presence of CRP.

3.6. Interactions of RafR with RNAP

The mechanism of RafR repression was examined by investigating the interactions of RafR with the RNAP.

The transcription assay was carried out to study the *rafP* repression *in vitro* (Figure 11). Purified components of the transcription reaction were sufficient for the RafR to exert its function even in the presence of CRP. The *in vitro* transcription assay also confirmed melibiose as a raffinose operon inducer.

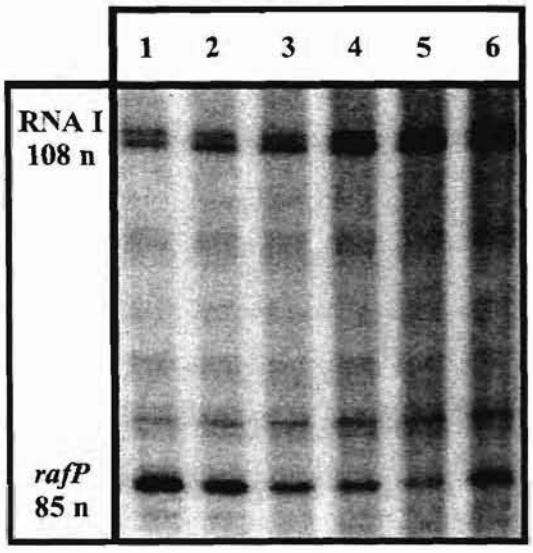


Figure 11. *In vitro* transcription from *rafP* using purified components with RafR titration. RNasin at the concentration of 5 u is added to each reaction and reaction time is reduced to 3 min., in order to minimize the enzymatic degradation of nascent RNA molecules. DNA concentration is 2.5 nM; RNAP is 16.7 nM; CRP is 20 nM. cAMP is 20 μM. The lane 1 contains no RafR, lane 2 – 2.5 ng RafR, lane 3 – 5 ng RafR, lane 4 – 10 ng RafR, lane 5 – 20 ng RafR, lane 6 – 20 ng RafR + 10 mM melibiose. The *rafP*-specific 85 n transcript is marked on the left, along with the 108 n long reference transcript RNA I.

In order to investigate the **mechanism of RafR action at the *rafP***, EMSA of RNAP – *rafP* complexes in presence of RafR and $KMnO_4$ footprinting of RNAP – *rafP* complexes in the presence of RafR were employed (Figures 12 and 13, respectively). Although the RafR binding sites closely overlap the *rafP* –35 and –10 elements, which could lead to the most straightforward mechanism of repression by steric hindrance (Choy & Adhya, 1996), this is not the case with the *rafP*. The fact, that RafR and RNAP protect the same region of *rafP* in DNase I footprinting assay

(Figure 2 in Muiznieks *et al.*, 1999 and Figure 9, this text), does not necessarily imply mutual exclusion of binding. RNAP contacts with the –10 region, as well as with–35 element, which is not completely protected by RafR (Figure 2 in Muiznieks *et al.*, 1999), are feasible. The titration of the saturated 2xRafR – *rafP* complexes with RNAP (Figure 12) lead to accumulation of the lower mobility band, which should contain RNAP bound to the 2xRafR – *rafP* complex. Moreover, the binding of RNAP to the *rafP* was significantly enhanced in the presence of RafR, which may result from protein – protein interactions between RNAP and RafR. Since RafR facilitates the binding of RNAP to the *rafP*, it must repress the transcription initiation step other than formation of closed complex.

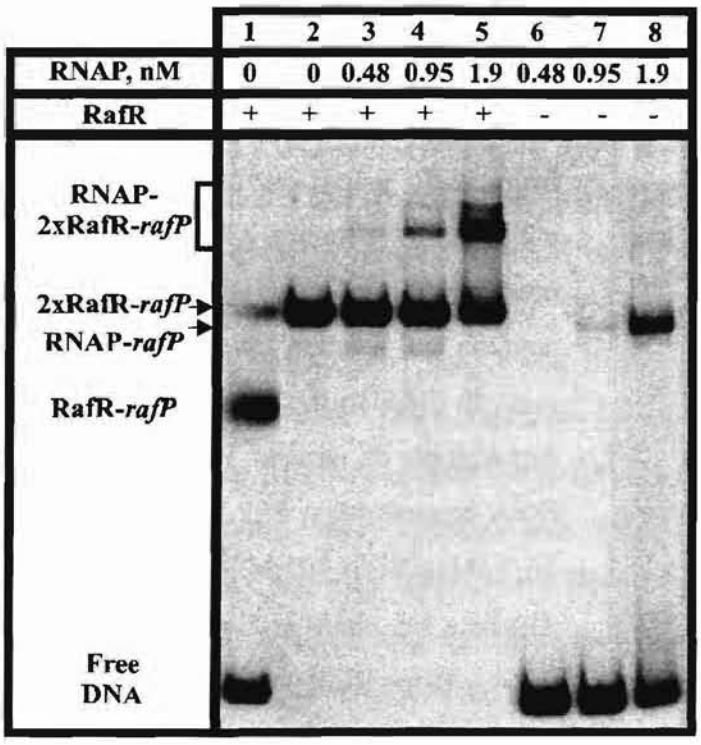


Figure 12. EMSA of RNAP titration of 2xRafR-*rafP* complexes. DNA concentration is 0.3 nM. RNAP concentration is as indicated above the figure. RafR is 0.2 ng (~0.14 nM) in the lane 1 and 1 ng (~0.68 nM) in lanes 2 – 5. Different protein – DNA complexes are indicated on the left.

The effect of RafR on formation of the open promoter complexes was tested by $KMnO_4$ probing technique of the *rafP* non-template strand (Figure 13).

The results presented in the Figure 13 together with the results of EMSA (Figure 12) strongly favor the scheme of the simultaneous binding of RafR and RNAP at *rafP*. Meanwhile, they demonstrate the inhibitory effect of RafR on the open complex formation. RafR repressed the promoter DNA strand separation both in the presence and absence of cAMP-CRP, which is consistent with data of *in vivo* repression in the crp^+ genetic background (Table 2 in Muiznieks *et al.*, 1999) and with

in vitro transcription assay (Figure 11). In the presence of cAMP, the repression was much more pronounced at the low concentration of RafR (Figure 13, compare lanes 2 and 5). At 2 ng concentration (ca. 1.36 nM of RafR and 1.2 nM of *rafP* DNA) RafR almost completely inhibited the melting of the DNA around the transcription start site in the presence of cAMP, whereas the same concentration of RafR had little effect in the absence of cAMP. The differences in the structure of open complex in the presence and absence of cAMP-CRP should be noted (see also Figure 10).

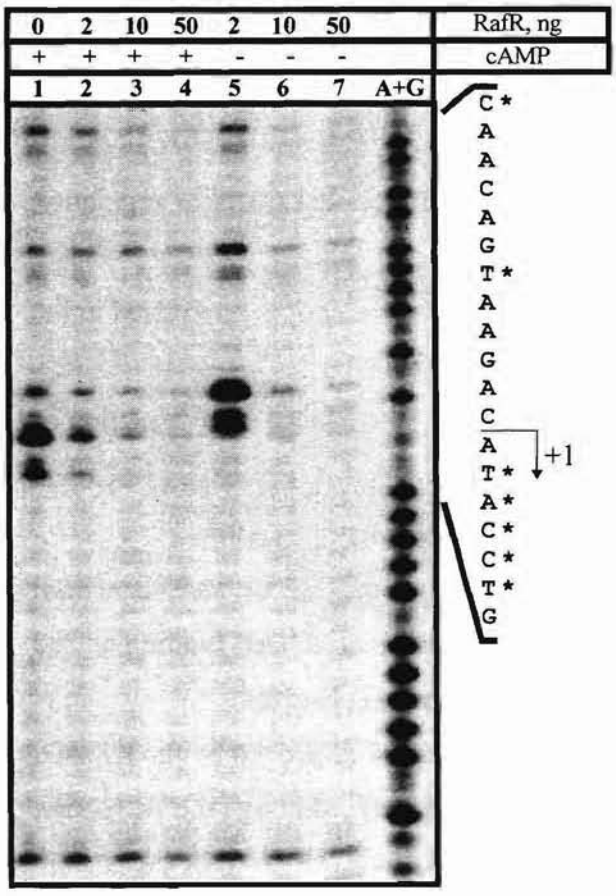


Figure 13. Impact of RafR on the formation of open complex at the *rafP* in the presence and absence of cAMP as probed by $KMnO_4$. RafR concentrations are indicated above the picture. RNAP is 38.2 nM. CRP is included in all reaction at 20 nM. cAMP is 100 μM, if included. DNA is 1.2 nM. DNA sequence of reactive region is shown on right, with hypersensitive nucleotides marked by asterisks.

The model of RNAP interactions with CRP and RafR, as well as the building of different nucleoprotein complexes for repression and activation will be considered in the next chapter.

3.7. Transcription regulation of the rafP

The transcription initiation at the *rafP* of *E. coli* plasmid-borne *raf* operon is both repressed and activated in response to the changes in composition and

concentration of available carbohydrates. If raffinose and melibiose are not present in the growth environment, the *rafP* is repressed by RafR. Gradual depletion of carbohydrates in the environment leads to the rise of intracellular cAMP concentration, which serves as a cofactor for CRP. RafR and CRP may bind *rafP* together, thus facilitating the repression of the *rafP* in the conditions of increasing carbon/energy starvation (Figure 14B). Induction of the *rafP* is mediated by the product of raffinose breakdown, melibiose. The increase of the intracellular melibiose concentration impairs the ability of RafR to bind to the promoter both in the presence and absence of CRP. At low level of cAMP the induction leads only to the basal expression level of the *raf* operon. The induction at high level of cAMP leads to activated level of the expression of the operon to ensure the immediate utilization of the available carbon/energy source.

Our results reveal the following peculiarities in the mechanism of regulation of gene expression by *rafP*.

1. Simultaneous binding of CRP and two RafR dimers at the *rafP* is implicated in the tight repression of transcription (Muiznieks *et al.*, 1999).
2. RNAP binding to the *rafP* is enhanced by CRP, which stimulates the formation of the closed promoter complex and, in addition, influences the formation of open promoter complex by facilitating the DNA wrapping around RNAP.
3. Not only the simultaneous binding of RNAP and RafR to the *rafP* has been suggested by EMSA, but also the clear enhancement of RNAP binding in the presence of promoter-saturating concentration of RafR is observed.
4. RafR represses the formation of open promoter complex as proved by the $KMnO_4$ probing both in the presence and absence of cAMP-CRP, although the repression in the presence of CRP is enhanced.

The schematic localization of the different protein factors, which take part in transcription initiation at *rafP*, is depicted in the Figure 14A.

The *rafP* repression complex, which involves both CRP and two RafR dimers, is shown in Figure 14B. Strong DNA bending by protein factors, which bind to their cognate sites, creates a complex nucleoprotein structure, which ensures 1200-fold repression of *rafP in vivo*. The localization of RNAP in the repression complex is not shown, although the binding of RNAP to the *rafP* saturated with RafR has been proven. Together with $KMnO_4$ probing data, these results strongly indicate on

repression of *rafP* at the stage of formation of open promoter complex. The binding of RNAP together with RafR may contribute to the maintenance of high local concentration of RNAP on the silent promoter, which is important for rapid and efficient induction.

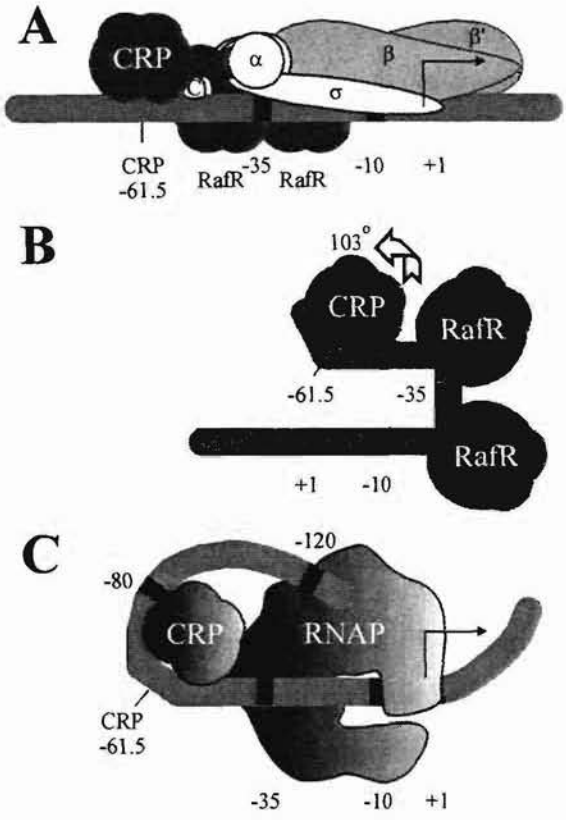


Figure 14. Schematic representation of the different protein complexes formed at the *rafP* and their effect upon promoter DNA structure. A, The approximate localization of the proteins affecting the transcription initiation at the *rafP* on the linear promoter sequence. RafR dimers are shown on the backside of the DNA double helix. CRP is rotated by 103° in respect to RafR because the distance between the CRP and O_1 site is 18 bp. RNAP location on the promoter is approximate and the subunit composition is adapted from Busby & Ebright, 1997. Promoter –35 and –10 elements, the transcription start site and the center of the CRP site are indicated. B, CRP and two RafR dimers bound simultaneously to the *rafP*. Each RafR dimer bends DNA by approximately 90° in the plane of the page. CRP bends DNA by approximately 80° and the bending direction is ca. 103° out of the plane of the page. C, RNAP – CRP – *rafP* activation complex. The shape of RNAP is adapted from Polyakov *et al.*, 1995. CRP bends DNA away from the plane of the page and contacts RNAP. Bending of *rafP* DNA facilitates the wrapping of DNA around RNAP.

Putative protein configuration and DNA structure for *rafP* activation is drawn in the Figure 14C. CRP activation at the *rafP* involves not only the recruitment of RNAP to the promoter, but also the modification of the structure of open promoter complex. This may be achieved by facilitated wrapping of promoter DNA around the RNAP, as suggested by the extension of RNAP footprints both upstream and

downstream in the presence of CRP. Promoter DNA contacts with the back of RNAP are implied by several protected and hypersensitive regions in the upstream region of the *rafP* between positions –130 and –110.

3.8. Role of the promoter upstream element in regulation of rafP

The transcription of *Escherichia coli* raffinose operon in complex natural environments could occur at the basal level, because of the induction of repressor RafR by melibiose and exclusion of activator CRP by low level of cAMP. Under these conditions only the promoter DNA sequence and topology are responsible for the regulation of *rafP*. In order to determine the role of upstream promoter region in regulation of basal *rafP* activity, the mutagenesis and deletion analysis of the corresponding region was carried out (Figure 5).

The plasmids carrying *rafP* derivatives were tested in *rafR*⁻, *crp*⁻ and *cya*⁻ genetic environment in order to assess the basal expression level. The results showing the activity of *rafP* derivatives, as measured by α-Gal accumulation, are compiled in the Table 2 of the Manuscript 3.

The deletion upstream from the position –59 had no effect on promoter activity (pRU1303). Further consecutive deletions caused decrease in *rafP* activity by as much as 9-fold, when sequences upstream from the position –45 were removed. The deletion extending into –35 element (upstream from –28) diminished promoter activity 50-fold. The mutations in *rafP* –59/-39 region, including those, which were introduced to create the restriction enzyme sites (pRU1307E and pRU1307M), had detrimental effect upon *rafP* activity. Promoter proximal mutations and A or T nucleotide exchanges to G or C were more detrimental. Combination of the mutagenesis and deletion analyses suggested more than 9-fold activation of *rafP* by –52/–39 region.

Replacement of wt *rafP* upstream sequence with unrelated oligonucleotide in plasmid pRU1372 resulted in more than 50-fold decrease of α-Gal activity, although the –35 element of the promoter was not impaired.

Mutations introduced in the –35 box of the *rafP* (pRU1307B and pRU1307C) diminished the wt *rafP* activity 50- to 100-fold. This was comparable with the promoter activity, where –35 region is deleted (pRU1301).

The deletion and mutation analysis of the –52/–39 region discloses a distinct sequence element in *rafP*. Within this element individual point mutations display cumulative effect. Thus, the –52/–39 sequence in *rafP* show functional features of the promoter UP elements.

UP element and – 35 sequence at *rafP* act in concert and the promoter activity is decreased 50- to 100-fold if any of these elements is impaired. Both the elements are equally important for the expression of the promoter in the absence of CRP activation since their defects can not be mutually complemented.

The comparison of the wt *rafP* upstream region with the consensus sequence of UP element (Estrem *et al.*, 1998) revealed only vague homology, four nucleotide matches within 15 defined positions (Figure 15).

UP element consensus	59-NNAAAWWTWTTTTNNAAAANNN-38
(Estrem *et al.*, 1998)	
	* * * *
wt *rafP*	59-GATCACACAACCGAAACGTTTT-38
(Aslanidis *et al.*, 1989)	

Figure 15. Comparison of the *rafP* upstream region with the UP element consensus sequence. The sequence matches are marked by asterisks.

In order to assess the role of the DNA sequence and topology in the functional proficiency of UP element in *rafP*, the wt upstream region was replaced with the synthetic oligonucleotides containing $(T)_n(A)_n$ or $(A)_n(T)_n$ tracts (pRU1369 and pRU1371, respectively). Although the number of A and T residues within the UP element region of these plasmids was increased to 15 and 12 respectively, the match with the consensus sequence remained low (Figure 2 in Manuscript 3). None of the constructions displayed significant promoter activity (Table 2 in Manuscript 3). Mutagenesis and insertion of 4 and 5 nucleotides between the cloned oligonucleotide sequence and –35 element of *rafP* generated plasmids pRU1369+4 and pRU1371+5. Here the homology with the UP element consensus sequence was improved (10 matches from 15), but the number of A or T nucleotides within UP element region was reduced (Figure 2 in Manuscript 3). Although both the sequences were identical in 16 positions out of 22, the pRU1369+4 still possessed only marginal activity, but

the activity of pRU1371+5 was increased more than 25-fold in comparison with pRU1371.

These results can be interpreted in context of the promoter topology. According to the theoretical predictions (Ulanovsky & Trifonov, 1987), the oligonucleotide sequence in pRU1369 and 1369+4 does not contain significant sequence-directed DNA curvature, while the oligonucleotide in pRU1371 and pRU1371+5 should be statically bent at 80 – 90°. Electrophoretic mobility analysis in PAAG of 215 bp restriction fragment spanning *rafP* region from these plasmids confirmed the predictions. K_R value for pRU1369 and pRU1369+4 fragments was 1.02 and for pRU1371 and pRU1371+5 – 1.45 (results not shown). The function of UP element in pRU1371+5 is mimicked by intrinsically curved A/T rich DNA fragment. The same sequence was not functionally proficient in pRU1371 due to incorrect phasing of the curved DNA sequence and –35 element of the promoter. Our data indicate that for the recruitment of RNAP spatial structure of UP element is as important as the specific sequence determinants.

UP element may stabilize RNAP binding to –35 element of *rafP* by providing additional contacts with the C-terminal domain of the α subunit. This assumption is supported by the fact that the CRP activation of the promoter, which involves the interactions with the same domain of RNAP α subunit, to large extent abolishes the need for UP element. Mutations at the positions –40 and –45 of *rafP* UP element decreased promoter activity in crp^-, but not in crp^+ strain.

The functional significance of the UP element of the *rafP* should be considered in context with expression of the raffinose operon in the natural environments. If raffinose and/or melibiose are present in the growth medium, but the concentration of glucose is still high, UP element provides cell an opportunity to metabolize additional carbohydrates, which may offer selective advantages.

CONCLUSIONS

1. The expression of *Escherichia coli* raffinose catabolism operon genes is regulated by *raf* repressor (RafR) and cAMP receptor protein (CRP), which interact with RNA polymerase (RNAP) at the *rafP* promoter. Regulation of the transcription depends on the protein – DNA complexes of different composition and topology and not on the mutual exclusion of proteins from the promoter.
2. The *rafP* promoter belongs to the Class I of CRP-dependent promoters. The binding site of CRP is centered at the position –60.5 in respect to the mRNA start site. The 30– to 60-fold activation of *rafP* transcription is achieved by CRP both *in vivo* and *in vitro* in the absence of any additional protein factors.
3. The mechanism of CRP dependent activation of *rafP* transcription involves stimulation of the closed promoter complex formation and wrapping of DNA around RNAP. The influence of CRP on the open promoter complex formation at *rafP* is a novel feature of Class I CRP-dependent promoters.
4. CRP and two RafR dimers bind to *rafP* simultaneously and with DNA bending. The direction of DNA bending in CRP-*rafP* and RafR-*rafP* complexes is opposite in respect to the bound proteins, which postulates that RafR dimer bends DNA away from itself, since CRP wraps the DNA around itself.
5. The binding of RafR to its operator sites, which embrace the –35 promoter element, does not interfere with the RNAP binding to the *rafP* even at the promoter-saturating repressor concentrations. Moreover, the RNAP binding to the *rafP* is clearly favored in the presence of RafR.
6. The *rafP* activity is repressed by RafR at the step of the open promoter complex formation. CRP may play the role of co-repressor in the repression complex, which is formed by RafR and RNAP.
7. The basal expression of the *rafP* depends on the sequence located between –52 and –39 in respect to the mRNA start. This sequence functionally resembles UP element and acts in concert with –35 element. The promoter activity is decreased 50- to 100-fold, if any of these elements is impaired. Both the elements are equally important for the expression of the promoter in the absence of CRP activation since their defects can not be mutually complemented. The functional proficiency of UP element at *rafP* depends both on the nucleotide sequence and DNA topology.

ACKNOWLEDGEMENTS

The financial support from Latvian Council of Science, Councils of Science of Denmark, Fuhrmann Foundation Denmark and European Training Foundation (Tempus) is acknowledged.

I wish to express my sincere gratitude towards Prof. Indriķis Muižnieks for introducing me to the field of molecular microbiology and for his continuous support throughout the time of my studies and work at the Department of Microbiology and Biotechnology.

I am gratefull to late Prof. Valdis Bērziņš for guiding me through the diploma and master studies at the Department of Biochemistry and Molecular Biology.

Many thanks are due to Prof. Rüdiger Schmitt, University of Regensburg, for making my multiple visits in his lab a fruitful and pleasant experience. I look forward to continuing our collaboration.

I am sincerely thankful to Prof. Paul Valentin-Hansen for hosting my visit at the University of Odense and for teaching me how to make a "nice footprint".

The discussions with Dr. H. E. Choy are acknowledged. I am thankful for teaching me the method of and theory behind *in vitro* transcription.

I would like to thank all the friends and colleagues, who have been with me.

Last, but not least, I want to thank my family for their support and understanding.

REFERENCES

Aiyar, S.E., Gourse, R.L. & Ross, W. (1998). Upstream A-tracts increase bacterial promoter activity through interactions with the RNAP α-subunit. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14652-14657.

Aki, T. & Adhya, S. (1997). Repressor induced site-specific binding of HU for transcriptional regulation. *EMBO J.* **16**, 3666-3674.

Aki, T., Choy, H.E. & Adhya, S. (1996). Histone-like protein HU as a specific transcriptional regulator: co- factor role in repression of gal transcription by GAL repressor. *Genes Cells* **1**, 179-188.

Alberti, S., Oehler, S., von Wilcken-Bergmann, B., Kramer, H. & Muller-Hill, B. (1991). Dimer-to-tetramer assembly of Lac repressor involves a leucine heptad repeat. *New Biol.* **3**, 57-62.

Aslanidis, C. & Schmitt, R. (1990). Regulatory elements of the raffinose operon: nucleotide sequences of operator and repressor genes. *J. Bacteriol.* **172**, 2178-2180.

Aslanidis, C., Schmid, K. & Schmitt, R. (1989). Nucleotide sequences and operon structure of plasmid-borne genes mediating uptake and utilization of raffinose in *Escherichia coli*. *J. Bacteriol.* **171**, 6753-6763.

Aslanidis, C., Muiznieks, I. & Schmitt, R. (1990). Successive binding of *raf* repressor to adjacent *raf* operator sites *in vitro*. *Mol. Gen. Genet.* **223**, 297-304.

Attey, A., Belyaeva, T., Savery, N., Hoggett, J., Fujita, N., Ishihama, A. & Busby, S. (1994). Interactions between the cyclic AMP receptor protein and the alpha subunit of RNAP at the *Escherichia coli* galactose operon P1 promoter. *Nucleic Acids Res.* **22**, 4375-4380.

Ayers, D.G., Auble. D.T. & deHaseth, P.L. (1989). Promoter recognition by *Escherichia coli* RNAP. Role of the spacer DNA in functional complex formation. *J. Mol. Biol.* **207**, 749-756.

Barne, K.A., Bown, J.A., Busby, S.J. & Minchin, S.D. (1997). Region 2.5 of the *Escherichia coli* RNAP σ^{70} subunit is responsible for the recognition of the "extended -10" motif at promoters. *EMBO J.* **16**, 4034-4040.

Belyaeva, T.A., Bown, J.A., Fujita. N.. Ishihama, A. & Busby, S.J. (1996). Location of the C-terminal domain of the RNAP alpha subunit in different open complexes at the *Escherichia coli* galactose operon regulatory region. *Nucleic Acids Res.* **24**, 2242-2251.

Bokal, A.J., Ross, W. & Gourse, R.L. (1995). The transcriptional activator protein FIS: DNA interactions and cooperative interactions with RNAP at the *Escherichia coli rrnB P1* promoter. *J. Mol. Biol.* **245**, 197-207.

Bokal, A.J., Ross, W., Gaal, T., Johnson, R.C. & Gourse, R.L. (1997). Molecular anatomy of a transcription activation patch: FIS-RNAP interactions at the *Escherichia coli rrnB P1* promoter. *EMBO J.* **16**, 154-162.

Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. & Buc, H. (1989). Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli*. *EMBO J.* **8**, 4289-4296.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254.

Breul, A., Assmann, H., Golz, R., von Wilcken-Bergmann, B. & Muller-Hill, B. (1993). Mutants with substitutions for Glu171 in the catabolite activator protein (CAP) of *Escherichia coli* activate transcription from the *lac* promoter. *Mol. Gen. Genet.* **238**, 155-160.

Bullock, W.O., Fernandez, J.M. & Short, J.M. (1987). XL1-Blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *Biotechniques* **5**, 376-378.

Burkardt, H.J., Mattes, R., Schmid, K. & Schmitt, R. (1978). Properties of two conjugative plasmids mediating tetracycline resistance, raffinose catabolism and hydrogen sulfide production in *Escherichia coli*. *Mol. Gen. Genet.* **166**, 75-84.

Busby, S. & Ebright, R.H. (1997). Transcription activation at class II CAP-dependent promoters. *Mol. Microbiol.* **23**, 853-859.

Callaci, S., Heyduk, E. & Heyduk, T. (1999). Core RNAP from *E. coli* induces a major change in the domain arrangement of the σ^{70} Subunit. *Molecular Cell* **3**, 229-238.

Choy, H.E. & Adhya, S. (1992). Control of *gal* transcription through DNA looping: inhibition of the initial transcribing complex. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 11264-11268.

Choy, H.E. & Adhya, S. (1993). RNAP idling and clearance in *gal* promoters: use of supercoiled minicircle DNA template made *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 472-476.

Choy, H.E. & Adhya, S. (1996). Negative control. In *Escherichia coli and Salmonella. Cellular and Molecular Biology. 2nd ed.* (Neidhardt, F.C., ed. in chief), pp. 1287-1299, ASM Press, Washington, D.C.

Collado-Vides, J., Magasanik, B. & Gralla, J.D. (1991). Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* **55**, 371-394.

Darst, S.A., Kubalek, E.W. & Kornberg, R.D. (1989). Three-dimensional structure of *Escherichia coli* RNAP holoenzyme determined by electron crystallography. *Nature* **340**, 730-732.

Darst, S.A., Polyakov, A., Richter, C. & Zhang, G. (1998). Insights into *Escherichia coli* RNAP structure from a combination of X-ray and electron crystallography. *J. Struct. Biol.* **124**, 115-122.

deHaseth, P.L. & Helmann, J.D. (1995). Open complex formation by *Escherichia coli* RNAP: the mechanism of polymerase-induced strand separation of double helical DNA. *Mol. Microbiol.* **16**, 817-824.

deHaseth, P.L., Zupancic, M.L. & Record, M.T., Jr. (1998). RNAP-promoter interactions: the comings and goings of RNAP. *J. Bacteriol.* **180**, 3019-3025.

Dixon, W.J., Hayes, J.J., Levin, J.R., Weidner, M.F., Dombroski, B.A. & Tullius, T.D. (1991). Hydroxyl radical footprinting. In *Methods Enzymol.* Vol. 208 (Sauer, R.T., ed.), pp. 380-413, Academic Press, London.

Dombroski, A.J. (1997). Recognition of the -10 promoter sequence by a partial polypeptide of σ^{70} *in vitro*. *J. Biol. Chem.* **272**, 3487-3494.

Dombroski, A.J., Walter, W.A., Record, M.T., Jr., Siegele, D.A. & Gross, C.A. (1992). Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA. *Cell* **70**, 501-512.

Dombroski, A.J., Walter, W.A. & Gross, C.A. (1993). Amino-terminal amino acids modulate σ-factor DNA-binding activity. *Genes Dev.* **7**, 2446-2455.

Dove, S.L., Joung, J.K. & Hochschild, A. (1997). Activation of prokaryotic transcription through arbitrary protein- protein contacts. *Nature* **386**, 627-630.

Ebright, R.H. (1993). Transcription activation at Class I CAP-dependent promoters. *Mol. Microbiol.* **8**, 797-802.

Ebright, R.H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984). Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. *Nature* **311**, 232-235.

Estrem, S.T., Gaal, T., Ross, W. & Gourse, R.L. (1998). Identification of an UP element consensus sequence for bacterial promoters. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9761-9766.

Fried, M. & Crothers, D.M. (1981). Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **11**, 6505-6525.

Gaal, T., Ross, W., Blatter, E.E., Tang, H., Jia, X., Krishnan, V.V., Assa-Munt, N., Ebright, R.H. & Gourse, R.L. (1996). DNA-binding determinants of the α subunit of RNAP: novel DNA-binding domain architecture. *Genes Dev.* **10**, 16-26.

Garges, S., Choy, H.E. & Ryu, S. (1995). Techniques to isolate templates for *in vitro* transcription assays. In *Microbial Gene Techniques* (Adolph, K.W., ed.), pp. 389-399, Academic Press, London.

Gartenberg, M.R. & Crothers, D.M. (1991). Synthetic DNA bending sequences increase the rate of *in vitro* transcription initiation at the *Escherichia coli lac* promoter. *J. Mol. Biol.* **219**, 217-230.

Gaston, K., Bell, A., Kolb, A., Buc, H. & Busby, S. (1990). Stringent spacing requirements for transcription activation by CRP. *Cell* **62**, 733-743.

Ghosaini, L.R., Brown, A.M. & Sturtevant, J.M. (1988). Scanning calorimetric study of the thermal unfolding of catabolite activator protein from *Escherichia coli* in the absence and presence of cyclic mononucleotides. *Biochemistry* **27**, 5257-5261.

Gralla, J.D. & Collado-Vides, J. (1996). Organization and function of transcription regulatory elements. In *Escherichia coli and Salmonella. Cellular and Molecular Biology. 2nd ed.* (Neidhardt, F.C., ed. in chief), pp. 1232-1245, ASM Press, Washington, D.C.

Gralla, J.D., Hsieh, M. & Wong, C. (1993). Permanganate probing. In Footprinting of Nucleic Acid – Protein Complexes (Revzin, A., ed.), pp. 107-128, Academic Press, London.

Gunasekera, A., Ebright, Y.W. & Ebright, R.H. (1992). DNA sequence determinants for binding of the *Escherichia coli* catabolite gene activator protein. *J. Biol. Chem.* **267**, 14713-14720.

Haber, R. & Adhya, S. (1988). Interaction of spatially separated protein-DNA complexes for control of gene expression: operator conversions. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9683-9687.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557-568.

Hawley, D.K., Johnson, A.D. & McClure, W.R. (1985). Functional and physical characterization of transcription initiation complexes in the bacteriophage λ O_R region. *J. Biol. Chem.* **260**, 8618-8626.

Heumann, H., Ricchetti, M. & Werel, W. (1988). DNA-dependent RNAP of Escherichia coli induces bending or an increased flexibility of DNA by specific complex formation. *EMBO J.* **7**, 4379-4381.

Hudson, J.M. & Fried, M.G. (1990). Co-operative interactions between the catabolite gene activator protein and the *lac* repressor at the lactose promoter. *J. Mol. Biol.* **214**, 381-396.

Irani, M.H., Orosz, L. & Adhya, S. (1983). A control element within a structural gene: the *gal* operon of *Escherichia coli*. *Cell* **32**, 783-788.

Ishihama, A. (1981). Subunit assembly of *Escherichia coli* RNAP. *Adv. Biophys.* **14:1-35**, 1-35.

Ishihama, A. (1993). Protein-protein communication within the transcription apparatus. *J. Bacteriol.* **175**, 2483-2489.

Jaenicke, R., Muiznieks, I., Aslanidis, C. & Schmitt, R. (1990). Ultracentrifugal analysis of the quaternary structure of the *raf* repressor from *Escherichia coli*. *FEBS Lett.* **260**, 233-235.

Jeon, Y.H., Yamazaki, T., Otomo, T., Ishihama, A. & Kyogoku, Y. (1997). Flexible linker in the RNAP α subunit facilitates the independent motion of the C-terminal activator contact domain. *J. Mol. Biol.* **267**, 953-962.

Joung, J.K., Koepp, D.M. & Hochschild, A. (1994). Synergistic activation of transcription by bacteriophage lambda cI protein and *E. coli* cAMP receptor protein. *Science* **265**, 1863-1866.

Kahn, J.D. & Crothers, D.M. (1992). Protein-induced bending and DNA cyclization. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 6343-6347.

Kim, J., Zwieb, C., Wu, C. & Adhya, S. (1989). Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene* **85**, 15-23.

Kolb, A., Busby, S., Buc, H., Garges, S. & Adhya, S. (1993). Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **62**, 749-795.

Kubori, T. & Shimamoto, N. (1996). A branched pathway in the early stage of transcription by *Escherichia coli* RNAP. *J. Mol. Biol.* **256**, 449-457.

Kuldell, N. & Hochschild, A. (1994). Amino acid substitutions in the –35 recognition motif of σ^{70} that result in defects in phage λ repressor-stimulated transcription. *J. Bacteriol.* **176**, 2991-2998.

Kumar, A., Malloch, R.A., Fujita, N., Smillie, D.A., Ishihama, A. & Hayward, R.S. (1993). The –35 recognition region of *Escherichia coli* σ^{70} is inessential for initiation of transcription at an "extended –10" promoter. *J. Mol. Biol.* **232**, 406-418.

Landini, P., Bown, J.A., Volkert, M.R. & Busby, S.J. (1998). Ada protein-RNAP σ subunit interaction and α subunit-promoter DNA interaction are necessary at different steps in transcription initiation at the *Escherichia coli ada* and *aidB* promoters. *J. Biol. Chem.* **273**, 13307-13312.

Lee, J. & Goldfarb, A. (1991). *lac* repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter. *Cell* **66**, 793-798.

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G. & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247-1254.

Li, M., Moyle, H. & Susskind, M.M. (1994). Target of the transcriptional activation function of phage lambda cI protein. *Science* **263**, 75-77.

Lonetto, M.A. & Gross, C.A. (1996). Nomenclature of sigma factors from Escherichia coli and Salmonella typhimurium and relationships to sigma factors from other organisms. In *Escherichia coli and Salmonella. Cellular and Molecular Biology. 2nd ed.* (Neidhardt, F.C., ed. in chief), p. 821, ASM Press, Washington, D.C.

Mandal, N., Su, W., Haber, R., Adhya, S. & Echols, H. (1990). DNA looping in cellular repression of transcription of the galactose operon. *Genes Dev.* **4**, 410-418.

Maniatis, T., Fritsch, E.F. & Sambrook, J. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Marr, M.T. & Roberts, J.W. (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science* **276**, 1258-1260.

McCarthy, J.G., Williams, L.D. & Rich, A. (1990). Chemical reactivity of potassium permanganate and diethyl pyrocarbonate with B DNA: specific reactivity with short A-tracts. *Biochemistry* **29**, 6071-6081.

Miller, A., Wood, D., Ebright, R.H. & Rothman-Denes, L.B. (1997). RNAP β' subunit: a target of DNA binding-independent activation. *Science* **275**, 1655-1657.

Mollegaard, N.E., Rasmussen, P.B., Valentin-Hansen, P. & Nielsen, P.E. (1993). Characterization of promoter recognition complexes formed by CRP and CytR for repression and by CRP and RNAP for activation of transcription on the *Escherichia coli deoP2* promoter. *J. Biol. Chem.* **268**, 17471-17477.

Muiznieks, I. & Rostoks, N. (1998). Analysis of Specific Protein-DNA Interactions. In *Basic Cloning Procedures* (Berzins, V., ed.), pp. 68-103, Springer Verlag, Berlin, Heidelberg.

Muiznieks, I. & Schmitt, R. (1994). Role of two operators in regulating the plasmid-borne *raf* operon of *Escherichia coli*. *Mol. Gen. Genet.* **242**, 90-99.

Muiznieks, I., Rostoks, N. & Schmitt, R. (1999). Efficient control of *raf* gene expression by CAP and two Raf repressors that bend DNA in opposite directions. *Biol. Chem.* **380**, 19-29.

Murakami, K., Fujita, N. & Ishihama, A. (1996). Transcription factor recognition surface on the RNAP alpha subunit is involved in contact with the DNA enhancer element. *EMBO J.* **15**, 4358-4367.

Muro, A.F., Pesce, C.G. & Kornblihtt, A.R. (1993). DNA sequencing by the chemical method: a 10 minute procedure for the G+A reaction. *Trends Genet.* **9**, 337-338.

Newlands, J.T., Josaitis, C.A., Ross, W. & Gourse, R.L. (1992). Both FIS-dependent and factor-independent upstream activation of the *rrnB P1* promoter are face of the helix dependent. *Nucleic Acids Res.* **20**, 719-726.

Niu, W., Kim, Y., Tau, G., Heyduk, T. & Ebright, R.H. (1996). Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNAP. *Cell* **87**, 1123-1134.

Oehler, S., Eismann, E.R., Kramer, H. & Muller-Hill, B. (1990). The three operators of the *lac* operon cooperate in repression. *EMBO J.* **9**, 973-979.

Orskov, I. & Orskov, F. (1973). Plasmid-determined H_2S character in *Escherichia coli* and its relation to plasmid-carried raffinose fermentation and tetracycline resistance characters. Examination of 32 H_2S-positive strains isolated during the years 1950 to 1971. *J. Gen. Microbiol.* **77**, 487-499.

Pedersen, H., Sogaard-Andersen, L., Holst, B. & Valentin-Hansen, P. (1991). Heterologous cooperativity in *Escherichia coli*. The CytR repressor both contacts DNA and the cAMP receptor protein when binding to the *deoP2* promoter. *J. Biol. Chem.* **266**, 17804-17808.

Pérez-Martin, J. & de Lorenzo, V. (1997). Clues and consequences of DNA bending in transcription. *Annu. Rev. Microbiol.* **51**, 593-628.

Perros, M. & Steitz, T.A. (1996). DNA looping and Lac repressor – CAP interaction. *Science* **274**, 1929-1930.

Polyakov, A., Severinova, E. & Darst, S.A. (1995). Three-dimensional structure of *E. coli* core RNAP: promoter binding and elongation conformations of the enzyme. *Cell* **83**, 365-373.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. *Nature* **322**, 697-701.

Ptashne, M. & Gann, A. (1997). Transcriptional activation by recruitment. *Nature* **386**, 569-577.

Rao, L., Ross, W., Appleman, J.A., Gaal, T., Leirmo, S., Schlax, P.J., Record, M.T., Jr. & Gourse, R.L. (1994). Factor independent activation of *rrnB P1*. An "extended" promoter with an upstream element that dramatically increases promoter strength. *J. Mol. Biol.* **235**, 1421-1435.

Record, M.T., Jr., Reznikoff, W.S., Craig, M.L., McQuade, K.L. & Schlax, P.J. (1996). Escherichia coli RNAP ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcription initiation. In *Escherichia coli and Salmonella. Cellular and Molecular Biology. 2nd ed.* (Neidhardt, F.C., ed. in chief), pp. 792-820, ASM Press, Washington, D.C.

Rees, W.A., Keller, R.W., Vesenka, J.P., Yang, G. & Bustamante, C. (1993). Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. *Science* **260**, 1646-1649.

Reznikoff, W.S., Winter, R.B. & Hurley, C.K. (1974). The location of the repressor binding sites in the *lac* operon. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 2314-2318.

Rhodius, V.A., West, D.M., Webster, C.L., Busby, S.J.W. & Savery, N.J. (1997). Transcription activation at class II CRP-dependent promoters: the role of different activating regions. *Nucleic Acids Res.* **25**, 326-332.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. & Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNAP. *Science* **262**, 1407-1413.

Ross, W., Aiyar, S.E., Salomon, J. & Gourse, R.L. (1998). *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. *J. Bacteriol.* **180**, 5375-5383.

Sabourin, D. & Beckwith, J. (1975). Deletion of the *Escherichia coli crp* gene. *J. Bacteriol.* **122**, 338-340.

Salvo, J.J. & Grindley, N.D. (1987). Helical phasing between DNA bends and the determination of bend direction. *Nucleic Acids Res.* **15**, 9771-9779.

Savery, N.J., Lloyd, G.S., Kainz, M., Gaal, T., Ross, W., Ebright, R.H., Gourse, R.L. & Busby, S.J. (1998). Transcription activation at Class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNAP alpha subunit. *EMBO J.* **17**, 3439-3447.

Schlax, P.J., Capp, M.W. & Record, M.T., Jr. (1995). Inhibition of transcription initiation by *lac* repressor. *J. Mol. Biol.* **245**, 331-350.

Schmid, M.B. (1990). More than just "histone-like" proteins. *Cell* **63**, 451-453.

Schultz, S.C., Shields, G.C. & Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science* **253**, 1001-1007.

Schumacher, M.A., Choi, K.Y., Zalkin, H. & Brennan, R.G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science* **266**, 763-770.

Scott, S., Busby, S. & Beacham, I. (1995). Transcriptional co-activation at the *ansB* promoters: involvement of the activating regions of CRP and FNR when bound in tandem. *Mol. Microbiol.* **18**, 521-531.

Straney, S.B. & Crothers, D.M. (1987). Lac repressor is a transient gene-activating protein. *Cell* **51**, 699-707.

Stueber, D. & Bujard, H. (1982). Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes. *EMBO J.* **1**, 1399-1404.

Su, T.T. & McClure, W.R. (1994). Selective binding of *Escherichia coli* RNAP to topoisomers of minicircles carrying the TAC16 and TAC17 promoters. *J. Biol. Chem.* **269**, 13511-13521.

Su, T.Z., Qi, S., Yun, W.H. & Xiu, L. (1989). [Regulation in the expression of α-galactosidase gene in *raf* operon in *Escherichia coli*] (in chinese). *Wei. Sheng. Wu. Hsueh. Pao.* **29**, 180-186.

Tao, K., Zou, C., Fujita, N. & Ishihama, A. (1995). Mapping of the OxyR protein contact site in the C-terminal region of RNAP alpha subunit. *J. Bacteriol.* **177**, 6740-6744.

Thompson, J.F. & Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucleic Acids Res.* **16**, 9687-9705.

Tomizawa, J., Itoh, T., Selzer, G. & Som, T. (1981). Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 1421-1425.

Tullius, T.D. & Dombroski, B.A. (1986). Hydroxyl radical "footprinting": high-resolution information about DNA - protein contacts and application to lambda repressor and Cro protein. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 5469-5473.

Ulanovsky, L.E. & Trifonov, E.N. (1987). Estimation of wedge components in curved DNA. *Nature* **326**, 720-722.

Ushida, C. & Aiba, H. (1990). Helical phase dependent action of CRP: effect of the distance between the CRP site and the –35 region on promoter activity. *Nucleic Acids Res.* **18**, 6325-6330.

Valentin-Hansen, P., Sogaard-Andersen, L. & Pedersen, H. (1996). A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Mol. Microbiol.* **20**, 461-466.

Weickert, M.J. & Adhya, S. (1992). A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* **267**, 15869-15874.

Weickert, M.J. & Adhya, S. (1993). The galactose regulon of *Escherichia coli*. *Mol. Microbiol.* **10**, 245-251.

Wu, H.M. & Crothers, D.M. (1984). The locus of sequence-directed and protein-induced DNA bending. *Nature* **308**, 509-513.

Zaychikov, E., Martin, E., Denissova, L., Kozlov, M., Markovtsov, V., Kashlev, M., Heumann, H., Nikiforov, V., Goldfarb, A. & Mustaev, A. (1996). Mapping of catalytic residues in the RNAP active center. *Science* **273**, 107-109.

Zaychikov, E., Denissova, L., Meier, T., Gotte, M. & Heumann, H. (1997). Influence of Mg^{2-} and temperature on formation of the transcription bubble. *J. Biol. Chem.* **272**, 2259-2267.

Zhou, Y., Busby, S. & Ebright, R.H. (1993). Identification of the functional subunit of a dimeric transcription activator protein by use of oriented heterodimers. *Cell* **73**, 375-379.

Zhou, Y., Merkel, T.J. & Ebright, R.H. (1994). Characterization of the activating region of *Escherichia coli* catabolite gene activator protein (CAP). II. Role at Class I and class II CAP-dependent promoters. *J. Mol. Biol.* **243**, 603-610.

Zwieb, C., Kim, J. & Adhya, S. (1989). DNA bending by negative regulatory proteins: Gal and Lac repressors. *Genes Dev.* **3**, 606-611.

LIST OF PUBLICATIONS

1. Muiznieks, I. & Rostoks, N. (1998). Analysis of specific protein-DNA interactions. In *Basic Cloning Procedures* (Berzins, V., ed.), pp. 68-103, Springer Verlag, Berlin, Heidelberg.

2. Muiznieks, I., Rostoks, N. & Schmitt, R. (1999). Efficient control of *raf* gene expression by CAP and two Raf repressors that bend DNA in opposite directions. *Biol. Chem.* **380**, 19-29.

3. Rostoks, N., Muiznieks, I. & Schmitt, R. The third promoter element regulating the basal expression level of plasmid-borne raffinose operon promoter *rafP*. Manuscript 3.

4. Muiznieks, I., Matjushkova, N., Rostoks, N., Makarenkova, G. & Dumpe G. (1995). Long distance regulation through replicon topology in plasmid gene expression. In *Abstracts of the 2nd Genetical Congress of the Baltic States, Proc. Latvian Acad. Sci.* **5/6**, A8.

5. Matjushkova, N. & Rostoks N. (1992). A spontaneous insertion of transposon Tn 1000 in *rafA* gene of plasmid pRU1307. In *Proceedings of the 1st Genetical Congress of the Baltic States, Experimental Biology* **3/4**, 151-153.

1

Analysis of Specific Protein-DNA Interactions

INDRIKIS MUIZNIEKS[*,1] AND NILS ROSTOKS

Introduction

The central issue in the regulation of genome functions is the mechanism of sequence-specific protein-nucleic acid interactions. Gene expression, replication, recombination and DNA condensation in chromatin are steered by binding of regulatory protein ligands to specific sites in DNA. Numerous methods have been developed to study protein-DNA interactions. In this chapter we discuss two widely used and straightforward approaches to address this problem.

Electrophoretic mobility shift assay (EMSA), or gel retardation, or band shift assay characterize the capability of proteins to bind DNA fragments and to form the complexes which are stable in non-denaturing polyacrylamide gels and move slower than free DNA in electrophoresis.

DNA footprinting or DNA protection against the attack of degrading agents by the bound proteins allows the identification of the specific nucleotide sequences which are involved in binding. Additionally, both methods yield information about the structure of the complex and quantitative data about the kinetics of the interactions.

While countless modifications of the methods originally described have been published, appropriate adjustments are needed in every special case. Here, we would like to introduce our system that has worked well for the studies of the regulation of bacterial α-galactosidase (*raf*A) gene promoter P_{rafA}. Within 80 nucleotide base pairs (bp) P_{rafA} carries: (1) the binding site for RNA-polymerase; (2) the recognition sequence for the non-

* Corresponding author: Indrikis Muiznieks, phone: +371-7-322914; fax: +371-7-325657; e-mail: indrikis@acad.latnet.lv
[1] University of Latvia, Faculty of Biology, Kronvalda 4, LV1586 Riga, Latvia

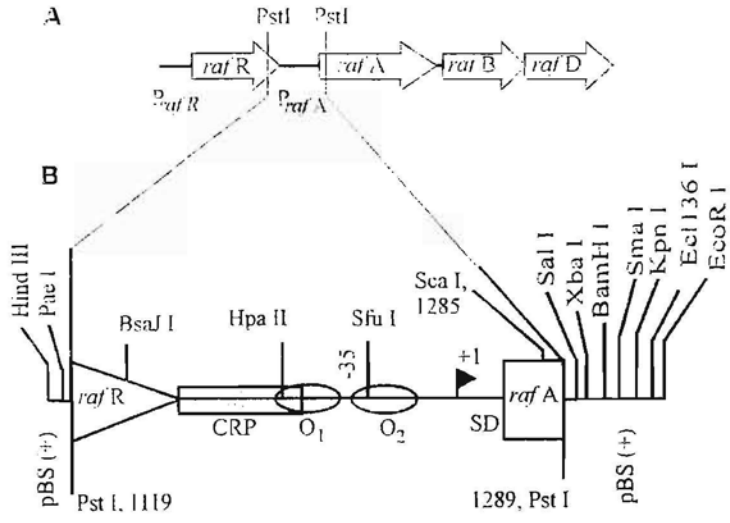


Fig. 5.1A–B. Structure of bacterial plasmid-borne *raf*-operon and P_{rafA} promoter. **A** The *raf*-operon encodes functions required for the inducible uptake and utilization of raffinose in *Escherichia coli*. The expression of three structural genes is negatively controlled by the product of *raf*R gene, the RafR repressor. Homodimers of 36.8 kDa RafR bind to two operator sites O_1 and O_2, that flank the *−35* sequence of the *raf* promoter P_{rafA} (Aslanidis 1989, 1990). The *raf*A gene codes for α-galactosidase, a protein of estimated M_r 81.2 kDa, which is active in tetrameric form. Further members of the *raf*-operon are Raf permease (*raf*B) and sucrose hydrolase (*raf*D). The disaccharide melibiose is a natural inducer of the *raf*-operon. The expression of *raf*-genes is activated by cAMP receptor protein CRP in absence of glucose. The binding sequence for the CRP protein is immediately flanking the O_1 operator. As in the *lac*Z gene promoter, the centre of CRP binding site in P_{rafA} is 61.5 bp 5' to the start point of mRNA synthesis. **B** The 170-bp *Pst*I fragment, base pairs 1119 to 1289 of *raf* operon according to Aslanidis and Schmitt (1990), which carries the P_{rafA} was cloned in polylinker of the phagemid pBS(+) creating plasmid pRU1330. Polylinker sites are indicated with *vertically rotated* names of the restriction enzymes. The polylinker sites were useful in creating differentially labelled DNA ends. RafR binding operator sequences O_1 and O_2 are depicted as *open ellipses*, the site for CRP protein binding is shown as a *shaded box* which partially overlaps with O_1. *Banner* shows the relative position of mRNA start. In the same experiment tandem dimer of the P_{rafA} promoter fragment in *Pst*I site of pBS(+) was obtained. Cutting the tandem dimer with restriction enzymes which are indicated within P_{rafA} sequence generated permutated fragments of identical length, but with different positioning of protein-binding sites within the DNA molecule. The cloning in pBS(+) was intended also to obtain single-stranded DNA of P_{rafA} for site-directed mutagenesis. The Amersham Sculptor in vitro mutagenesis system was used to create the mutations which abrogate binding capacity of either O_1 or O_2 (Muiznieks and Schmitt 1994). The sole *Hpa*II site within P_{rafA} is the product of mutagenesis. In P_{rafA} derivatives which carry *Hpa*II site the operator sequence O_1 has lost the ability to bind RafR, while CRP and O_2 sites are functional

specific activator, cAMP-receptor protein (CRP); (3) two binding sites for the cognate repressor (RafR). The structure of P_{rafA} is shown in Fig. 5.1. Protein binding at P_{rafA} reflects the basic principles of bacterial gene regulation. Meanwhile, the studies of multiple protein interactions within the condensed space of P_{rafA} provide novel insights for the characterization of the role of DNA topology in building protein-DNA complexes.

This chapter gives detailed protocols for making necessary protein and DNA preparations and carrying out the EMSA and DNase I footprinting analysis of a bacterial gene promoter, P_{rafA}.

Electrophoretic Mobility Shift Assay (EMSA)

Principle and Applications

EMSA was developed in the early 1980s (Fried and Crothers 1981; Garner and Revzin 1981) and since then has undergone many modifications that allowed it to become a primary tool in a number of molecular biology applications (for reviews see Carey 1991; Lane 1992; Kerr 1995).

Advantages of this method are its relative simplicity, ability to resolve multiple protein-DNA complexes (Fig. 5.2) and the possibility to work with subpicomolar amounts of material. EMSA may be used for a variety of purposes such as: (1) analysing the ability of a DNA sequence to bind some protein factors; (2) finding an unknown protein factor that binds to a certain DNA sequence or vice versa; (3) studying structural and topological changes in protein or DNA caused by the molecular interactions; (4) exploring the thermodynamic and kinetic parameters of protein-DNA binding.

EMSA is usually performed in polyacrylamide gels (PAAG). Modifications of the method for agarose gels are described, but they are used mostly for observing DNA band shifts with very large protein complexes (Lieberman and Berk 1991). PAAG resolving power sets the limit for the length of DNA fragments used in EMSA. Considering the reduced electrophoretic mobility of protein-DNA complexes, the use of DNA fragments longer than 400–500 bp should be avoided. The lower limit of the DNA length for EMSA is within the oligonucleotide range. It is determined by the size of the specific protein-binding site on DNA, ca. 30 bp.

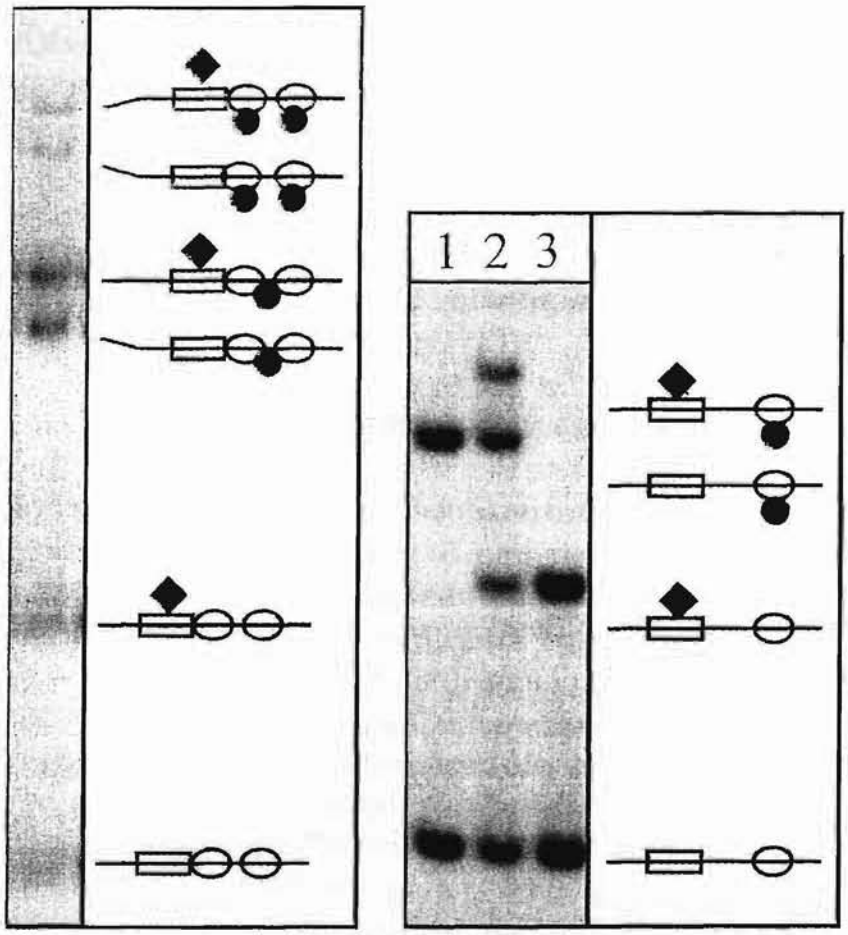


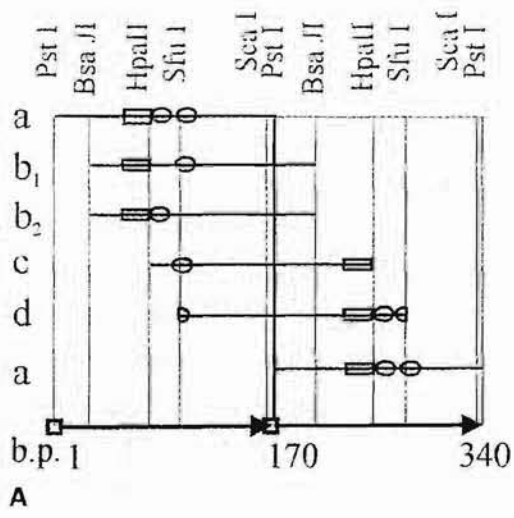
Fig. 5.2A–B. Analysis of protein-DNA complexes by EMSA. **A** The *wt* P_{rafA} may form five different complexes with RafR and CRP. All types of the complexes can be resolved in 20-cm-long 4% PAAG as exemplified by the autoradiogram on the *left* side of the panel. The structure of the complexes is explained on the *right*. RafR is shown as *filled circle* (·), CRP as a *filled diamond* (■). The *line* depicts the 221-bp *Eco*RI/*Hin*dIII fragment from pBS(+) which was isolated and labelled with [α-^{32}P]dATP and Klenow enzyme. *Open ellipses* stand for operator sites, the *box* for the CRP binding sites. If only one repressor dimer is binding to the DNA, no defined affinity for either operator is observed (Muiznieks and Schmitt 1994). Therefore in the pictogram the repressor is depicted in *intermediary position*. The distance between the centres of operator sites is 21 bp, exactly two turns of the helix in B-DNA. Both RafR dimers interact with the DNA from the same side of the molecule. The distance between the centres of CRP and RafR binding sites is 17–18 bp. CRP and RafR are mutually rotated at about 120–150° around the surface of the DNA tube. In the pictograms the proteins are depicted on the *opposite sides* of DNA. **B** EMSA with P_{rafA} where O_1 has been inactivated by point mutation creating *Hpa*II restriction site. *Lane 1* 221-bp *Eco*RI/*Hin*dIII fragment bound only with RafR; *lane 2* with CRP and RafR; *lane 3* with CRP only. Calculations according to Eq. (1) give the value of $\omega \sim 1$, consequently, there is neither interference nor cooperativity in binding of CRP and RafR to the mutated P_{rafA}.The structures of protein-DNA complexes are explained in the *right side* of the panel

Specific applications of EMSA that have been important for our research are: (1) the determination of binding co-operativity of protein factors to an individual DNA fragment (Fig. 5.2); (2) the characterization of protein-induced DNA bending (Fig. 5.3).

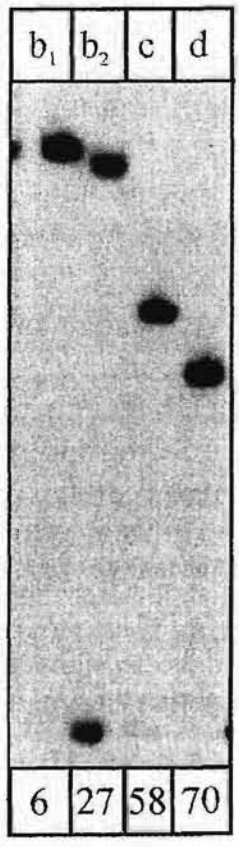
Binding Co-operativity. The transcription of the majority of prokaryotic promoters is either repressed or activated by some protein factors. These protein factors may contact their binding sites independently or they may interact with each other and thus exhibit some hindrance or co-operativity at binding. In the case of hindrance the affinity of binding of each individual protein factor is higher than that of their joint binding. In the case of co-operativity the affinity of joint binding is higher than that for any individual protein.

An example for hindrance may be the binding of the *raf* repressor to two operator sites in the plasmid-borne raffinose operon promoter (Muiznieks and Schmitt 1994). Both operator sites are bound with the same affinity, however, if one of them is already bound by the repressor the other one is bound with ca. 13-times lower affinity.

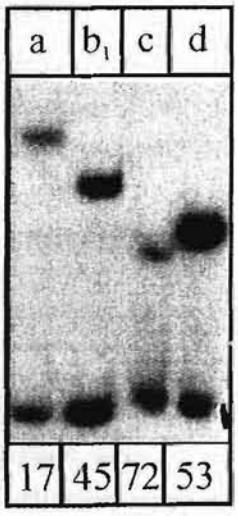
Fig. 5.3A–C. Bending of DNA by the regulatory proteins. **A** Structure of the tandem dimer of the 170-bp *PstI* fragment carrying *wt* and mutant P_{rafA}. The restriction enzymes which cut only once per P_{rafA} *PstI* fragment monomer and were used to obtain permutated sequences are given *above* the figure. The permutated fragments were end-phosphorylated with PNK and [γ-^{32}P]ATP. For labelling, the *PstI* fragment, which has a 5'-recessed end, was substituted by the *ScaI* fragment (blunt end). The distance between *ScaI* and *PstI* sites in the 5'-part of the *rafA* gene is only 4 bp (Fig. 5.1). The localization of protein-binding sites within *ScaI* fragment from the P_{rafA} dimer is nearly the same as in non-permutated sequence of *PstI* fragment of P_{rafA}. The relative localization of the protein-binding sites within the fragments is shown by the same *symbols* as in Fig. 5.2. The types of permutated fragments are denoted with *letters* on the *left margin* of the figure. **B** Bending of P_{rafA} by RafR. EMSA was prepared with 0.2 ng of RafR and permutated P_{rafA} fragments which are denoted *above* the lanes and in the *left margin* of Fig. 5.3A. The distance in bp between the centre of the binding site and the fragment midpoint is given *below* the lanes. The most pronounced mobility shifts are produced with those DNA fragments where the repressor binds near the midpoint of the fragment. **C** Bending of P_{rafA} by CRP. EMSA was prepared with 0.12 ng of affinity-purified CRP and permutated P_{rafA} fragments. The distance in bp between the centre of the binding site and the fragment midpoint is given *below* the lanes. The data show that both RafR and CRP bend P_{rafA} DNA upon binding. Using Eq. (2) (see text), we have estimated that in the P_{rafA} sequence the binding of CRP induces an angle of 85°, but the binding of one RafR repressor dimer an angle of 110° ▶


Pst I
Bsa JI
HpaII
Sfu I
Sca I
Pst I
Bsa JI
HpaII
Sfu I
Sca I
Pst I
a
b1
b2
c
d
a
b.p. 1
170
340

A


b1
b2
c
d
6
27
58
70

B


a
b1
c
d
17
45
72
53

C

The co-operative binding of heterogeneous proteins is exemplified by the binding of CRP and *lac* repressor to their respective primary sites in the *lac* operon promoter (Hudson and Fried 1990; Vossen 1996). However, there are numerous papers on studies of protein binding co-operativity and protein-protein interactions employing the EMSA (see Pedersen 1992, Søgaard-Andersen and Valentin-Hansen 1993, Mao 1994, Kristensen 1996 and references therein). Two approaches have been developed to study binding co-operativity of protein factors: (1) protein distribution analysis (Fried and Crothers 1981); (2) binding competition assay. For detailed calculations and formulae see Hudson and Fried (1990) and Senear and Brenowitz (1991).

Both approaches characterize the relative protein-DNA binding constants through relative probabilities of formation of the corresponding complex. Since the protein-DNA complexes resolved in the native PAA gel contain large numbers of molecules, the probability of formation of each complex may be approximated by the frequency of its occurrence corresponding to the intensity of the band in autoradiography.

By protein distribution analysis, the co-operativity parameter ω_{P1P2} may be calculated according to the equation:

$$\frac{I_0 I_{P1P2}}{I_{P1} I_{P2}} = \omega_{P1P2} \qquad (1)$$

Here, I_0, I_{P1}, I_{P2} and I_{P1P2} are the intensities of the autoradiography bands of free DNA; the Protein 1-DNA complex; the Protein 2-DNA complex and the double complex of both proteins with DNA, respectively. The proteins co-operate at binding DNA if ω_{P1P2} is >1, behave neutrally if ω_{P1P2} is ~ 1 and interfere with each others' binding if ω_{P1P2} is <1.

In binding competition assay, the multiple protein-DNA complexes are incubated with competing, non-specific DNA. To employ this method one of the protein factors must possess lower binding affinity for its site or, alternatively, the binding affinity of the same protein for another site must be different. Because one protein is more weakly bound than another, it is preferably transferred to competing DNA. In case of binding co-operativity this transfer is reduced in the presence of the second protein. To obtain the results one should compare the autoradiography band intensities in two binding assays: (1) with the

both proteins; and (2) with the protein which is transferred to the competitor DNA more easily.

Protein-Induced DNA Bending. Protein-induced DNA bending plays an important role in building the spatial structure of the transcription complex. In EMSA, the protein binding-dependent mobility shift of DNA fragments is further increased by protein-induced DNA bending. Maximal mobility anomaly is observed when the bend is localized in the centre of the fragment and minimal when the bend is at the end of the fragment (Kolb 1983; Wu and Crothers 1984). In the "circular permutation assay" (Fig. 5.3), a tandem repeat of the DNA fragment containing a protein-binding site is cleaved with the restriction enzymes which cut only once per fragment monomer. Thus, a set of DNA fragments of the same length but with different protein-binding site location is obtained. The DNA fragments are complexed with the protein and run on native PAAG. The relative mobilities of the protein-DNA complexes are plotted against the position of the restriction sites in the DNA fragment (5→3'). The apex of the curve indicates the centre of the bend.

The relative bending angles can be calculated according to the equation:

$$\mu_M/\mu_E = \cos \alpha/2, \quad (2)$$

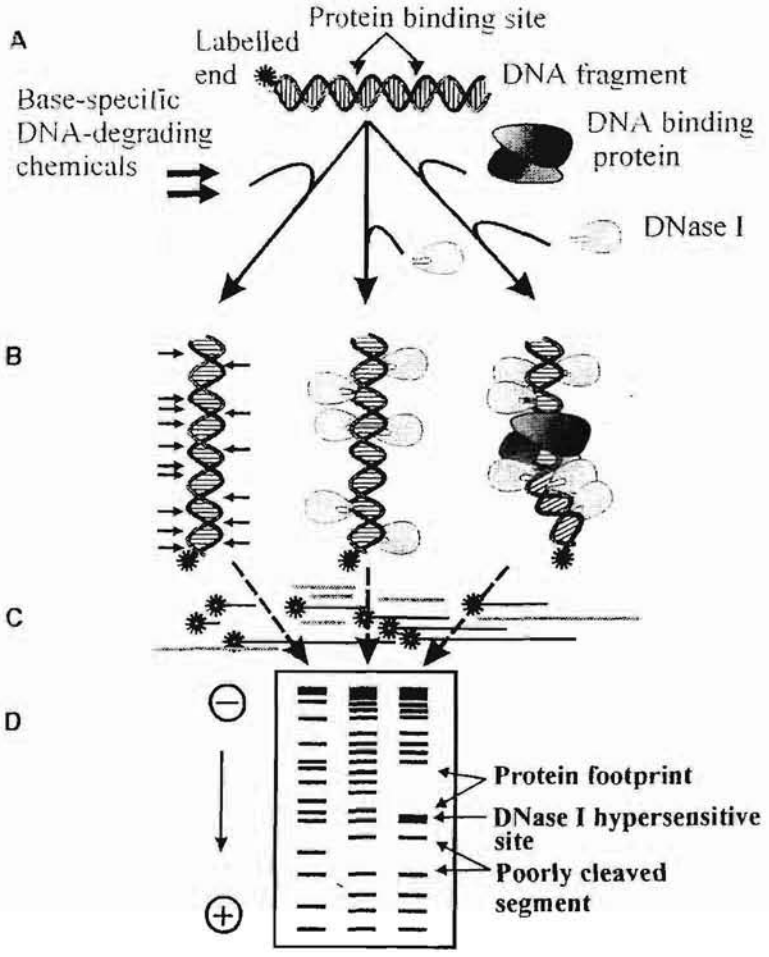
where μ_M is the mobility of the complex with protein bound at the centre of the DNA and μ_E=mobility of the complex with the protein bound at the end (Kim 1989; Thompson and Landy 1988).

Alternatively, DNA bending angles in protein-DNA complexes may be evaluated by comparing their electrophoretic mobility to a set of fragments which carry the standard DNA curvature elements, adenine tract determined bends (Zinkel and Crothers 1990).

DNA Footprinting

EMSA is helpful for characterizing the binding of specific proteins to DNA fragments which carry cognate recognition sites, but it does not provide sequence information about the structure of these sites. The analysis of protein-dependent DNA protection against non-processive degradation, DNA footprinting, is used to identify the sequences which are directly interacting with proteins. The principle of the method is depicted in Fig. 5.4.

Fig. 5.4A–D. Scheme of the footprinting experiment. A The components needed for DNA footprinting: singly end-labeiled DNA fragment, DNA-binding protein, DNase I and chemicals for nucleotide base-specific degradation of DNA. Their preparation is described in Sections 5.1 to 5.3. B The protein-DNA binding reaction and partial degradation of unprotected part of the DNA fragment (Sect. 5.4). The protein-DNA complex is shown in the *right* part of the panel. The protein binding induces DNA bending. In the control reaction, the protein-unprotected fragment is subjected to DNase I degradation as shown in the *middle* part of the panel. To identify the DNA motif which is interacting with the protein, chemical sequencing reactions are carried out in parallel with the same fragment. Stochastic mixture of labelled and unlabelled DNA degradation products is generated. The concentration of DNA-degrading agents which are used in the reaction should produce one or less than one chain break per DNA molecule. Considering the huge number of molecules in the reaction (range of 10^9) this should result in statistically even distribution of degradation events over all the accessible sites for degradation. More than one attack of degrading agents per DNA molecule is depicted in picture just to make clear that numerous fragments are produced in the reaction. C The reaction is stopped, partially degraded DNA fragments are extracted, concentrated, denatured and electrophoresed in sequencing gel (see Sect. 5.4). Only the fragments which are produced from the labelled chain of the DNA will be visualized in the gel. The label from the other chain usually is removed by cutting away terminal 10–20 bases with an appropriate restriction enzyme. The small labelled fragment which is formed in this reaction does not produce interfering bands. The electrophoresis is carried out so that the bands shorter than 25–30 bp leave the gel. D Autoradiography and analysis of the footprint (Sect. 5.5). The analysis of DNA degradation patterns reveal: (1) regions of the protein-specific protection of DNA, the "windows", where the bands are missing due to the presence of DNA binding factor in footprinting reaction; (2) regions of DNA which are poorly cleaved by DNase I also in the absence of the protein due to some local structural features of the fragment, e.g. narrowed minor groove in oligo-T tracts; (3) DNase I hypersensitive sites which may be created by protein-DNA interactions usually as a consequence of protein-induced DNA bending and changing the configuration of the grooves. Poorly cleaved DNA regions are recognized in footprints as empty zones across all the lanes, both in the reactions with DNA binding proteins and in the controls. They may interfere with precise localization of the borders of the specific footprints. Minute structural modifications may influence the DNA sensitivity to DNase I degradation. The methylation of C residue in CG dinucleotide enhances the DNase I susceptibility of the neighbouring 5' phosphodiester bond, although the methyl-group of cytosine is not exposed in the minor groove of the DNA (Kochanek 1993)

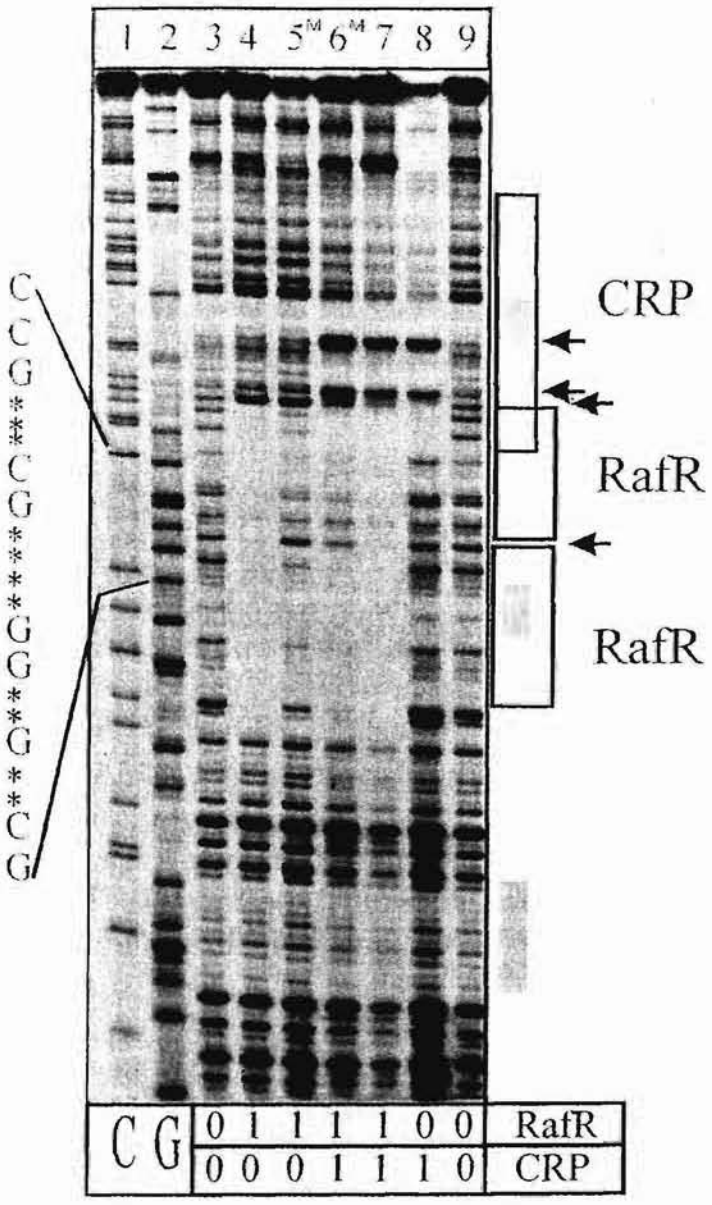


The method was proposed by D.J. Galas and A. Schmitz in 1978. The founder-fathers of DNA footprinting used DNase I as the DNA degrading tool. Since then a plenitude of degrading agents have proven their usefulness in DNA footprinting analysis. They may be classified into three main groups: (1) DNA degrading enzymes; (2) chemical and physical agents which produce free radicals; (3) chemical agents which modify nucleotides and prepare DNA for subsequent cleavage at the places of modification.

DNase I. High specific activity, stability at storage and reproducibility of the results obtained still make DNase I the enzyme of choice for the majority of DNA footprinting experiments (for recent review see Leblanc and Moss 1994). DNase I is an endonuclease which attacks DNA in the minor groove. The protein of 31 kDa molecular mass is active in the monomer form, and its structure is stabilized by Ca^{2+} ions (Lizarraga 1978). The degradation is non-processive. The enzyme cleaves preferen-

tially after pyrimidine bases. Depending upon subtle modulations of DNA structure, e.g. minor alternations in the minor groove width or flexibility, DNase I will cleave some nucleotide motifs more easily or, on the other hand, less easily (Hogan 1989; Kochanek 1993). This can be considered as either a drawback of the method or, vice versa, can be employed to obtain additional information about the quest structures (Fig. 5.5). The molecular dimensions of DNase I are comparable to those of the proteins usually involved in DNA binding. This implies that the boundaries of the protected regions will be drawn with some extension. Some nucleotides immediately next to the bound proteins will remain inaccessible to the DNase I action.

Fig. 5.5. DNase I protection experiments (footprinting) of P_{rafA} with RafR and CRP proteins. The *Eco*RI/*Hin*dIII fragment from pRU1330 was end-phosphorylated with PNK and subsequently digested with *Pae*I. The experiments were performed with 0.5 ng of DNase I per reaction. Approximately one half of the DNA molecules was not cleaved by the nuclease. They build a *thick zone* on *top* of the gel. *Lanes 1* and *2* show chemical sequencing reactions with the P_{rafA} fragment using C- and G-specific modification reactions as denoted *below the lanes* (Maxam and Gilbert 1980). The *left margin* of the panel demonstrates the tracing of specific G and C pattern within the P_{rafA} sequence (Aslanidis and Schmitt 1990) which permits the precise localization of the protein-binding motifs. A number "1" *below* the figure denotes the presence of 10 ng of purified RafR or ca. 10 ng CRP from crude *E. coli* cell extract. "0" denotes the absence of the particular protein. *Lanes 3* and *9* are controls where the DNA was subjected to DNase I attack without protein protection. The regions of intrinsic resistance to DNase I attack are marked with *grey blocks* in the *right margin*. In *lanes 4* and *8* the DNase protection patterns with RafR and CRP, respectively, are shown. CRP produces pronounced bands of DNase I hypersensitivity in the centre of protected DNA segment, while RafR generates only minor bands of enhanced cleavage at the outer borders of the binding sequence. DNase I hypersensitive sites which are generated by the protein binding are marked in the *right margin* of the panel by *arrowheads*. EMSA (Fig. 5.3) has shown that both the proteins bend DNA. The differences in the DNase I hypersensitivity patterns imply that the manner of DNA bending by RafR and CRP is different. *Lane 7* demonstrates that both operators and CRP binding sequence of P_{rafA} can be occupied by the cognate proteins simultaneously. *Narrow white block* on the *right margin* of the panel shows the borders of CRP binding site. Two *broader white blocks* span the binding sequences of RafR. In *lanes 5* and *6* which are indexed by *M* the inducer of the *raf*-operon, melibiose, was added in the binding mixture to final concentration 10 mM. As expected the binding of RafR is weakened in the presence of the inducer (cf. *lanes 4* and *5*). The RafR-CRP-DNA complex is slightly more stable in the presence of melibiose than the RafR-DNA complex alone (*lanes 5* and *6*)



DNase II, Micrococcal Nuclease. DNase II and micrococcal nuclease are further enzymes which are employed for DNA footprinting. Micrococcal nuclease cuts almost exclusively at pA and pT bonds showing greater activity at $(A\text{-}T)_n$ than in homopolymeric runs of A and T (Fox and Waring 1987). These sequences are poorly cut by DNase I. The use of both of the enzymes provides mutually complementary data.

Hydroxyl Radical. The hydroxyl radical, generated by the reduction of hydrogen peroxide by iron(II), was first used to cut the DNA in footprinting assays by D.T. Tullius and B.A. Dombrovski in 1986. Numerous systems for the generation of free radicals for

DNA structure analysis were developed later: methydiumpropyl-EDTA.Fe(II), ortho-phenanthroline.CuI, photoinduction, etc. (Tullius 1991; Macgregor 1992; Baily and Waring 1995). In contrast to the bulky enzymes, the hydroxyl radicals can protrude into the closest boundaries of DNA and protein interactions, and they are less sequence specific than the nucleases, although they demonstrate some DNA secondary structure specificity. Some possibility still exists that unspecific denaturation of protein or protein-DNA complexes may take place during the time when the reactions which generate free radicals are initiated.

Dimethyl Sulphate. Dimethyl sulphate is the most widely used chemical for footprinting analysis among the nucleotide base-modifying reagents. The capability to penetrate through the cellular membranes makes it especially useful for in vivo genomic footprinting applications (Saluz and Jost 1993).

Others. Diethylpyrocarbonate, psoralen, osmium tetroxide and potassium permanganate can be considered as conformation-sensitive probes in DNA-degradation analysis (Runkel and Nordheim 1986; McCarthy and Rich 1991). In the absence of specific binding proteins these agents will preferably attack the regions of highly supercoiled, undertwisted DNA, partially single-stranded, melted DNA regions and four-way junctions. These agents are esspecially useful in the analysis of promoter structures.

A further development of the basic DNA-protection analysis technique is in situ footprinting of protein-DNA complexes following EMSA.

Quantitative analysis of protein-DNA interactions by means of DNA footprinting is a feasible, although seldom used approach (Rehfuss 1990).

Materials

Equipment

- Devices for vertical PAAG electrophoresis, e.g., Minigel-Twin and Maxigel (Biometra)
- Device for horizontal agarose gel electrophoresis GNA100 (Pharmacia)
- UV transilluminator B89196 (Bioblock Sci.)
- Microcentrifuge, e.g., Beckman Microfuge E (Beckman Instruments, Inc.)

- Sorvall centrifuge RC-3B (Du Pont)
- French pressure cell press (American Instrument Co.)
- Ultra Turrax dispersing tool T25 (IKA Labortechnik)
- Beckman scintillation counter LS5800 (Beckman Instruments, Inc.)

- *Escherichia coli* DNA Polymerase I Large Fragment (Klenow Fragment) (Boehringer Mannheim) **Materials**
- DNase I, RNase-free or grade I (Boehringer Mannheim)
- T4 Polynucleotide Kinase (any available commercial vendor, e.g. New England Biolabs, MBI Fermentas)
- restriction enzymes (any available commercial vendor)
- shrimp alkaline phosphatase (Amersham-USB)
- 100 mM solutions of ultrapure dNTPs (Pharmacia)
- [α-^{32}P] dNTPs and [γ-^{32}P]ATP (at 3000 Ci/mM, 10 mCi/ml) (Amersham)

Note. All the chemicals should be highest purity available!

- NucTrap Probe Purification Columns (Stratagene)
- plasmid DNA purification kits (Qiagen)
- X-ray films - Fuji RX
- intensifying screens for autoradiography - DuPoint Cronex Lightning Plus

- XL1-Blue Stratagene **Bacterial Strain**
  Genotype - *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q}*Z*ΔM15 Tn*10* (Tet^{r})]

- 2× TY medium **Bacterial Growth Medium**
  16 g/l bacto-tryptone
  10 g/l yeast extract
  5 g/l NaCl

Note. Sterilize by autoclaving for 20 min at 121 °C

- Acrylamide stock for EMSA gels **Solutions**
  39 % acrylamide
  1 % bis-acrylamide

Note. Store in dark bottle at 4 °C

Caution. Acrylamide is a neurotoxin and is readily absorbed through the skin. Always wear gloves when working with acrylamide and its solutions!

- Acrylamide stock for sequencing gels
  5.7 % acrylamide
  0.3 % bis-acrylamide
  7 M ultrapure urea
  10 % 10× TBE buffer

Note. Store in dark bottle at 4 °C

- Ammonium acetate
  10 M CH_3COONH_4, pH 8.0
- Ammonium persulphate (APS)
  10 % solution

Note. Store in aliquots at −20 °C as the APS is unstable in aqueous solutions. Use the aliquot only once

- Ca^{2+}/Mg^{2+} Solution
  5 mM $CaCl_2$
  10 mM $MgCl_2$
- cAMP
  10 mM Adenosine 3': 5'-Cyclic Monophosphate in TE Buffer

Note. Filter sterilize, store in aliquots at −20 °C

- Competitor DNA
  10 μg/ml of poly(dI):poly(dC) or poly(dA):poly(dT) in TE Buffer.

Note. Store in aliquots at −20 °C

- EDTA
  0.5 M ethylenediaminetetraacetic acid disodium salt dihydrate, pH 8.0

Note. EDTA will not dissolve until the pH 8.0 is set by the NaOH

- Ethanol
  100 % and 70 % (v/v) solutions
- Ethidium bromide
  10^{-4}% in 1× TAE

Note. Store in dark bottle at room temperature (20 °C)

Caution. Ethidium bromide is a powerful mutagen. Always wear gloves while handling gels or solutions containing the dye!

- Loading Solution I
  50 % glycerol
  0.05 % bromphenol blue

0.05 % xylene cyanol
- Loading Solution II
  0.5 M NaOH
  50 % formamide
  0.1 % bromphenol blue
  0.1 % xylene cyanol
- Melibiose
  100 mM solution in ddH_2O

Note. Filter sterilize

- PhenolCIA
  Phenol:chloroform:isoamyl alcohol (25:24:1)
  Store at 4 °C

Note. Phenol is preliminarily equilibrated with 0.5 M Tris-HCl and 0.5 M NaCl!

Caution. Phenol can cause severe burns. Always wear gloves when working with it. Any areas of skin that come into contact with phenol should be washed with soap and water. Do not use ethanol

- SDS
  10 % sodium dodecyl sulphate
- Sodium acetate
  3 M CH_3COONa, pH 4.8
- Stop solution
  0.5 % SDS
  0.13 M NaCl
  30 µg yeast tRNA per ml

Buffers

- Buffer I
  25 mM Tris-HCl, pH 8.0
  50 mM NaCl
  1 mM EDTA
  0.25 mg/ml lysozyme

Note. Prepare buffer without lysozyme and add it to the buffer just before use

- Buffer II
  2 % (v/v) Triton X-100
  40 mM Tris-HCl, pH 6.5
  0.5 M NaCl
  8 mM EDTA

- Buffer III
  100 mM Tris-HCl, pH 6.5
  20 mM EDTA
- Buffer D
  50 mM K phosphate, pH 7.5
  2 mM EDTA
  2 mM dithiothreitol (DTT)
  5 % (v/v) glycerol
- Buffer S
  10 mM Na phosphate, pH 6.8
  0.1 mM EDTA
  0.1 M NaCl
  50 % (v/v) glycerol
- Buffer W
  0.5 M K phosphate pH 7.5
  2 mM EDTA
  2 mM DTT
  5 % (v/v) glycerol
- DNA elution buffer
  0.5 M ammonium acetate
  10 mM magnesium acetate
  1 mM EDTA, pH 8.0
  0.1 % SDS
- DNase I stock and dilution buffer
  50 mM KCl
  50 mM Tris-HCl, pH 8.0
  1 mM DTT
  100 µg/ml bovine serum albumin
  50 % (v/v) glycerol
- Imidazole buffer, 10×
  0.5 M imidazole HCl, pH 6.4
  180 mM $MgCl_2$
  50 mM DTT
- Klenow enzyme buffer, 10×
  0.5 M Tris-HCl, pH 7.6
  0.1 M $MgCl_2$
- Polynucleotide Kinase buffer, 10×
  0.5 M Tris-HCl, pH 7.6
  0.1 M $MgCl_2$
  50 mM DTT
  1 mM spermidine
  1 mM EDTA, pH 8.0
- Protein-DNA binding buffer, 10×

100 mM Tris-HCl, pH8.0
100 mM KCl
10 mM EDTA
10 mM DTT
0.5 mg/ml BSA
30 % (v/v) glycerol
0.1 % Nonidet P-40

Note. If the binding of CRP to the DNA is to be tested, include 2 mM cAMP in the 10× binding buffer

- TAE electrophoresis buffer, 50×
  2 M Tris-acetate, pH 8.0
  0.05 M EDTA
- TBE electrophoresis buffer, 10×
  0.89 M Tris-Borate, pH 8.0
  0.02 M EDTA
- TE buffer
  10 mM Tris-HCl, pH 8.0
  1 mM EDTA
- TE 0.1 buffer
  10 mM Tris-HCl, pH 8.0
  0.1 mM EDTA

5.1 Isolation of DNA Fragments

Procedure

Basic Protocol

A variety of methods for purifying DNA fragments exist either from PAAG or agarose gel. In our hands the best results have been obtained with the modification of the method which is described by Sambrook (1989).

1. Set up the incubation of 5 µg of Qiagen column-purified plasmid DNA which contains P_{rafA} with appropriate restriction enzyme(s) in 50 µl total volume of the reaction mixture.
2. While the plasmid DNA is digested, prepare a 5 % PAAG. In a measuring cylinder mix:
   4.0 ml acrylamide stock for EMSA, 0.64 ml 50× TAE, 0.25 ml 10 % APS, ddH_2O to 32 ml.
   Add 30 µl of TEMED and mix well. Pour a 120×120×1.2 mm gel

with 25-mm-wide slots, pre run 0.5 h with 15 V/cm at room temperature (20 °C).

3. Add 5 µl of Loading Solution I to the digested DNA probes, mix, and load them into the gel slots. Continue the electrophoresis until the bromphenol blue marker has migrated 2/3 of the gel length.
4. Stain the gel with ethidium bromide and visualize the DNA fragments in reflected UV light. Cut out the band of interest.
5. Transfer the gel slice into a 1.5-ml Eppendorf tube and add 400 µl of DNA elution buffer.
6. Incubate the tube overnight at 37 °C, if possible on the rotator.
7. Spin the tube briefly in microcentrifuge to collect the condensation from tube walls and collect the elution buffer carefully trying not to transfer PAAG pieces.
8. Elute the DNA fragment once more with half the volume of elution buffer for a couple of hours. Collect the elution buffer as previously and combine with the first one.
9. Precipitate the DNA with isopropanol (1.0 vol) or 100 % ethanol (2.5 vol) and collect the DNA pellet by centrifugation in micro-centrifuge for 15 min at maximal speed.
10. Discard the supernatant, wash the pellet with 70 % ethanol and centrifuge for 5 min.
11. Dry the pellet at 37 °C and dissolve in 200 µl of ddH_2O. Add 25 µl of 3 M sodium acetate, pH 4.8, precipitate, wash and dry as previously.
12. Dissolve the DNA pellet in water or TE 0.1 buffer.

The procedure should yield ca. 0.2 µg (1.5 pM) of 170–210 bp fragments. Run a 1/10 aliquot of the fragment preparation in 1.5 % agarose gel and determine the DNA concentration by comparing the ethidium bromide fluorescence of the fragment with the fluorescence of the equal length bands containing 10 and 20 ng DNA.

Options Good quality DNA can be recovered also from agarose gels by using electroelution or by centrifugation of agarose gel slice through glass wool.

Numerous reagent kits for purification of DNA fragments from agarose gels are commercially available, e.g., QIAquick Gel Extraction Kit from Qiagen or Sephaglas BandPrep Kit from Pharmacia.

In some cases, for EMSA, but not for footprinting assays, one can omit isolation of the DNA fragment and proceed directly with DNA labelling, if there are no other DNA fragments in the restriction hydrolysate that might be confused with protein-DNA complex(es).

It is also possible to label the DNA fragments first, directly in the restriction mixture and then purify labelled fragments from agarose or PAAG. DNA bands in the gel can be located by autoradiography and purified by any of above-mentioned procedures.

5.2 Preparation of Protein Factors for EMSA

EMSA is based on specific protein-DNA interactions often simulating the in vivo conditions when only several protein molecules (~10 as in *lac* repressor case) specifically bind to a single DNA site per genome. Consequently, not only highly purified proteins, but also crude cell extracts may be used in EMSA. The gene cloning approaches allow one to obtain recombinant bacteria which produce high amounts of the proteins of interest. This allows one to use dilute crude extracts in binding reactions, thereby preventing the interference of endogenous nucleases and proteases. Binding buffers without magnesium ions restrict the activity of most nucleases.

Procedure

RafR

The repressor of the plasmid-borne bacterial raffinose catabolism operon is encoded by the *raf*R gene (Aslanidis and Schmitt 1990) which has been subcloned into high copy number plasmid pUC8 under the control of the *lac*Z promoter. Repressor protein expression is induced upon addition of IPTG and protein is accumulated in the form of inclusion bodies that are easily prepared, purified and reconstituted in the active form (Aslanidis 1990).

1. Transform the *Escherichia coli* strain XL1-Blue with the plasmid pRU984 (Aslanidis 1990).
2. Transfer a single colony into 300 ml of 2× TY medium supplemented with ampicillin (100 μg/ml) and IPTG (1 mM) and grow the culture overnight at 37 °C with shaking.
3. Harvest the cells by centrifugation for 30 min at 5000 rpm in the Sorvall centrifuge, discard the supernatant, resuspend the cells in 5 ml of Buffer I and incubate for 30 min on ice.
4. Disrupt the cells by three passages through the French press. Allow the lysate to cool down on ice between the passages.
5. Sediment the inclusion bodies by centrifugation for 15 min at 10 000 rpm, 4 °C and discard the supernatant.
6. Resuspend the inclusion bodies in 5 ml of Buffer II, homogenize using an Ultra Turrax dispersing tool T25 three times for 20 s and keep the homogenate on ice for 30 min.

Note. An alternative to the Ultra Turrax dispersing tool may be sonication three times for 20 s at 22 kHz, however, very strong foaming is observed due to presence of Triton X-100.

7. Repeat step 5.
8. Resuspend the inclusion bodies in 5 ml of Buffer III, homogenize three times with Ultra Turrax and keep 30 min on ice. Repeat the centrifugation/homogenization step three times.
9. Resuspend the inclusion bodies in 0.5 ml of Buffer I without the lysozyme. Add SDS to the final concentration of 0.1 % and allow the solubilization of the RafR protein to proceed overnight at 4 °C.
10. Pellet the non-soluble inclusion bodies by centrifugation in the micro-centrifuge for 10 min at maximal speed, dispense the supernatant containing the functional repressor in 20 μl aliquots and store at −70 °C.

Protein preparation is analysed using 10 % PAA-SDS gel as described in Sambrook 1989. Average concentration of protein preparation is about 2 μg/μl as judged by SDS-PAGE and the preparation contains only minor contaminants (see Fig. 2B in Aslanidis 1990).

CRP

The gene for CRP is cloned and constitutively overexpressed in a high copy-number plasmid pBG2 (Breul 1993).

Basic Protocol

1. Transform the *E. coli* strain XL1-Blue with the plasmid pBG2.
2. Transfer a single colony into 300 ml of 2× TY medium supplemented with ampicillin (100 µg/ml) and grow the culture overnight at 37 °C with shaking.
3. Harvest the cells by centrifugation for 30 min at 5000 rpm in the Sorvall centrifuge, discard the supernatant, resuspend the cells in Buffer I to give $OD_{600} = 100$, incubate for 30 min on ice, freeze at −70 °C and thaw on ice.
4. Disrupt the cells by three passages through the French press. Allow the lysate to cool down on ice between the passages.
5. Centrifuge the lysate for 15 min at 10 000 rpm, 4 °C and discard the pellet containing cell debris. Aliquot the supernatant and keep at −70 °C.

Chromatography Through the cAMP-Agarose

Option

1. Dialyse 3 ml of the supernatant from the step 5 overnight against two changes of 500 ml of Buffer D. Centrifuge the dialysate at 4 °C, 30 min, 15 000 rpm to sediment the precipitated protein.
2. Make a 5 ml column of cAMP-agarose (Pharmacia), pre-wash it with 5 column volumes of Buffer W +1 M NaCl and equilibrate with 10 vol of buffer D at flow rate 12 ml/h.
3. Load the dialysate to the column at flow rate 3.5 ml/h and wash with 10 column volumes of Buffer W (12 ml/h).
4. Elute CRP with Buffer W +2 mM cAMP at flow rate 2 ml/h. Collect 0.5 ml fractions and check the protein concentration. CRP begins to elute at the end of the first column volume and leaves the column in ½ of its volume.
5. Dialyse the peak fractions of CRP overnight at 4 °C against 100 vol of Buffer S, aliquot and store at −70 °C.

In our hands, the use of more than 1000-fold diluted crude CRP extracts and affinity-purified CRP preparations gave similar results in EMSA and footprinting.

5.3
Labelling of DNA Fragments for EMSA and Footprinting

Although many non-radioactive DNA labelling methods exist today, radioactive labelling is still a widely used, fast and convenient method ensuring the highest quality and sensitivity for various molecular biology applications. The particles emitted by the decaying radioactive isotope penetrate the photographic film, collide with silver halide crystals and generate precipitates of silver atoms. The isotope of choice to prepare radioactively labelled DNA fragments for EMSA is ^{32}P because its β-emission energy is much stronger than that of other often used isotopes (1.709 MeV compared to 0.167 MeV for ^{35}S). Particles emitted by ^{35}S can penetrate the film emulsion only to a depth of 0.25 mm which is not enough when wet gel is covered with Saran Wrap. In contrast, ^{32}P generates β particles which penetrate water or plastic to a depth of 6 mm and pass completely through an X-ray film. This allows us to take autoradiograms of wet gels covered with Saran Wrap as well as make use of intensifying screens that enhance the image ca. fivefold.

Several approaches have been developed to produce labelled DNA fragments with high specific activity but not all of them are suitable for EMSA or footprinting. Random priming method yields DNA probe with specific activity $>10^9$ cpm/μg, however, the label is spread within the fragment and the length of the fragments is not uniform. We recommend two approaches that may be used to generate labelled DNA fragments for EMSA with a specific activity $>10^7$ cpm/μg DNA: (1) labelling 3'-recessed ends with *E. coli* DNA Polymerase I Klenow fragment and appropriate [α-^{32}P]dNTP; and (2) labelling dephosphorylated 5'-ends with T4 polynucleotide kinase and [γ-^{32}P]ATP. The choice of method depends on the ends of the DNA fragment produced by different restriction enzymes. Only recessed 3'-ends are labelled by the Klenow enzyme. T4 polynucleotide kinase may be used to label protruding 5'-ends as well as blunt ends. Recessed 5'-ends are labelled with low efficiency. In this case, the use of imidazole buffer instead of the standard kinase buffer may improve the efficiency of labelling.

Procedure

Labelling with the Klenow Enzyme

The Klenow enzyme adds complementary deoxynucleotides to the hydroxyl groups at the recessed 3' ends of the DNA fragment. In contrast to *E. coli* DNA Polymerase I the Klenow enzyme possesses only 5' to 3' polymerase and 3' to 5' exonuclease activities. If one of the deoxynucleotides in the reaction is substituted by its [α-^{32}P] analogue, the reaction product will be a DNA fragment with one or both ends labelled depending on the restriction enzyme(s) used and on the labelled deoxynucleotide included in the reaction. In our experiments we labelled both ends of P_{rafA} carrying the *Eco*RI/*Hind*III fragment from the plasmid pRU1330 (Fig. 5.1) with [α-^{32}P]dATP.

1. Mix in the Eppendorf tube following reaction components:
   up to 1.5 pM of DNA fragment,
   2.5 µl of 10× Klenow Buffer,
   10 µCi (~3.3 pM) of [α-^{32}P]dATP,
   1 µl of mix of other dNTPs (2 mM each) to fill the ends of DNA fragment,
   2 U of Klenow enzyme (labelling grade),
   ddH_2O to 25 µl.
2. Incubate for 30 min at 37 °C.
3. Increase the reaction volume to 50 µl with ddH_2O and add 1 µl of 5 M NaCl. Extract twice with PhenolCIA.
4. Increase the volume of the aqueous phase to 75 µl and add 25 µl of 10 M ammonium acetate. Precipitate with 250 µl of 100 % ethanol.
5. Spin for 10 min in microcentrifuge, wash the pellet with 70 % ethanol, air-dry.
6. Repeat steps 4 and 5. Control the radioactivity of supernatant with Geiger counter. Two ethanol precipitation/washing steps in presence of 2.5 M ammonium acetate remove more than 95 % of unincorporated label.
7. Perform the Cherenkov counting of the dry DNA pellet. Normally, the specific activity of the sample should be $2-5\times10^7$ cpm/µg.

8. Dissolve the sample in TE 0.1 buffer to make specific activity 1×10^5 cpm/μl.

Note. To avoid the loss of pellet during ethanol precipitation, place tubes in centrifuge with cap hinge at the top and note the position of pellet. Draw off ethanol carefully so as not to disturb the pellet.

Option A frequently employed method to remove the unincorporated label from the reaction mix is gel filtration through Sephadex G-50 columns (Sambrook 1989). Instead of self-made columns it is recommended to use commercially available NucTrap Probe Purification Columns from Stratagene and follow the manufacturer's instructions.

Labelling with T4 Polynucleotide Kinase

T4 polynucleotide kinase (PNK) catalyses the transfer of the γ-phosphate group of ATP to a 5'-OH terminus of the DNA; therefore it is possible to label DNA using [γ-^{32}P]ATP. PNK can catalyze either the forward reaction, namely transfer of a phosphate to a 5'-OH group, or drive the exchange reaction, causing the transfer of the terminal 5'-phosphate group of DNA to ADP and afterwards the rephosphorylation of DNA by transfer of labelled γ-phosphate to DNA. ADP must be in excess amount. Here, we describe only the forward reaction for which the DNA must first be dephosphorylated. To calculate the concentrations of the termini of nucleic acid molecules to be labelled use Table 5.1.

Dephosphorylation There are several enzymes that catalyse cleavage of 5'-phosphate groups from DNA fragments leaving hydroxyl groups necessary for PNK. Those are bacterial alkaline phosphatase (BAP), calf intestinal alkaline phosphatase (CAP) and shrimp alkaline phos-

Table 5.1. Size-concentration relationship of linear double-stranded DNA

Size of double-stranded DNA (in base pairs)	Amount of DNA required to contribute 1 pM of 5' termini (in μg)
50	1.7×10^{-2}
100	3.3×10^{-2}
250	8.4×10^{-2}
500	1.7×10^{-1}

phatase (SAP). All of them carry out the same reaction, however, CAP and SAP have an essential advantage over BAP – they can be completely inactivated by heating, whereas BAP can be inactivated only by multiple phenol/chloroform extractions.

1. Mix the following components in an Eppendorf tube:
   DNA fragment to be dephosphorylated (1.5 pM), 5 µl of 10× SAP buffer (supplied with SAP), 1.0 U of SAP (0.2 U catalyze removal of phosphate from 1 pM of DNA ends), ddH_2O to 50 µl.
2. Incubate for 1 h at 37 °C.
3. Heat the reaction mix for 15 min at 65 °C, then increase the volume till 100 µl and add 2 µl of 5 M NaCl.
4. Extract twice with PhenolCIA.
5. Precipitate the aqueous phase with 2.5 vol of 100 % ethanol for 20 min on ice.
6. Spin in a micro-centrifuge for 10 min at maximal speed, wash the pellet with cold 70 % ethanol and air dry.

Labelling

1. Add to the dephosphorylated dry DNA (up to 1.5 pM of 5' ends; see Table 5.1 to calculate concentration of your DNA fragment) the following components:
   2 µl of 10× PNK buffer, 10 µCi (~3.3 pM) of [γ-^{32}P]ATP, 20–30 U of PNK, ddH_2O to 20 µl.
2. Incubate for 30 min at 37 °C.
3. Increase the volume of reaction mix to 100 µl, add 2 µl 5 M NaCl and extract once with PhenolCIA.
4. Separate the DNA from unincorporated [γ-^{32}P]ATP as described above.
5. Perform the Cherenkov counting of the dry DNA pellet. Normally, the specific activity of the sample should be $>2-5\times10^7$ cpm/µg.
6. Dissolve the sample in TE 0.1 buffer to specific activity 1×10^5 cpm/µl.

Note. Ammonium ions are inhibitors of PNK therefore the DNA fragment after preparation from PAAG according to Sambrook (1989) must be carefully purified. DNA molecules with blunt or

5'-recessed ends are labelled less efficiently than those with 3'-recessed ends therefore it is recommended to increase the amount of PNK to 30–40 units. To obtain the effective phosphorylation of 5'-recessed ends it is also recommended to use imidazole buffer and include in the reaction mixture polyethylene glycol (PEG 8000) in concentrations ranging between 4 and 10 %.

Depending on the number of protein-DNA complexes formed in EMSA, good bands on X-ray film will be produced by 500–1000 cpm of the labelled fragment. To obtain the sufficiently strong signal in autoradiography it is advisable to take for one EMSA reaction the amount of labelled DNA corresponding to 500–1000 cpm, depending on the number of bands expected.

Preparation of the Fragment for Footprinting

Footprinting experiments require a *singly end-labelled* DNA fragment. To obtain this by phosphorylation with PNK, the labelled fragment must be additionally digested with a restriction enzyme which releases a small portion from one end. This procedure with the P_{rafA}-carrying plasmid pRU1330 is facilitated by the flanking sites from pBS(+) polylinker (Fig. 5.1). The fragment was cloned into the *Pst*I site, the *Eco*RI/*Hin*dIII fragment was isolated, labelled with PNK + [γ-^{32}P]ATP and digested with *Ecl*136II or *Pae*I to remove the label from one end of the fragment.

1. The following reaction was assembled in the microcentrifuge tube:
   20 µl of phosphorylated DNA fragment from above, 5 µl of 10× appropriate restriction buffer, 40 U of *Pae*I or *Ecl*136II, ddH_2O to 50 µl
2. Incubate at 37 °C for 1 h. It is important to remove one end of the fragment completely. This is ensured by excess enzyme in the reaction.
3. Extract twice with PhenolCIA.
4. Precipitate and wash the DNA with ethanol. Air dry and count the radioactivity.
5. Dissolve in TE 0.1 to final specific activity 1×10^5 cpm/µl. Store at 4 °C.

For one footprinting reaction ca. 30 000 cpm of labelled DNA are needed.

5.4 Protein-DNA Binding Reactions

Protein-DNA complexes are formed by mixing stoichiometric amounts of the DNA fragment and active protein. If the specific activity of the DNA is in the range of 3×10^7 cpm/μg, as little as 0.1–1.0 ng or several femtomoles of fragment are used in the reaction.

The protein concentration in the reaction depends on the quality of protein preparation and on the number of its binding sites on the DNA fragment. Usually not all protein molecules have retained their binding ability after purification. If the DNA carries more than one binding site, the amount of the cognate protein must be increased accordingly. In most cases, subsaturating protein concentrations when the band of free DNA is still visible in the gel are optimal for the interpretation of EMSA. The right protein amount for binding reactions can be found only empirically by titration of labelled DNA fragments with series of dilutions of the protein preparation.

Efficiency of the formation of protein-DNA complexes is strongly influenced by the composition of binding buffer. In our work addition of DTT, BSA and 0.1 % Nonidet P-40 to the binding reaction favoured protein-DNA interactions. Glycerol in the binding buffer not only had a positive effect on protein binding but also allowed direct loading of the incubations' mix on PAAG without much mixing/pipetting. On the other hand, addition of Mg^{2+} ions in the binding buffer had no positive effect, moreover, it caused degradation of the DNA fragment especially when crude cell lysates with CRP were used. It is not possible to give any general recommendations concerning the buffer composition, mostly because of different binding conditions for different proteins (see Hassanain 1993).

In our hands the protein-binding buffer composition which is given above has worked well both with RafR and CRP, in EMSA and in footprinting assays. However, we would like to note that even for the very well studied CRP-*lac* promoter interactions binding conditions described by different authors may vary. Binding of particular protein factors to their sites demands that

special ligands be added, e.g., specific binding of CRP occurs only in the presence of cAMP. Studies of bacterial repressor-operator interactions may require the addition of specific inducers that cause the dissociation of the complex. cAMP must be included not only in the binding buffer (200 μM) but also in the electrophoresis buffer (20 μM end concentration).

All the nucleic acid binding proteins exhibit some degree of non-specific affinity. This can cause smearing of retarded bands or even appearance of non-specific protein-DNA complexes. This is especially important when crude cell lysates are used. Unspecific binding to the quest DNA may be avoided by increasing the salt concentration in the reaction mixture. A more frequently used approach to overcome this problem is adding of "non-specific" competitor DNA. For this purpose sonicated fish sperm or calf thymus DNA is often employed, although these DNAs may carry sequences which mimick specific binding sites. The use of synthetic competitor DNAs, e.g., poly(dA):poly(dT) or poly(dI):poly(dC), is preferable.

Formation of the protein-DNA complexes is usually carried out at room temperature, however, this may vary with the purpose of the experiment. Incubations at 4 °C or at 37 °C are described also.

The time for complex formation is usually chosen between 5 and 30 min. However, when binding of several consecutively added proteins is investigated reaction times may be increased up to 1 h to allow the binding to reach equilibrium.

All these variations in procedures just point out the necessity of empirical determination of individual reaction conditions for different protein factors and DNA sequences.

We describe here the protocol for the binding reaction which worked well with the proteins involved in the regulation of the *raf*-operon. Similar conditions were applicable also for the binding of crude and affinity-purified human transcription factor AP-2 to the cognate DNA sequences.

Procedure

Binding Reaction in EMSA

1. Mix in the Eppendorf tube:
   3000 cpm or ca. 1 fM of labelled DNA fragment, an appropriate amount of the protein, 1 μl of 10× protein-binding buffer,

cofactors (cAMP, 200 µM; melibiose 10 µM – 1 mM), 1 µg of competitor DNA [poly(dA):poly(dT)]
ddH_2O to 10 µl.

2. Incubate binding reactions for 20 min at room temperature (~20 °C).

3. Load the binding reactions directly on the non-denaturing PAAG without addition of dyes (this is possible due to the presence of 3 % glycerol in the binding reactions) and load 1 µl of Loading Buffer I in the side lanes of the gel to control the migration of the samples in the gel.
In control reactions DNA fragments without proteins or without cofactors are incubated. Additional control is provided by the reaction without competitor DNA. In most reactions with purified RafR and CRP the competitor DNA was omitted since its presence did not influence the binding of the proteins to P_{rafA}.
Typical amounts of the proteins used in our work were about 1 ng of purified RafR and CRP, 1:5000 dilution of the crude extract of CRP over-producing bacterial cells which were disrupted in French Press at $OD_{600}=100$ (Figs. 5.2–5.3).
Specific binding of CRP to DNA is observed only in the presence of cAMP. Melibiose is the natural inducer of the *raf*-operon. This disaccharide abrogates RafR binding and it was used to study the protein-protein interactions at P_{rafA}.

Binding and Footprinting Reaction

1. Mix in the Eppendorf tube:
30 000 cpm or ca. 10 fM of labelled DNA fragment, appropriate amount of the protein, 5 µl of 10× protein-binding buffer, cofactors (cAMP, 200 µM; melibiose 10 µM – 1 mM)
ddH_2O to 50 µl.

2. Incubate binding reactions for 20 min at room temperature (~20 °C).

3. While the incubation is in progress, heat the Stop Solution to 37 °C and mix well.

4. Treat each reaction identically in the following manipulations.

Note. Process no more than three samples simultaneously to achieve similar results. Add 50 µl of Ca^{2+}/Mg^{2+} Solution and

incubate at room temperature for 1 min. Add 3 µl of appropriate DNase I dilution, mix gently, but thoroughly, and incubate at room temperature for 1 min.

5. Terminate the reaction by adding 100 µl of Stop Solution. Mix well.

6. Extract the reaction with 200 µl PhenolCIA.

7. Transfer the upper, aqueous phase to a fresh tube, add 2 µl of 3 M Na-acetate and 500 µl of 100 % ethanol. Precipitate on ice for 20 min.

8. Spin down the DNA in microcentrifuge at maximal speed for 10 min. Carefully remove the supernatant, wash with 70 % ethanol, and air dry.

9. Resuspend the pellet in 4 µl Loading Solution II by vortexing and flicking the tube. Heat at 95 °C for 2 min and chill on ice for at least 2 min.

10. Load onto a 6 % polyacrylamide sequencing gel. Run the gel at 1200–1500 V in 1× TBE buffer until the bromphenol blue is at the bottom of the gel.
The optimal amount of DNase I for the footprinting reaction is to be determined empirically in pilot experiments. The typical DNase I concentrations added per reaction vary between 0.2–1.0 ng. This should allow approximately one random nick per labelled DNA molecule. The dilutions of the grade I DNase I are made in 50 % glycerol-containing buffer from 1 mg/ml stock solution. The dilutions and the stock solution may be stored at least 6 months at −20 °C.
To save the chemicals, the preliminary footprinting reactions can be performed in 1/3 of the described scale with no binding proteins added to the reaction and analysed on 120×200×0.5 mm denaturing PAA gels.
In control reactions DNA fragments are incubated without binding proteins or without cofactors which are needed for binding.
Unlike EMSA, in footprinting assay the DNA should be completely saturated with the binding proteins. For this purpose we used 10 ng of purified RafR and CAP per footprinting reaction (Fig. 5.5). When competitor DNA was added in footprinting with crude CRP extracts, the DNase I concentration per reaction was increased.

To localize the protein-binding sites, the footprinting samples are run in the sequencing gel in parallel with base-specific chemical degradation products of the same DNA fragment (Maxam and Gilbert 1980).

5.5 Electrophoretic Analysis

Procedure

Electrophoresis in EMSA

Dimensions of the PAAG for EMSA depend mostly on the number of expected shifted bands and on the molecular weight of the complexes. If only a couple of low molecular weight bands are formed, it is possible to use short gels (ca. 10 cm). However, if several large protein-DNA complexes are expected and, especially, if they have similar sizes, it is advisable to make longer gels (20 cm or more).

Long gels are also recommended when EMSA is used to assess topological features of DNA which are induced upon binding of protein factor, e.g., protein-induced DNA bending.

The standard PAAG thickness in EMSA is 1 mm. Thinner gels are more easily dried after electrophoresis, but the size of the wells may become too small. Gels which are 2–2.5 mm thick may be used when it is essential to load enough cell lysate, e.g., to detect DNA bound protein by Western blot.

Four to 5% PAAG are most frequently used in EMSA. Such gels have a pore diameter ca. 16–20 nm and provide sufficient frictional force on protein-DNA complexes to resolve them according to their molecular mass and/or structural peculiarities, e.g., bent DNA structures.

PAAG pore diameter depends also on the degree of polymer cross-linking. Instead of the standard 29:1 or 38:2 acrylamide to bis-acrylamide ratio, EMSA gels usually have lower cross-linking at acrylamide to bis-acrylamide ratios of 39:1 or 75:1. In our experiments we used PAA gels with acrylamide to bis ratio 39:1. This is sufficient to separate on a 20 cm long gel all the protein-DNA complexes formed by RafR and CRP with P_{rafA}.

Our experience shows that electrophoresis in 1× TAE buffer gives sharper bands and better resolution than in 0.5× TBE. To avoid buffer exhaustion during prolonged runs, an electrophore-

Table 5.2. Migration of marker dyes in the native PAA gels

Percentage of gel	Bromphenol blue[a]	Xylene cyanol[a]
3.5	100	460
5.0	65	260
8.0	45	160
12.0	20	70
20.0	12	45

[a] The numbers are the approximate sizes of DNA fragments (in nucleotide pairs) with which the dyes would comigrate.

sis chamber with buffer recirculation should be used. Electrophoresis is usually carried out at the same or lower temperature as binding reactions. At least a 0.5 h long pre-run is recommended to guarantee even distribution of ligands in the buffer and gel and to allow stabilization of the current.

Electrophoresis is carried out at ca. 10 V/cm until the samples have migrated appropriate distances. To observe the progression of the electrophoresis, dye markers are added in the lanes next to the binding reactions (see Table 5.2). When the samples have migrated the desired distance in the gel, the current is stopped, the glass plates are disassembled and the gel is either vacuum-dried or covered with Saran Wrap and directly subjected to autoradiography.

Electrophoresis of the Footprints

The footprinting analysis is performed in denaturing sequencing gels and include all the usual steps for processing these gels.

The length of the gel run depends on the DNA fragment size and the localization of the protein-binding site within the fragment. Fragments of 100–600 bp may be used in footprinting, with the protein-binding site not closer than 30 bp from the labelled end. Protein-binding sites as far as 400 bp from the labelled end can be used, but require longer electrophoresis, and the bands are not so sharp.

Good results can be obtained with any type of sequencing gel. To ensure a uniform load of radioactivity in every slot it is recommended to count every probe before dissolving and to adjust the volume of the Loading Solution II according to the amount of cpm in the tube. The optimal width of the slots is 6–8 mm, with gel thickness at the top ~0.2 mm.

Gel Autoradiography

The standard method for the detection of radioactively labelled nucleic acids is autoradiography, although recently phosphor-imaging systems have become available allowing direct scanning of gels and blots without use of X-ray films. Even when employing phosphor-imaging systems, it is advisable to make also an autoradiogram of the gel for your record.

Both wet and dried gels can be subjected to autoradiography. Autoradiography of wet gels covered with Saran Wrap is preferred when further manipulations with the gel are planned, e.g., when localized bands are cut out and radioactivity is counted or when localized protein-DNA complexes are excised and their footprints are made. Dry gels, however, provide higher sensitivity and better quality pictures.

Autoradiography of EMSA gels is usually carried out with two intensifying screens to shorten the exposure times which depend on distribution of labelled DNA in bands and which may vary from several hours to several days.

When quantitative experiments are carried out and densitometry of exposed films is planned it may be reasonable to use pre-flashed X-ray films for autoradiography. Films are pre-exposed to a short (<1 ms) flash of light that activates the silver halide crystals in the emulsion. Crystals in these films have a prolonged linear response to emitted β particles and fluorescent light of the intensifying screens.

Sequencing gels should be dried before autoradiography. With gels which contain more than 20 000 cpm per slot we have not noticed significant differences between autoradiographs done at room temperature or at $-70\,^{\circ}C$.

References

Aslanidis C, Schmid K, Schmitt R (1989) Nucleotide sequence and operon structure of plasmid-borne genes mediating uptake and utilization of raffinose in *Escherichia coli*. J Bacteriol 171:6753–6761

Aslanidis C, Muiznieks I, Schmitt R (1990) Successive binding of *raf* repressor to adjacent *raf* operator sites in vitro. Mol Gen Genet 223:297–304

Aslanidis C, Schmitt R (1990) Regulatory elements of the raffinose operon: nucleotide sequence of operator and repressor genes. J Bacteriol 172:2178–2180

Baily C, Waring MJ (1995) Comparison of different footprinting methodologies for detecting binding sites for small ligand on DNA. J Biomol Struct Dyn 12:869–898

Breul A, Assmann H, Golz R, von Wilcken-Bergmann B, Müller-Hill B (1993) Mutants with substitutions for Glu-171 in the catabolite activation protein (CAP) of *Escherichia coli* activate transcription from the *lac* promoter. Mol Gen Genet 238:155–160

Carey J (1991) Gel retardation. In: Sauer RT (ed) Protein-DNA Interactions. Methods Enzymol. vol 208. Academic Press, London, pp 103–117

Fried M, Crothers DM (1981) Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res 9:6505–6525

Galas DJ, Schmitz A (1978) DNAse footprinting: a simple method for the detection of protein- DNA binding specificity. Nucleic Acids Res 5:3157–3170

Garner MM, Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res 9:3047–3060

Hassanain HH, Dai W, Gupta SL (1993) Enhanced gel mobility shift assay for DNA-binding factors. Anal Biochem 213:162–167

Hogan ME, Robertson MW, Austin RH (1989) DNA flexibility variation may dominate DNase I cleavage. Proc Natl Acad Sci USA 86:9273–9277

Hudson JM, Fried MG (1990) Co-operative interactions between the catabolite gene activator protein and the *lac* repressor at the lactose promoter. J Mol Biol 214:381–396

Kerr LD (1995) Electrophoretic mobility shift assay. Methods Enzymol 254:619–632

Kim J, Zwieb C, Wu C, Adhya S (1989) Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene 85:15–21

Kochanek S, Renz D, Doerfler W (1993) Differences in the accessibility of methylated and unmethylated DNA to DNase I. Nucleic Acids Res 21:5843–5845

Kolb A, Spassky A., Chapon C, Blazy B, Buc H (1983) On the different binding affinities of CRP at the *lac*, *gal* and *mal*T regions. Nucleic Acids Res 11:7833–7852

Kristensen H-H, Valentin-Hansen P, Søgaard-Andersen L (1996) CytR/cAMP-CRP nucleoprotein formation in *E. coli*: the CytR repressor binds its operator as a stable dimer in a ternary complex with cAMP-CRP. J Mol Biol 260:113–119

Lane D, Prentki P, Chandler M (1992) Use of gel retardation to analyze protein-nucleic acid interactions. Microbiol Rev 56:509–528

Lieberman PM,. Berk AJ (1991) The Zta *trans*-activator protein stabilizes TFIID association with promoter DNA by direct protein-protein interaction. Genes Dev 5:2441–2454

Lizzaraga B, Sanchez-Romero D, Gil A, Melgar E (1978) The role of Ca^{++} on pH-induced hydrodynamic changes of bovine pancreatic deoxyribonuclease A. J Biol Chem 253:3191–3195

Mao C, Carlson NG, Little JW (1994) Cooperative DNA-protein interactions. Effects of changing the spacing between adjacent binding sites. J Mol Biol 235:532–544

McCarthy JG, Rich A (1991) Detection of an usual distortion in A tract DNA using $KMnO_4$: effect of temperature and dystamycin on the altered conformation. Nucleic Acids Res 19:5421–5429

Macgregor RB (1992) Photogeneration of hydroxyl radicals for footprinting. Anal Biochem 204:324–327

Maxam AM, Gilbert W (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. In: Grossman L, Moldave K (eds) Nucleic acids. Methods Enzymol, vol 65. Academic Press, London, pp 499–599

Muiznieks I, Schmitt R (1994) Role of two operators in regulating the plasmid borne *raf* operon of *Escherichia coli*. Mol Gen Genet 242:90–99

Pedersen H, Søgaard-Andersen L, Holst B, Gerlach P, Bremer E, Valentin-Hansen P (1992) cAMP-CRP activator complex and the CytR repressor protein bind co-operatively to the *cytRP* promoter in *Escherichia coli* and CytR antagonizes the cAMP-CRP-induced DNA bend. J Mol Biol 227:396–406

Rehfuss R, Goidisman J, Dabrowiak JC (1990) Quantitative footprinting analysis. Binding to a single site. Biochemistry 29:777–781

Runkel L, Nordheim A (1986) Chemical footprinting of the interaction between left-handed Z-DNA and anti-Z-DNA antibodies by diethylpyrocarbonate carbethoxylation. J Mol Biol 189: 487–501

Saluz HP, Jost JP (1993) Approaches to characterize protein-DNA interactions in vivo. Crit Rev Eukaryot Gene Expr 3:1–29

Sambrook J, Fritch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Senear DF, Brenowitz M (1991) Determination of binding constants for cooperative site-specific protein-DNA interactions using gel mobility shift assay. J Biol Chem 266:13661–13671

Søgaard-Andersen L, Valentin-Hansen P (1993) Protein-protein interactions in gene regulation: the cAMP-CRP complex sets the specificity of a second DNA-binding protein, the CytR repressor. Cell 75:557–566

Thompson JF, Landy A (1988) Empirical estimation of protein-induced DNA bending angles: applications to l site-specific recombination complexes. Nucleic Acids Res 20:9687–9705

Tullius TD (1991) DNA footprinting with the hydroxyl radical. Free Radic Res Commun 12–13 Pt 2:521–529

Tullius TD, Dombrovski BA (1986) Hydroxyl radical "footprinting": high resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. Proc Natl Acad Sci USA 83:5469–5473

Vossen KM, Stickle DF, Fried MG (1996) The mechanism of CAP-*lac* repressor binding cooperativity at the *E. coli* lactose promoter. J Mol Biol 255:44–54

Wu H-M, Crothers DM (1984) The locus of sequence-directed and protein-induced DNA bending. Nature 308:509–513

Zinkel S, Crothers DM (1990) Comparative gel electrophoresis measurement of the DNA bend angle induced by the catabolite activator protein. Biopolymers 29:29–38

2

Biol. Chem., Vol. 380, pp. 19–29, January 1999

Efficient Control of *raf* Gene Expression by CAP and Two Raf Repressors that Bend DNA in Opposite Directions

Indrikis Muiznieks[1], Nils Rostoks[1] and Rüdiger Schmitt[2,*]

[1] Faculty of Biology, University of Latvia, Kronvalda Blvd. 4, Riga, LV-1586, Republic of Latvia
[2] Lehrstuhl für Genetik, Universität Regensburg, D-93040 Regensburg, Germany

* Corresponding author

The plasmid-borne *raf* operon of *Escherichia coli* encodes proteins involved in the uptake and utilisation of the trisaccharide raffinose. The operon is subject to dual regulation; to negative control by the binding of RafR repressor to twin operators, O_1 and O_2, and to positive control by the cAMP-binding protein, CAP. We have identified the CAP binding site (CBS) as a 22 bp palindromic sequence with incomplete dyad symmetry by deletion analysis, DNaseI footprinting and electrophoretic mobility shift assays (EMSA) of CAP-DNA complexes. The CBS is centred 60.5 bp upstream of the transcription start point and partially overlaps O_1. *In vivo*, CAP increases *rafA* (α-galactosidase) gene expression up to 50-fold. The 28 bp spacing between the centres of CBS and the – 35 box is essential, since insertions of 4, 8, 12 or 16 bp completely eliminated *rafA* gene expression. *In vitro* binding studies revealed that the CBS, O_1 and O_2 sites, can be simultaneously occupied by their cognate proteins. However, no cooperativity between binding of CAP and RafR was detected. EMSA with circularly permuted DNA fragments demonstrated that CAP and RafR proteins bend *raf* promoter *(rafP)* DNA by 75° ± 5° and 95° ± 5°, respectively, in opposite directions. Among sugar catabolic operons, the compact arrangement of three protein-binding sites, a CBS and two operators bounding the – 35 promoter box, is unique and provides a sensitive and highly efficient device for transcriptional control.

Key words: CAP protein / DNA bending / *rafP* topology / *raf* gene expression / Raf repressor.

Introduction

Utilisation of the trisaccharide raffinose as sole carbon source by certain strains of *Escherichia coli* is mediated by the plasmid-borne *raf* operon (Schmitt *et al.*, 1979; Schmid *et al.*, 1979). The operon encodes an α-galactosidase (α-Gal, *rafA* gene), a permease (*rafB* gene) and an invertase (*rafD* gene) needed for inducible uptake and degradation of raffinose (Aslanidis *et al.*, 1989). The recently identified porin-encoding gene, *rafY*, is not a part of the *raf* operon (Ulmke *et al.*, 1997). The *raf* genes are co-transcribed from a common promoter upon induction by endogenous melibiose. The promoter is controlled by the RafR repressor which is encoded by the upstream *rafR* gene (Aslanidis and Schmitt, 1990; Aslanidis *et al.*, 1990). RafR binds to the 18 bp palindromic operator sites, O_1 and O_2, separated by 3 bp of the – 35 box (Figure 1). Binding of RafR to O_1 and O_2 reduced *in vivo* transcription of *rafA* by a factor of 1200, whereas RafR binding to O_2 alone resulted in a 540-fold (45%) and to O_1 alone a 72-fold (6%) reduction in *rafA* expression. This difference has been ascribed to the positioning of the operator sites relative to the promoter elements, hence, the dominance of O_2, which is located between the – 35 and – 10 promoter boxes (Lanzer and Bujard, 1988; Muiznieks and Schmitt, 1994). The binding of the second RafR dimer to O_1, however, is crucial for complete repression.

An earlier study (Su *et al.*, 1989) of α-Gal activity in *E. coli* mutants deficient in CAP *(crp)* or adenylate cyclase *(cya)* suggested that the *raf* operon is subject to positive transcription control by the catabolite gene activator protein, CAP (also termed CRP, for cyclic AMP receptor protein). The cAMP-CAP complex (hereafter referred to as CAP) activates transcription of ca. 30 operons in *E. coli* by binding to a cognate regulatory site upstream from the – 35 and – 10 promoter boxes (Collado-Vides *et al.*, 1991; Kolb *et al.*, 1993). CAP binds to DNA by a helix-turn-helix motif in the carboxy-terminal domain of the protein (Schultz *et al.*, 1991; Parkinson *et al.*, 1996). The nine amino acid recognition helix binds to one half of the palindromic CAP binding sequence. CBS (de Crombrugghe *et al.*, 1984; Gunasekera *et al.*, 1992). Figure 1 depicts the consensus CBS and indicates the positions essential for CAP recognition in bold type. The center of CBS with respect to the transcription start site varies from – 40 to – 200 in different promoters. For three promoter types, where CBS is centered at – 41.5 *(gal)*, – 61.5 *(lac)* and – 70.5 *(malT)*, CAP alone is sufficient to stimulate transcription initiation by RNA polymerase. Gel shift experiments using a set of circularly permuted DNA fragments have demonstrated that CAP binding induces bending of DNA by approximately 90° (Gronenborn *et al.*, 1984; Gartenberg and Crothers, 1988; Kim *et al.*, 1989; Zinkel and Crothers, 1990). Crystal structure analysis of the CAP-DNA complex indicated that the curvature results mainly from two kinks in the consensus sequence, when DNA is wrapped around the protein (Schultz *et al.*, 1991; Parkinson *et al.*, 1996). Transcription

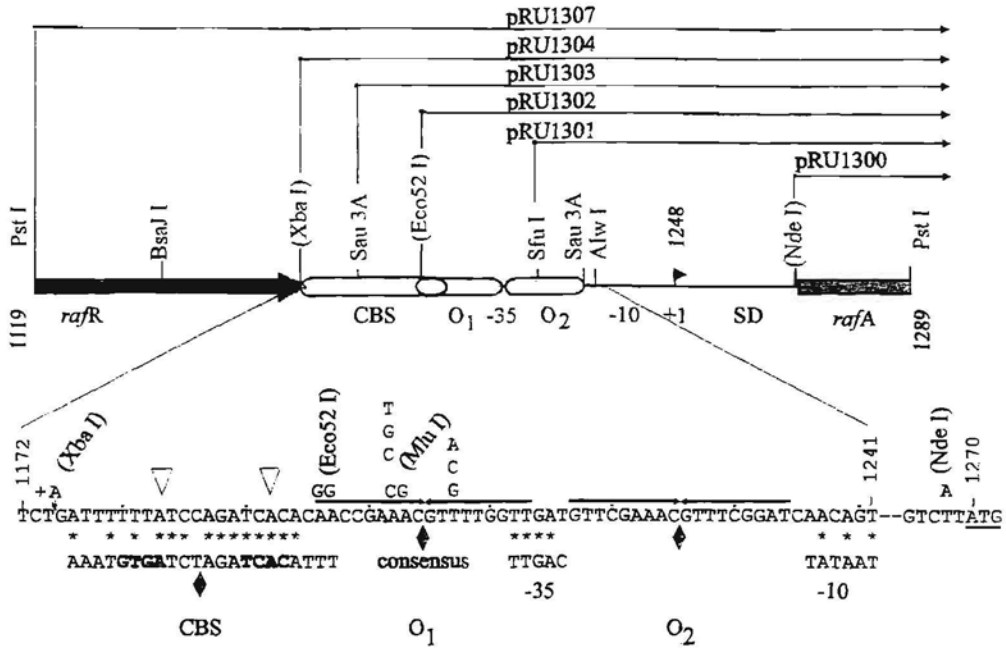


Fig. 1 Organisation of Sequence of *rafP* DNA.
Physical and genetic map of a 170 bp *Pst*I fragment containing the *raf* regulatory region *(rafP)* intervening between *rafR* (repressor gene) and *rafA* (α-gal gene); CBS, catabolite activator protein binding site; O_1, O_2 *raf* operators; −35; −10, *rafP* promoter boxes; + 1, transcription start point; SD, Shine-Dalgarno sequence. A set of nested deletions (pRU1300–pRU1304) and the native plasmid, pRU1307 (Table 2) extending from the left into *rafP* is drawn above the map. Relevant restriction sites are shown, those introduced by mutagenesis are in parentheses. The nucleotide sequence encompassing the CAP, O_1 and O_2 sites is shown below, on an expanded scale. Nucleotide numbering is according to Aslanidis and Schmitt (1990). Inverted arrows indicate dyad symmetry of *raf* repressor binding sequences, filled diamonds show the symmetry axes of protein binding sites and solid arrowheads mark sites hypersensitive to DNaseI upon CAP binding. *In vitro* nucleotide substitutions and one insertion (+A) are shown above, consensus sequences and nucleotide identities (*) are given below the native sequence. ATG (underlined) at position 1270 marks the translation start of *rafA*.

may be activated by CAP through direct protein-protein interactions between one of the α-subunits of RNA polymerase (RNAP) and the promoter-proximal subunit of CAP (Irwin and Ptashne, 1987; Heyduk *et al.*, 1993; Ebright, 1993). CAP-RNAP contacts are necessary, but may not be sufficient for activation of transcription (Ryu *et al.*, 1994). DNA bending is clearly important, since intrinsically curved DNA can partially substitute for CBS and induce then CAP-specific effects both *in vivo* and *in vitro* (Bracco *et al.*, 1989; Gartenberg and Crothes, 1991; Perez-Martin and Espinosa, 1993, 1994; Perez-Martin *et al.*, 1994).

Further complexity in the regulation of sugar catabolic operons, as exemplified by the *gal* and *lac* systems, is introduced by the formation of promoter DNA-RNAP-repressor tertiary complexes (Straney and Crothers, 1987; Kuhnke *et al.*, 1989; Lee and Goldfarb, 1991) and DNA-RNAP-CAP-repressor quaternary complexes (Nick and Gilbert, 1985; Goodrich and McClure, 1992; Dalma-Weiszhausz and Brenowitz, 1996). The binding of repressor proteins, LacI and GalR, induces bending of cognate DNA (Zwieb *et al.*, 1989; Kuhnke *et al.*, 1989). X-ray structure analyses have shown that LacI binding induces distortion in the operator sequence causing it to bend away from the protein (Lewis *et al.*, 1996). It is thus believed that gene expression is regulated through interactions of numerous proteins during formation of the transcription initiation complex and not simply by mutual exclusion of effector proteins from the promoter DNA.

This study was initiated to define the *raf* CBS and to analyse the effects of CAP and RafR binding on *raf* regulatory DNA structure (hereafter referred to as *rafP*) and on *rafA* gene expression. We report here a uniquely compact array of regulatory sites, with CAP binding to a quasi-palindromic sequence centered at −60.5 bp and with CAP and two RafR dimers binding simultaneously, but non-cooperatively, to their cognate sites, CBS, O_1 and O_2. Both, CAP and RafR proteins induce DNA bending, but to different degrees and in opposite directions.

Results

Mapping of the CAP Site

The recognition helix of CAP binds to one half-site of the palindromic CAP consensus DNA binding site (de Crombrugghe *et al.*, 1984; Ebright *et al.*, 1984; Gunasekara *et al.*, 1992). Alignment with the *rafP* (Aslanidis and Schmitt, 1990; Aslanidis *et al.*, 1990) revealed a 64% sequence match upstream adjacent to the O_1 operator site (Figure 1). This presumptive CBS exhibits incomplete dyad symmetry with a canonical right-hand core and an imperfect left-hand half-site. For functional classification of the site,

deletion mapping and DNaseI footprinting have been used.

The native *rafP* sequence (pRU1307) and a set of five overlapping deletions extending from the left into the regulatory region were compared for *rafA* gene expression in CAP-proficient and CAP-deficient *E. coli* hosts. Deletions were introduced by endonuclease cleavage at already existing or artificially introduced restriction sites, namely, *Xba*I (T/CTAGA at pos. 1173), *Sau*3A (/GATC at pos. 1188), *Eco*52I (C/GGCCG at pos. 1196), *Sfu*I (TT/CGAA at pos. 1220) and *Nde*I (CA/TATG at pos. 1296), as diagrammed in Figure 1 (recognition sequences and assignments of post-cleavage 5′ nucleotide positions in parentheses). α-Gal activities determined in three *E. coli* hosts, CA8000 (*cya*$^+$, *crp*$^+$), CA8306 *(cya)* and XA3D1 *(crp)*, that were transformed with the deletion constructs, are listed in Table 1. These data revealed that

(i) the integrity of the putative CAP site is essential for *rafA* transcription, since a deletion extending to position 1173 (pRU1304) still retained more than 90% of wild-type activity;

(ii) however, a deletion extending only into the right-hand half of the palindromic site (pRU1303) displayed just 20% of wild-type activity;

(iii) all deletions beyond this site showed only background activity.

Moreover, the two *E. coli* hosts deficient in CAP (XA3D1) and in adenylate cyclase (CA8306) expressed less than 5% of wild-type α-Gal activity throughout, suggesting that it is the lack of CAP binding to its cognate site which causes the reduction in *rafA* gene expression.

The sequence bound by CAP was further defined by DNaseI footprinting. The binding of RafR alone and together with CAP was probed in this experiment to determine whether simultaneous binding of both regulatory proteins was feasible. The results presented in Figure 2 demonstrate that CAP protects a 30 bp region covering

Table 1 α-Galactosidase Activity of pRU1307 Derivatives with and without Catabolite Gene Activation.

Plasmid strain	α-Gal activity in *E. coli* strains[a]		
	CA8000 (*crp*$^+$ *cya*$^+$)	CA8306 (*crp*$^-$ *cya*$^-$)	XA3D1 (*crp*$^-$ *cya*$^+$)
pRU1307[b]	1780	34	62
pRU1304[b]	1650	33	65
pRU1303[b]	370	32	67
pRU1302[b]	15	17	60
pRU1301[b]	< 1	< 1	< 1
pRU1300[b]	< 1	< 1	< 1
pRU1324 (O_1^-)	< 1	2	ND[c]

[a] Enzyme activities are expressed as arbitrary units determined in the late exponential phase of growth (Muiznieks and Schmitt, 1994). Values were averaged from three independent experiments. The estimated standard deviation is ± 20%.

[b] Configurations of pRU1300 through pRU1307 are shown in Figure 1.

[c] ND, not determined.

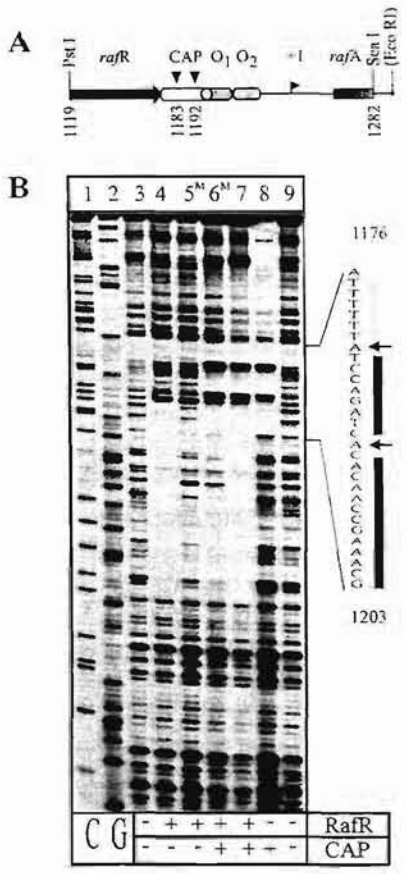


Fig. 2 DNaseI Footprinting of CAP and RafR Binding to *rafP* DNA.

(A) Diagram of the 198 bp *Eco*RI/*Pst*I fragment containing *rafP* that was [^{32}P]-labelled and subjected to DNaseI digestion. Symbols are as in Figure 1. (B) Lanes 1, 2; sequencing markers generated by the C or G chemical sequencing reactions; lanes 3, 9: free DNA treated with DNaseI; lanes 4 – 8: DNaseI digestion after binding of RafR, CAP, or both as indicated below (+/–); the superscript 'M' indicates the presence of 1 mM melibiose (inducer) in the binding reaction. Protection by CAP is illustrated by bars at the right-hand margin; shading indicates potential protection not seen due to an intrinsic DNaseI-insensitivity of the oligo T tract.

the entire CBS consensus with two DNaseI-hypersensitive sites, one in each half-site of the palindrome (Figure 2A, arrowheads). Their positions, 3 bp and 5 bp away from the dyad symmetry axis (at position 1186/1187), are slightly shifted to the right half-site, perhaps as a consequence of the non-ideal left-half sequence of the CBS (Figure 1). Moreover, footprinting demonstrated that concurrent binding of one CAP dimer and two RafR dimers to their cognate sites was feasible (Figure 2B, lanes 6 and 7). The patterns seen with various combinations of bound proteins in the presence and absence of 1 mM melibiose (inducer) can be accounted for as follows:

(i) Two RafR dimers cover the region between positions 1194 and 1244 containing O_1, O_2 and the – 35 promoter box (Figure 2B, lane 4). Unlike CAP, bound RafR completely protects the core sequences of O_1 and O_2, but it also introduces a new DNaseI-hypersensitive site at position 1194, where the CAP and O_1 sites overlap (Figure 2B, 4 – 7).

(ii) 1 mM melibiose reverses the operator-binding of RafR (Figure 2B, 5). The binding of CAP stabilises the repressor-DNA complex in the presence of melibiose (Figure 2B, 6) suggesting that under these conditions CAP may act as co-repressor. This assumption is further corroborated by the fact that pRU1324-specified α-Gal synthesis is repressed by RafR more efficiently in wild type (*cya*$^+$) than in an adenylate cyclase-deficient mutant (Table 1).

(iii) Footprinting defines the distance between the centres of CBS and O_1 as 18 bp, and between the axes of O_1 and O_2 as 21 bp (Figures 1 and 2). Therefore, the two RafR dimers bind at the same face of the DNA double helix, whereas the CBS is twisted by 103°.

Differential Binding of CAP and RafR

The gel shifts shown in Figure 3 further demonstrate that the CBS, O_1 and O_2 sites can be occupied either separately or simultaneously by cognate regulatory proteins. Five predictable combinations of protein-DNA complexes were formed, the upper two bands, however, at reduced intensity (Figure 3B). Complexes represented by these bands are the only combinations where both operators are occupied by RafR dimers. As established by Fried and Crothers (1981), the relative intensities of shifted bands at low protein concentration reflect both the probability of complex formation and the relative affinities of binding. The data shown in Figure 3B thus confirm our earlier observations that the close spacing of O_1 and O_2 causes mutual 'binding obstruction' between the two bound RafR dimers and that stable occupation of both sites requires a change in DNA conformation (Muiznieks and Schmitt, 1994). By contrast, no mutual obstruction between CAP and RafR binding was observed with mutant DNAs, where one or the other operator site was non-functional (Figure 3C, D). The observed differences in band intensities, which correspond to singly and doubly occupied promoter DNAs, where only one operator site is functional, do not support an interpretation in terms of binding co-operativity between CAP and RafR (Kleinschmidt *et al.*, 1991). This suggests that the relative twist of 103° between the dyad axes of the CBS and O_1 supports mutually independent binding of CAP and RafR proteins, whereas the simultaneous binding of the two RafR dimers to O_1 and O_2 at the same face of the double helix is impeded by steric hindrance and requires DNA bending to accommodate the two repressor molecules (Muiznieks and Schmitt, 1994).

Intrinsic and CAP/RafR-Induced DNA Bending

Various promoters are known to contain sequence-inherent bends that may participate in transcription initiation (Plaskon and Wartell, 1987; Lozinski *et al.*, 1991; Kim *et al.*, 1995). It has been shown that the affinity of CAP binding to DNA largely depends on such pre-existing bends and that

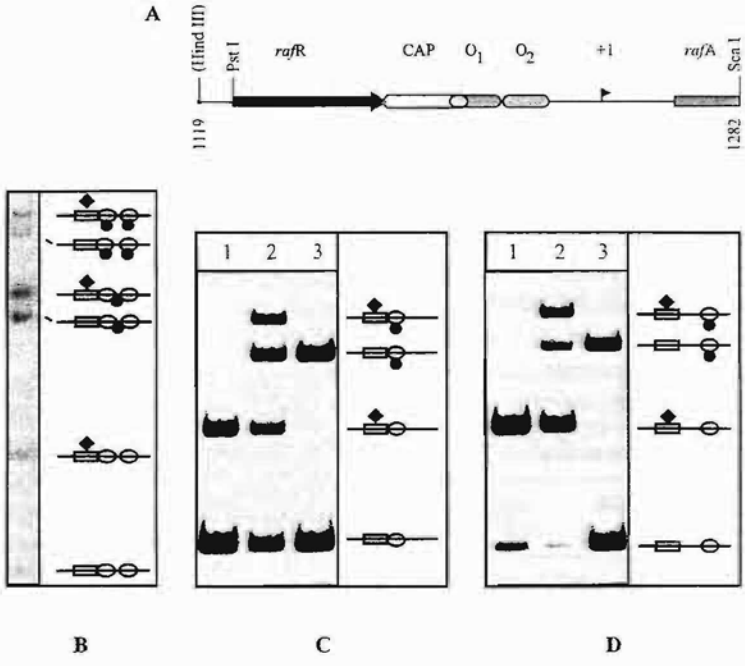


Fig. 3 Gel Shift Assays Demonstrating the Binding of CAP and/or RafR to *rafP*-DNA.
(A) Diagram of the 180 bp *Hin*dIII/*Sca*I fragment of pRU1307 used in these gel shift assays. Symbols as in Figure 1. (B) CAP and RafR proteins bound to wild-type DNA. Samples of 15 fmol 3′-labelled, gel-purified DNA incubated with RafR (5 ng) and CAP (5 ng) at 21 °C for 20 min, were separated on 4% PAGE. Pictograms to the right indicate the DNA-protein complexes represented by the respective band. Open rectangle, CAP binding site; open ovals, *raf* operators; solid diamond, CAP protein dimer; solid circle, RafR dimer. At non-saturating concentration, RafR has no preference for either O_1 or O_2, as symbolized by the solid circle placed between the operator sites. (C) CAP and RafR proteins bound to *rafP* DNA lacking a functional O_2. [^{32}P]-labelled pRU1324 DNA (15 fmol) was incubated with 5 ng RafR (lane 1), with 5 ng RafR plus 5 ng CAP (lane 2) or with 5 ng CAP (lane 3). (D) CAP and RafR proteins bound to *rafP* DNA lacking a functional O_1. Labelled pRU1324 DNA (5 fmol) was incubated with 5 ng RafR (lane 1), with 5 ng RafR plus 5 ng CAP (lane 2), or with 5 ng CAP (lane 3).

CAP tends to stabilise and enhance DNA bending (Kahn and Crothers, 1992).

To examine intrinsic and induced bending of DNA in the *raf* system, we have used the circular permutation EMSA (Wu and Crothers, 1984) both in the absence and in the presence of binding proteins. The 170 bp *Pst*I fragment containing the entire *rafP* (Figure 1) was circularly permuted by employing pre-existing or newly generated restriction sites, as shown in Figure 4A. The electrophoretic mobilities of fragments A, D and E (Figure 4A, lanes *Pst*I, *Sfu*I, *Alw*I) are considerably lower than those of B and C (Figure 4A, lanes *Bsa*JI, *Eco*52I) as illustrated by plotting their relative mobilities (Figure 4B). These data suggest that an intrinsic bend is centrally located in fragment D, ca. 40 bp upstream from the CBS dyad axis. However, this upstream curvature is not essential for *rafA* gene expression, since a deletion of this sequence (pRU1304; Figure 1) resulted in wild-type levels of α-Gal (Table 1).

X-ray structure analyses of CAP-CBS co-crystals (using the CBS consensus sequence) revealed that the DNA was bent by 90°, had two major kinks at the hypersensitive sites (Figure 1, arrowheads) and was wrapped around the protein (Schultz *et al.*, 1991). By contrast, Lac repressor binding to *lac*O induces a distortion in the operator sequence that causes the DNA to bend away from the protein (Lewis *et al.*, 1996).

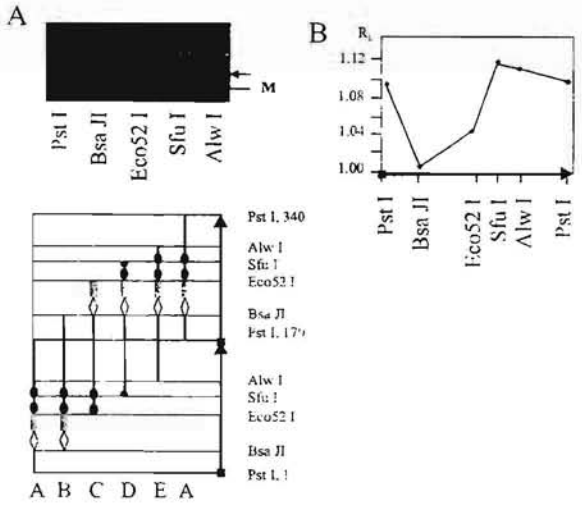


Fig. 4 Circular Permutation Gel Shift Analysis of Intrinsic Bending of *rafP* DNA.
(A) The electrophoretic mobility of a 340 bp tandem dimer of the *Pst*I fragment containing the *rafP* (Figure 1) was determined by non-denaturing 8% PAGE at 4 °C. By restriction with the enzymes indicated below the lanes, five 170 bp circular permutations with different locations of CAP (solid rectangle), O_1 and O_2 (shaded ovals) were obtained, as diagrammed below the gel. The site of intrinsic bending (diamond) was assigned according to Figure 4B. M, 165 bp marker fragment. Arrow denotes virtual position of the unbent 170 bp fragment. (B) Fragment relative lengths (R_L) plotted against the sequence permutation. R_L is the ratio of the mobility-derived apparent length of each fragment vs. the sequence length of corresponding fragments. R_L values are inversely correlated with the fragment end-to-end distance. This locates the center of bending 40 bp upstream of the CBS as depicted in Figure 4A.

Circular permutation EMSA with fragments A–E (Figure 4A) was used in the presence of CAP or RafR proteins. In these experiments, all fragments tested with bound RafR (Figure 5A) had only one functional operator owing to point mutations (fragments B_1, B_2) or to the position of the cleavage site used for permutation (fragments C, D). This resulted in clear-cut band shifts and facilitated a calculation of the degree of RafR-induced DNA bending. Electrophoretic mobilities of the permuted fragments binding either RafR (Figure 5A) or CAP protein (Figure 5B) and the distance between the fragment center and the dyad axis of the binding site (given below each lane) were directly correlated. Protein binding near the center of the DNA fragment led to considerably lower mobilities (Figure 5A, lane B1; Figure 5B, lane A) than binding near the end of the DNA fragment (Figure 5A, lane D; Figure 5B, lane C).

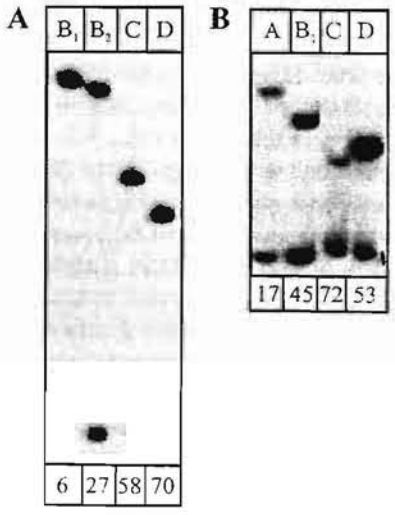


Fig. 5 Circular Permutation Gel Shift Analysis of *rafP* DNA Bending by CAP and RafR.
(A) Bending by RafR. DNA fragments (5 fmol) were [^{32}P]-end-labelled and incubated with RafR (5 ng) as in Figure 3. Circular permutations and letter code as in Figure 4A. B_1 (from pRU1324; Table 3) contains a non-functional O_1; B_2 (from pRU1327; Table 3) contains a non-functional O_2. The distance between the midpoint of each fragment and the dyad axis of the functional operator are given below (in bp). (B) Bending by CAP. Nomenclature of DNA fragments as in Figure 4, labelling and protein binding conditions as in Figure 3. The distances between the midpoint of each fragment and the dyad axis of the CBS are indicated below (in bp).

We have estimated the protein-induced bending angles in *rafP* by assuming that the electrophoretic mobility of a rigid DNA fragment is related to its end-to-end distance, *L*, which equals

$$L = L \cos \alpha/2$$

with L being the length of unbent DNA and α the angle, by which the DNA is bent away from linearity (Thompson and Landy, 1988). Since the end-to-end distance of a fragment with a protein bound to one end is essentially identical to L, it follows that

$$\mu_M/\mu_E = (L \times \cos \alpha/2)/L = \cos \alpha/2$$

where μ_M is the mobility of the complex with the protein bound at the center of DNA and μ_E is the mobility of the complex with the protein bound at one end (Kim *et al.*, 1989).

Accordingly, one RafR dimer bends DNA by an estimated 95° ± 5°, which is similar to the value reported for the GalR-induced DNA bend (Zwieb *et al.*, 1989). CAP on the other hand bends *rafP* by 75° ± 5°, which is somewhat lower than the reported 96° DNA bending, when CAP binds to its cognate site at *lac*P (Kim *et al.*, 1989). The 75° ± 5° bending, however, becomes plausible in view of the imperfect dyad symmetry of the *raf* CBS as opposed to its near-perfect counterpart in *lacP*.

Direction of DNA Bending

The relative direction of CAP- and RafR-induced DNA bends was determined by assessing the relative mobilities of CAP and RafR complexed with a set of five DNA fragments each with a different spacing between the CBS and O_2 (O_1 being non-functional). EMSA of the DNA-protein complexes shows that electrophoretic mobility shifts of the three-component, CAP, RafR, *rafP* DNA complex fluctuate according to the phasing of CBS and O_2. Placing CAP and RafR on the same face of the DNA double helix results in a smaller shift (Figure 6B, lane 3) than placing them on opposite faces (Figure 6B, lane 4). Changes in fragment length and nucleotide composition brought about variations of electrophoretic mobilities of the *rafP* derivatives, both as 'naked' DNA (Figure 6B, mark 0) and as CAP- or RafR-bound complexes (Figure 6B, marks C and R). To distinguish between the effects of spacer length, fragment length and nucleotide composition of DNA on mobility, the following equation was used:

$$N_{CR} = CR/(C/R)$$

where *CR*, *C* and *R* are the electrophoretic mobilities of CAP-RafR-DNA, CAP-DNA and RafR-DNA complexes, respectively, relative to the mobility of free DNA. The normalised relative mobilities of the three-component complexes *CR/(C/R)* were plotted against the phasing of the two binding sites (Figure 6C). The results suggest that the bends imposed on *rafP* DNA by CAP and RafR have opposite directions. None of the plasmids pRU1341–pRU1344, characterized by increasing distances between CBS and O_2, produced detectable amounts of α-Gal upon transformation of *E. coli* cells (results not shown). We thus conclude that the correct spacing of CBS and the −35/−10 promoter elements is an essential criterion for *raf* gene transcription.

Discussion

Spatial Structure and Regulation of *rafP*

We have defined and analysed the compound structure of the *raf* regulatory region featuring three closely spaced binding sites for one CAP and two RafR dimers. The −35 box is bounded by two operators, O_1 and O_2, that bind RafR at the same surface of the double helix, downstream of the CBS, whose centre is twisted by 103° towards O_1 and O_2. The three protein dimers bind non-cooperatively to DNA and bend it in opposite directions, namely, by 75° towards CAP and by 90° each, away from the RafR molecules. This unique and highly efficient regulatory device is diagrammed in Figure 7 and compared to equivalent structures of two other prominent catabolic operons, *lac* and *gal*.

A comparison of these schematic representations suggests that the compact arrangement of positive and negative regulators in *rafP* accomplishes a tight transcription control that in the *lac* and *gal* systems can be only achieved by repressor tetramerisation accompanied by DNA looping (Figure 7). This is further highlighted by comparing the *in vivo* contribution of corresponding regulatory elements to gene control among the *raf*, *lac* and *gal* operons (Table 2). It may be seen that

(i) CAP binding to *raf* CBS results in a distinctly higher increase in gene expression (50-fold) than CAP binding to *lac* CBS (20-fold) or *gal* CBS (18-fold);
(ii) control over gene expression by repressor binding to either the main or the auxiliary operators alone is between one and two orders of magnitude stronger in the *raf* than in the *lac* and *gal* regulatory systems;
(iii) only the concerted binding of repressors to the main and auxiliary operators results in a similarly tight transcription control for the *raf* and *lac* systems, whereas the effect on *gal* transcription is still less stringent (Table 2).

As portrayed in Figure 7, these regulatory systems follow two quite different strategies: the *lac* and *gal* controls with distantly spaced operators (which enclose the CBS and −35/−10 promoter boxes) have to rely on DNA looping and interaction of bound repressor molecules for maximum repression (Fried and Hudson, 1996; Lewis, 1996; Choy and Adhya, 1996). To the same end, the *raf* system uses an array of three tightly spaced binding sites, CBS – O_1 – O_2, interlocked with the −35/−10 promoter boxes and, thus, does not require DNA looping. Among sugar catabolic operons, this compact spatial structure of regulatory elements and divergent bending of DNA by bound proteins is unique and ensures a highly efficient transcription control by positive and negative regulatory proteins. – The arrangement of two operators and two divergent promoters in the *tet* regulatory region is somewhat reminiscent of the *rafP* configuration, also with respect to fine tuning of gene expression (Klock *et al.*, 1985; Tovar *et al.*, 1988); however, in view of lacking positive control by CAP the molecular mechanism regulating transcription is clearly different. Experiments reported here (Figure 2 and 3B) demonstrate that the three *raf* regulatory sites, CBS, O_1 and O_2, can be simultaneously occupied by their cognate proteins. This is different from *lacP*, where the O_3 site isomerisation and repressor displacement are needed to

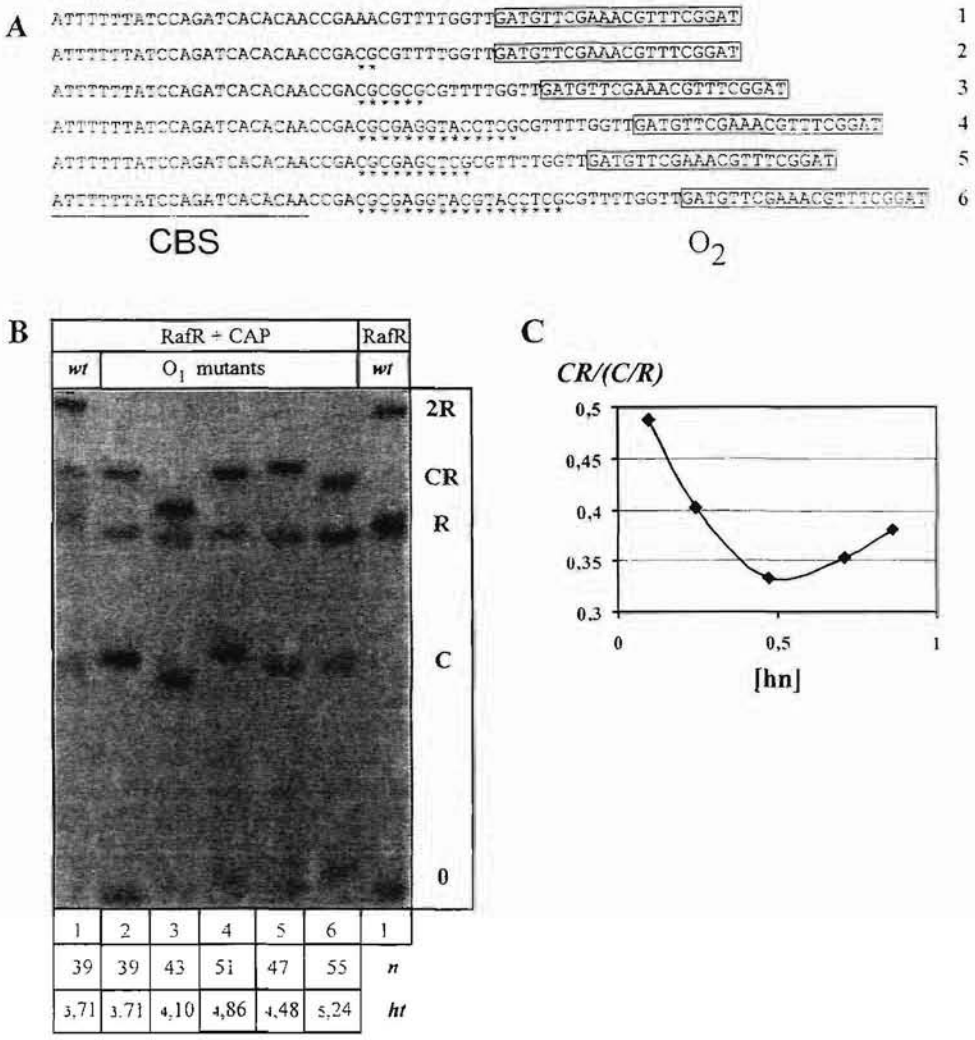


Fig. 6 Determination of Relative Directions of CAP- and RafR-Induced DNA Bends.
(A) Wild-type (1, pRU 1307) and five mutant rafP sequences (2, pRU1307M; 3, pRU1341; 4, pRU1342; 5, pRU1343; 6, pRU1344; as specified in Table 3) lacking functional O_1 with different spacings between the CBS (underlined) and O_2 (boxed) sites. Altered or inserted nucleotides are marked with asterisks. (B) Gel shift assays of the DNA fragments 1–6 from 6A. These were [^{32}P]-labelled and incubated with CAP and RafR proteins as in Figure 3. Symbols: *n*, number of bp between the dyad axes of CBS and O_2; *ht*, distance in helical turns, assuming B conformation of DNA with 10.5 bp per helical turn; 0, free DNA; C, DNA complexed with CAP protein; R, DNA complexed with RafR. CR, DNA complexed with CAP and RafR. (C) Normalized relative mobilities, *CR/(C/R)*, of the CAP-RafR-DNA three-component complex plotted against the phasing of CBS and O_2 binding sites, expressed as a decimal part of the number of helical turns separating the dyad axes of these sites. Data were taken from the experiment shown in Figure 6B. For details see text.

Table 2 Influence of Regulatory Sites and Their Cognate Proteins on Gene Expression Compared among the *raf*, *lac* and *gal* Regulatory Systems.

Regulatory system:	Mutation			Percent gene expression[c]		
Configuration[a]	*raf*	*lac*	*gal*	*raf*[c]	*lac*[d]	*gal*[e]
Constitutive	*rafR*	*lacI*	*galR*	100	100	100
Constitutive, CAP inactive	*rafR, cya*	*lacI, cya*	*galR, cya*	1.9	5.1	5.7
Main and auxiliary operators	*raf*$^+$	*lac*$^-$	*gal*$^+$	< 0.1	< 0.1	5.7
Only main operator	*raf*O_1	*lac*O_2O_3	*gal*O_E	0.2	5.6	34.5
Only auxiliary operator (s)	*raf*O_2	*lac*O_1	*gal*O_1	1.5	52.6	40.2

[a] Unless stated otherwise, CBS and repressor proteins are functional. Genetic configuration of regulatory elements is shown in Figure 7.
[b] Assessed by relative enzyme activities of α-galactosidase *(raf)*, β-galactosidase *(lac)* and galactokinase *(gal)*.
[c] Calculated from Muiznieks and Schmitt (1994) and from this work.
[d] Calculated from Mandecki and Caruthers (1984) and from Oehler *et al.* (1990).
[e] Calculated for *galP1* from Adhya and Miller (1979) and from Irani *et al.* (1983).

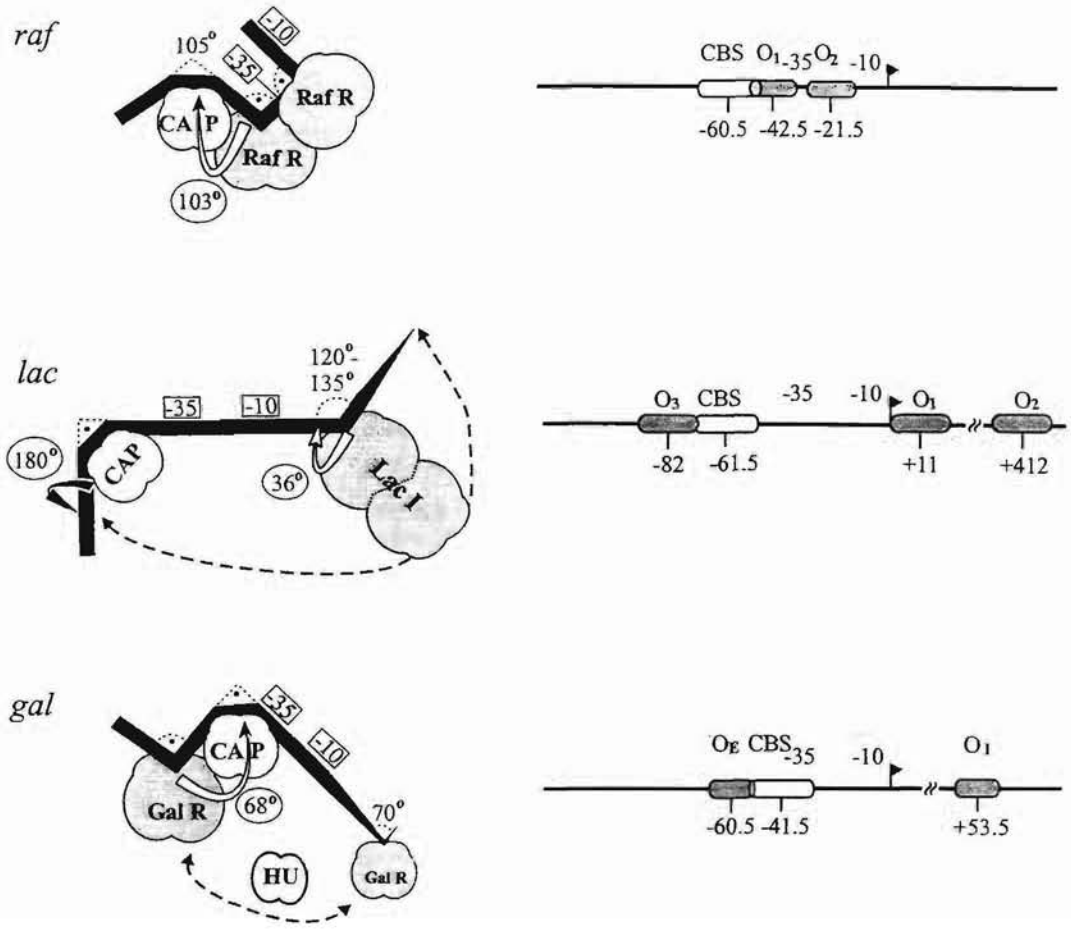


Fig. 7 Conformational Schemes (Left) and Genetic Organisation (Right) of the *raf*, *lac* and *gal* Regulatory Elements.
(Left) Spatial models of the dimeric (RafR, GalR) or tetrameric (LacI) repressor- and activator (CAP)-associated DNA complexes. Only the configurations where CAP is bound at its principal site are shown. Positions of the −35 and −10 promoter elements are assigned by boxed numbers. The degree of DNA bending (dashed arcs and triangles) by regulatory proteins was deduced from Wu and Crothers (1984), Zwieb *et al.* (1989), Kim *et al.* (1989), Schultz *et al.* (1991), Lewis *et al.* (1996) and from this work; (•) denotes a right angle (85–100°). The direction of bending by GalR has been included in analogy to the bending by RafR and LacI. Open arrows indicate the DNA twist between the centre of repressor binding and the centre of the CBS as deduced from the number of helical turns separating these sequence elements (assuming 10,5 bp per helical turn). Dashed lines with arrowheads indicate DNA looping. In *lac*, the possibility of DNA looping between O_1 and O_3 in the presence of CAP is disputed (Lewis *et al.*, 1996; Perros and Steitz, 1996; Lewis, 1996) and probably involves isomerisation of the O_3 site (filled arrow; Fried and Hudson, 1996). In *gal*, the histone-like protein, HU, is needed along with GalR for DNA looping and efficient repression (Aki *et al.*, 1996). The mode of Hu and GalR interactions is not known. (Right) Linear maps of regulatory sites of the three sugar-catabolic operons. Rounded boxes depict the positions of CBS (open) and operator (shaded) sites with their centres defined by bp numberings relative to the (+1) transcription start point (flag). Data on *lac* and *gal* regulatory sites were taken from Choy and Adhya (1996).

avoid steric collision of LacI with CAP (Fried and Hudson, 1996). The simultaneous binding of repressor and activator to *rafP* DNA is facilitated by the placement of overlapping CBS and O_1 sequences on twisted faces of the B-DNA helix. In footprinting experiments, the location of CAP- and RafR protected zones and the character of DNaseI-hypersensitive sites in footprinting assays did not reveal any interference of the two proteins (Figure 2B, lanes 4, 7, 8), nor did gel shift experiments indicate any cooperativity in CAP and RafR binding to *rafP* DNA (Figure 3). While it may appear paradoxical that the transcription activator, CAP, and the repressor, RafR, bind promoter DNA simultaneously, DNA bending induced by the joint action of CAP and RafR may be instrumental in recruiting RNAP to complete the transcription initiation complex.

The Regulatory Role of CAP

The axis of the *raf* CBS has been localised at position −60.5; its sequence matches the CBS consensus in 14 out of 22 positions (Figure 1). This resembles the CBS preceding the *lac* and *gal* promoters with 14 and 15 matches, respectively. However, the latter are near-evenly distributed along the binding sequence, whereas the *raf* CBS contains a canonical right half-site core, but conforms just in 6 out of 11 left-half positions. These deviations from the consensus sequence impede both, its binding and the bending of *rafP* DNA that was estimated as 75°, whereas the bending of *lac* DNA by CAP approximates 90° (Schultz *et al.*, 1991; Wu and Crothers, 1984; Kim *et al.*, 1989). The asymmetry of the *raf* CBS is thought to cause an asymmetric binding of CAP (Pyles and Lee, 1998), which, in

turn, reduces the degree of DNA bending. This corresponds to the five- to eight-fold lower affinity of CAP binding to *raf* CBS than to *lac* CBS (results not shown).

A deletion of the left half of the *raf* CBS (pRU1303, Figure 1), resulted in a 50-fold reduced binding efficiency of CAP compared to the native CBS. By contrast, *rafA* expression from the mutant plasmid pRU1303 was five- to ten-fold enhanced in a CAP-proficient *E. coli* compared to a *crp* or a *cya* mutant (Table 1). This discrepancy between *in vivo*-stimulation of gene expression by CAP and reduced *in vitro*-binding to and bending of DNA suggests that additional factors may be involved in transcription initiation in the cellular environment that have still to be identified.

Operater DNA Bending Induced by Repressor Binding

RafR induces prominent bends of the two adjacent operator sites, O_1 and O_2. From EMSA with circularly permutated *rafP* fragments, the extent of bending was calculated as 95° ± 5° for one functional operator (Figure 3). As inferred earlier (Muiznieks and Schmitt, 1994), sequence deterioration is needed for occupancy by RafR of two operator sites, whose centres are separated only by 21 bp. For this configuration, the additive angle of bending was indirectly determined. After introduction of a 31 bp synthetic DNA linker between O_1 and O_2, no significant increase in mobility shifts of the complex with two RafR dimers was observed over the native complex (results not shown), although increased spacing of the operator axes by three helical turns should have abolished any spatial hindrance. It is therefore proposed that two RafR dimers bound to the native O_1 and O_2, bend the DNA to about 95° ± 5° each, with the centres of bending located at the symmetry axes of the operators. The cumulative angle of DNA bending induced by two RafR dimers at *raf* O_1O_2 thus amounts to 190 ± 10° (Figure 7).

Materials and Methods

Bacterial Strains and Plasmids

Escherichia coli XL1-Blue (Stratagene Cloning Systems, La Jolla, CA, USA) was used for plasmid propagation as well as for RafR and CAP protein overexpression. *E. coli* CA8000 (*crp*$^+$ *cya*$^+$), CA8306 *(cya)* (Sabourin and Beckwith, 1975) and XA3D1 *(crp)* (Breul *et al.*, 1993) served to estimate the activation by CAP of different promoter mutants. Plasmids designed to test the effect of certain regulatory mutations on *rafA* gene expression and DNA bending are listed in Table 3.

Oligonucleotide-Directed Mutagenesis

The Amersham Sculptor *in vitro* mutagenesis system (Amersham Buchler, Braunschweig, Germany) was used as described before (Muiznieks and Schmitt, 1994).

Protein Preparation

RafR repressor was overproduced in *E. coli* XL1-Blue (pRU984) purified essentially as described by Aslanidis *et al.* (1990). The RafR inclusion bodies were solubilised in 0.1% SDS or 0.3% sarkosyl and 1:100 to 1:1000 dilutions were used directly for EMSA and DNaseI footprinting. Protein preparations were typically over 90% pure as estimated by densitometry of Coomassie Brilliant Blue R250-stained polyacrylamide gels. Lower SDS concentrations than 0.0001% did not interfere with subsequent protein-DNA binding assays.

CAP was overexpressed in *E. coli* XL1-Blue (pBG2) (Breul *et al.*, 1993). Both, crude cell lysates and affinity-purified CAP (Ghosaini *et al.*, 1988) were used for EMSA and DNaseI footprinting with no detectable differences in DNA binding. Binding buffer without Mg^{2+} ions was required, when CAP-enriched crude cell lysates were used. Protein concentrations were determined according to Bradford (1976).

Electrophoretic Mobility Shift Assay

Protein-DNA binding reactions and electrophoresis were performed as described by Aslanidis *et al.* (1990) and Muiznieks and Schmitt (1994) with slight modifications.

Table 3 Plasmids Used in This Study

Plasmid[a]	Source and derivations
pRU1300	*NdeI*/*Hin*dIII deletion of pRU1307N, coding sequence of *rafA* without promoter (Figure 1)
pRU1301	*Sfu*I (*Bst*BI)/*Hin*dIII deletion of pRU1307 retaining 28 bp of *rafP* upstream of the transcription start point (Figure 1)
pRU1302	*Eag*52I/*Hin*dIII deletion of pRU1307E retaining 52 bp of *rafP* upstream of the transcription start point (Figure 1)
pRU1303	*Sau*3A/*Hin*dIII deletion of pRU1307 retaining 60 bp of *rafP* upstream of the transcription start point (Figure 1)
pRU1304	*XbaI*/*Hin*dIII deletion of pRU1307 retaining 76 bp of *rafP* upstream of the transcription start point (Figure 1)
pRU1307	Native *rafP*-*rafA* configuration without *rafR* inserted into pUC8 (Muiznieks and Schmitt, 1994)
pRU1307X	Mutagenesis of pRU1307, A insertion between 1174 and 1175 bp to create an *Xba*I site
pRU1307E	Mutagenesis of pRU1307, transition of AA at 1196, 1197 bp to GG to create an *Eco*52I (*Eag*I) site
pRU1307M	Mutagenesis of pRU1307, substitution of AA at 1202, 1203 bp to CG to create a *Mlu*I site
pRU1307N	Mutagenesis of pRU1307, substitution of T at 1268 bp by A to create an *Nde*I site
pRU1324	Mutagenesis of pRU1307, G insertion between 1200 and 1201 bp to inactivate O_1
pRU1327	Mutagenesis of pRU1307, substitution of TT at 1227/1228 bp by GG to inactivate O_2
pRU1341	Filling-in of *Mlu*I overhanging ends in pRU1307M DNA resulting in a 4-bp addition between the CAP and O_1 sites
pRU1342	Ligation of the CGCGAGGTACCT linker in *Mlu*I site of pRU1307M introducing 12 bp between CAP and O_2
pRU1343	Filling-in of *Kpn*I overhanging ends in pRU1342 DNA resulting in an 8-bp addition between CAP and O_2
pRU1344	Filling-in of *Mlu*I overhanging ends in pRU1342 resulting in a 16 bp addition between CAP and O_2

[a] Construction of pRU984 and sequence numbering according to Aslanidis *et al.* (1990) and Muiznieks and Schmitt (1994).

Approximately 10 fmol DNA fragment and 1–100 ng cognate protein in binding buffer (10 mM Tris-HCl pH 8.0, 10 mM KCl, 1 mM DTT, 50 μg/ml BSA, 0.01% NP40, 3% glycerol) were allowed to react at room temperature for 20 min. Where needed, cAMP was included at 200 μM. The addition of 3% glycerol to the binding buffer allowed direct loading of samples on the gel and facilitated protein binding to DNA. Mg^{2+} ions were excluded from the reaction to prevent DNA degradation by nucleases from crude cell lysates sometimes used as a source of CAP.

DNaseI Protection Assay

DNaseI protection assays followed the protocol of Aslanidis *et al.* (1990). Protein-DNA complexes were allowed to form as for EMSA in 100 μl reaction mixture. 7 μl DNaseI, diluted in binding buffer plus 2.5 mM Ca^{2+} – 3 mM Mg^{2+} were added quickly and after 20 seconds the reaction was terminated by the addition of 20 μl DNaseI stop solution (2 M ammonium acetate, 20 mM EDTA, 0.2 mg/ml carrier DNA). The mixture was extracted with phenol and with phenol:chloroform (1:1), and DNA was then ethanol precipitated, washed in 70% ethanol, vacuum-dried and redissolved in 10 μl loading buffer (80% formamide, 1 mM EDTA, 0.1% xylene cyanol and bromophenol blue), incubated at 100 °C for 5 min and chilled on ice; 3 μl of the solution were loaded on an 8% polyacrylamide sequencing gel and run in 90 mM TRIS-borate (pH 8.5), 2.5 mM EDTA, 7 M urea.

α-Galactosidase Assay

α-Gal activity was determined as described by Muiznieks and Schmitt (1994). Activity is expressed as the maximum increment of specific activity (units per h) during exponential and early stationary growth. One unit of enzyme activity corresponds to a change of 0.01 OD_{420} (owing to hydrolysis of the chromogenic substrate *p*-nitrophenyl-α-D-galactopyranoside) per min per OD_{600}. Relative plasmid dosages were determined according to Stueber and Bujard (1982) for normalising the values of enzyme activity.

Acknowledgements

We thank Klaus Stark for the artwork. This study was supported by grants from the Latvian Council of Science to I.M. (grant 96-923) and from the Deutsche Forschungsgemeinschaft to R.S. (Schm 68/24-3). N.R. was recipient of a doctoral fellowship from the Latvian Council of Science, a grant from Danish Councils of Science, and a TEMPUS travel grant (Project S_JEP-09273-95).

References

Adhya, S., and Miller, W. (1979). Modulation of the two promoters of the galactose operon of *Escherichia coli*. Nature *279*, 492–494.

Aki, T., Choy, H.E., and Adhya, S. (1996). Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of *gal* transcription by Gal repressor. Genes to Cells *1*, 179–188.

Aslanidis, C., and Schmitt, R. (1990). Regulatory elements of the raffinose operon: nucleotide sequences of operator and repressor genes. J. Bacteriol. *172*, 2178–2180.

Aslanidis, C., Schmid, K., and Schmitt, R. (1989). Nucleotide sequences and operon structure of plasmid-borne genes mediating uptake and utilization of raffinose in *Escherichia coli*. J. Bacteriol. *171*, 6753–6763.

Aslanidis, C., Muiznieks, I., and Schmitt, R. (1990). Successive binding of *raf* repressor to adjacent *raf* operator sites *in vitro*. Mol. Gen. Genet. *223*, 297–304.

Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S., and Buc, H. (1989). Synthetic curved DNA sequences can act as transcriptional activator in *Escherichia coli*. EMBO J. *8*, 4289–4296.

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principles of protein-dye binding. Anal. Biochem. *72*, 248–254.

Breul, A., Assmann, H., Golz, R., von Wilcken-Bergmann, B., and Müller-Hill, B. (1993). Mutants with substitutions for Glu171 in the catabolite activator protein (CAP) of *Escherichia coli* activate transcription from the *lac* promoter. Mol. Gen. Genet. *238*, 155–160.

Choy, H., and Adhya, S. (1996). Negative regulation. In: *Escherichia coli* and *Salmonella typhimurium*. 2nd ed; F.C. Neidhardt *et al.*, eds. (Washington, DC, USA: ASM Press), pp. 1287–1299.

Collado-Vides, J., Magasanik, B., and Gralla, J.D. (1991). Control site location and transcription regulation in *Escherichia coli*. Microbiol. Rev. *55*, 371–394.

de Crombrugghe, B., Busby, S., and Buc, H. (1984). Cyclic AMP receptor protein: the role in transcription activation. Science *224*, 831–838.

Dalma-Weiszhausz, D.D., and Brenowitz, M. (1996). Interactions between RNA polymerase and the positive and negative regulators of transcription at the *Escherichia coli gal* operon. Biochemistry *35*, 3735–3745.

Ebright, R.H. (1993). Transcription activation at Class I CAP-dependent promoters. Mol. Microbiol. *8*, 797–802.

Ebright, R.H., Cossart, P., Gicquel-Sanzey, B., and Beckwith, J. (1984). Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. Nature *311*, 232–235.

Fried, M., and Crothers, D.M. (1981). Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nucl. Acids Res. *9*, 6505–6525.

Fried, M., and Hudson, J. (1996). DNA looping and *Lac* repressor-CAP interaction. Science *274*, 1930–1931.

Gartenberg, M.R., and Crothers, D.M. (1988). DNA sequence determinants of CAP-induced bending and protein binding affinity. Nature *333*, 824–829.

Gartenberg, M.R., and Crothers, D.M. (1991). Synthetic DNA bending sequences increase the rate of *in vitro* transcription initiation at the *Escherichia coli lac* promoter. J. Mol. Biol. *219*, 217–230.

Ghosaini, L.R., Brown, A.M., and Sturtevant, J.M. (1988). Scanning calorimetric study of the thermal unfolding of catabolite activator protein from *Escherichia coli* in the absence and presence of cyclic mononucleotides. Biochemistry *27*, 5257–5261.

Goodrich, J.A., and McClure, W.R. (1992). Regulation of open complex formation at the *Escherichia coli* galactose operon promoters. Simultaneous interaction of RNA polymerase, *gal* repressor and CAP/cAMP. J. Mol. Biol. *224*, 15–19.

Gronenborn, A.M., Nermut, M.V., Eason, P., and Clore, G.M. (1984). Visualization of cAMP receptor protein-induced DNA kinking by electron microscopy. J. Mol. Biol. *179*, 751–757.

Gunasekera, A., Ebright, Y.W., and Ebright, R.H. (1992). DNA sequence determinants for binding of the *Escherichia coli* catabolite gene activator protein. J. Biol. Chem. *267*, 14713–14720.

Heyduk, T., Lee, J.C., Ebright, Y.W., Blatter, E.E., Zhou, Y., and Ebright, R.H. (1993). CAP interacts with RNA polymerase in solution in the absence of promoter DNA. Nature *364*, 548–549.

Irani, M.H., Orozs, L., and Adhya, S. A control element within a structural gene: The *gal* operon of *Escherichia coli*. Cell *32*, 783–788.

Irwin, N., and Ptashne, M. (1987). Mutants of the catabolite activator protein of *Escherichia coli* that are specifically deficient in the gene-activation function. Proc. Natl. Acad. Sci. USA *23*, 8315–8319.

Kahn, J.D., and Crothers, D.M. (1992). Protein-induced bending and DNA cyclization. Proc. Natl. Acad. Sci. USA *89*, 6343–6347.

Kim, J., Zwieb, C., Wu, C., and Adhya, S. (1989). Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene *85*, 15–23.

Kleinschmidt, C., Tovar, K., and Hillen, W. (1991). Computer simulations and experimental studies of gel mobility patterns for weak and strong non-cooperative protein binding to two targets on the same DNA: application to binding of *tet* repressor variants to multiple and single Tet operator sites. Nucl. Acids Res. *19*, 1021–1028.

Klock, G., Unger, B., Gatz, C., Hillen, W., Altenbuchner, J., Schmid, K., and Schmitt, R. (1985). Heterologous repressor-operator recognition among four classes of tetracycline resistance determinants. J. Bacteriol. *161*, 326–332.

Kolb, A., Busby, S., Buc, H., Garg, S., and Adhya, S. (1993). Transcriptional regulation by cAMP and its receptor protein. Annu. Rev. Biochem. *62*, 749–795.

Kuhnke, G., Theres, C., Fritz, H.J., and Ehring, R. (1989). RNA polymerase and *gal* repressor bind simultaneously and with DNA bending to the control region of the *Escherichia coli* galactose operon. EMBO J. *8*, 1247–1255.

Lanzer, M., and Bujard, H. (1988). Promoters largely determine the efficiency of repressor action. Proc. Natl. Acad. Sci. USA *85*, 8973–8977.

Lee, J., and Goldfarb, A. (1991). *lac* repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter. Cell *66*, 793–798.

Lewis, M. (1996). DNA looping and *Lac* repressor-CAP interaction. Science *274*, 1931–1932.

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*, 1247–1254.

Lozinski, T., Adrych-Rozek, K., Markiewicz, W.T., and Wierzchowski, K. (1991). Effect of DNA bending in various regions of a consensus-like *Escherichia coli* promoter on its strength *in vivo* and structure of the open complex *in vitro*. Nucleic Acids Res. *19*, 2947–2953.

Mandecki, W., and Caruthers, M.H. (1984). Mutants of the *lac* promoter with large insertions and deletions between the CAP binding site and the –35 region. Gene *31*, 263–267.

Muiznieks, I., and Schmitt, R. (1994). The role of two operator sites in regulating the plasmid-borne *raf* operon of *Escherichia coli*. Mol. Gen. Genet. *242*, 90–99.

Nick, H., and Gilbert, W. (1985). Detection *in vivo* of protein DNA interactions with the *lac* operon of *Escherichia coli*. Nature *313*, 795–798.

Oehler, S., Eismann, E.R., Kramer, H., and Müller-Hill, B. (1990). The three operators of the *lac* operon cooperate in respression. EMBO J. *9*, 973–979.

Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y.W., Ebright, R.E., and Berman, H.M. (1996). Structure of the CAP-DNA complex at 2.5 Angstroms resolution: a complete picture of the protein-DNA interface. J. Mol. Biol. *260*, 395–408.

Perez-Martin, J., and Espinosa, M. (1993). Protein-induced bending as a transcriptional switch. Science *260*, 805–807.

Perez-Martin, J., and Espinosa, M. (1994). Correlation between DNA bending and transcriptional activation at a plasmid promoter. J. Mol. Biol. *241*, 7–17.

Perez-Martin, J., Rojo, F., and DeLorenzo, V. (1994). Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiol. Rev. *58*, 268–290.

Perros, M., and Steitz, T.A. (1996). DNA looping and *Lac* repressor-CAP interaction. Science *274*, 1929–1930.

Plaskon, R.R., and Wartell, R.M. (1987). Sequence distributions associated with DNA curvature are found upstream of strong *E. coli* promoters. Nucl. Acids Res. *15*, 785–796.

Pyles, E.A., and Lee, J.C. (1998). *Escherichia coli* cAMP receptor protein-DNA complexes. 2. Structural asymmetry of DNA bending. Biochemistry *37*, 5201–5210.

Ryu, S., Garges, S., and Adhya, S. (1994). An arcane role of DNA in transcription activation. Proc. Natl. Acad. Sci. USA *91*, 8582–8586.

Sabourin, D., and Beckwith, J. (1975). Deletion of the *Escherichia coli crp* gene. J. Bacteriol. *122*, 338–340.

Schmid, K., Ritschewald, S., and Schmitt, R. (1979). Relationships among raffinose plasmids determined by the immunochemical cross-reaction of their α-galactosidases. J. Gen. Microbiol. *114*, 477–481.

Schmitt, R., Mattes, R., Schmid, K., and Altenbuchner, J. (1979). Raf plasmids in strains of *Escherichia coli* and their possible role in enteropathogeny. In: Plasmids of Medical, Environmental and Commercial Interest, K.N. Timmis and A. Pühler, eds. (Amsterdam, Holland: Elsevier/North Holland Publishing Co.), pp. 199–210.

Schultz, S.C., Shields, G.C., and Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. Science *253*, 1001–1007.

Straney, S.B., and Crothers, D.M. (1987). Lac repressor is a transient gene-activating protein. Cell *51*, 699–707.

Stueber, D., and Bujard, H. (1982). Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid-specified genes. EMBO J. *1*, 1399–1404.

Su, T., Qi, S., Yun, W., and Xiu, L. (1989). Regulation in the expression of α-galactosidase gene in *raf* operon in *Escherichia coli*. Acta Microbiol. Sinica *29*, 180–186.

Thompson, J.F., and Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: application to λ-specific recombination complexes. Nucl. Acids Res. *20*, 9687–9705.

Tovar, K., Ernst, A., and Hillen, W. (1988). Identification and nucleotide sequence of the class E *tet* regulatory elements and operator and inducer binding of the encoded purified Tet repressor. Mol. Gen. Genet. *215*, 76–80.

Ulmke, C., Lengeler, J.W., and Schmid, K. (1997). Identification of a new porin, RafY, encoded by raffinose plasmid pRSD2 of *Escherichia coli*. J. Bacteriol. *179*, 5783–5788.

Wu, H.-M., and Crothers, D.M. (1984). The locus of sequence-directed and protein-induced DNA bending. Nature *308*, 509–513.

Zinkel, S.S., and Crothers, D.M. (1990). Comparative gel electrophoresis measurement of the DNA bend angle induced by the catabolite activator protein. Biopolymers *29*, 29–38.

Zwieb, C., Kim, J., and Adhya, S. (1989). DNA bending by negative regulatory proteins: Gal and Lac repressors. Genes Dev. *3*, 606–611.

Received September 9, 1998; accepted November 10, 1998

3

The third promoter element regulating the basal expression level of plasmid-borne raffinose operon promoter *rafP*

Nils Rostoks[a], Indrikis Muiznieks[a], Rüdiger Schmitt[b]*

[a] Faculty of Biology, University of Latvia, Riga, LV-1586, Republic of Latvia

[b] Lehrstuhl für Genetik, Universität Regensburg, D-93040 Regensburg, Germany

Keywords: Raffinose operon, transcription regulation, UP element

*Correspondence to:

Dr. Rüdiger Schmitt

Lehrstuhl für Genetik, Universität Regensburg

D-93040 Regensburg, Germany

Tel. +49-941-9433162

Fax +49-941-9433163

e-mail: rudy.schmitt@biologie.uni-regensburg.de

Abstract

We show that in *rafP*, the promoter of *Escherichia coli* plasmid born raffinose catabolism operon, the sequence, which is located between the positions –52 and -39 in respect to the mRNA start site is indispensable for non-activated, basal level of expression. This sequence and the conventional –35 promoter element both are equally important for the transcription of *rafP* in the absence of cAMP receptor protein, CRP. Functionally the -52/-39 sequence resembles an UP element of constitutively expressed bacterial promoters. However, it has week sequence homology to the UP element consensus. The sequence and DNA topology determinants essential for the *rafP* UP element are discussed.

1. Introduction

Our understanding of the regulation of bacterial gene expression has been obtained mainly from the studies, which model extreme levels of promoter activation or repression due to the binding of specific transcription factors. This hardly reflects the mode of gene transcription in bacterial cells growing in complex natural environments. The maintenance of cellular household functions depends upon the basal activity of constitutive promoters, which are primarily regulated by the promoter sequence elements [1] and DNA topology [2]. Conventional –35 and –10 promoter boxes are extensively characterized [1]. For the group of promoters, which regulate ribosomal RNA synthesis, an A/T rich sequence or UP element located immediately in 5'-direction from the –35 element is critical for effective transcription without any additional protein factors [3, 4]. Similar elements have been found also in other promoters [5-7]. The consensus sequence of UP element has been deduced [8]. UP element-containing promoters may be activated also by different transcription factors. In the case of *lacP1* [9, 10], UP element and the cAMP receptor protein (hereafter referred to as CRP) do not show additive effect. The same amino acids in the C-terminal domain of RNA polymerase (RNAP) α subunit are supposed to interact both with CRP and with UP element. Binding of CRP may hide away the UP element sequence from RNAP α subunit [11]. This implies that UP elements are needed mostly to ensure the basal transcription level.

The plasmid born raffinose catabolism operon (*raf* operon) of *Escherichia coli* includes *rafA* (α-galactosidase, α-Gal), *rafB* (permease) and *rafD* (invertase) genes, which are needed for the uptake and utilization of the trisaccharide raffinose [12]. It belongs to the group of sugar catabolism operons, which are induced by the presence

of the substrate and activated by CRP in response to glucose depletion in the growth environment. The *raf* genes are co-transcribed from a common promoter, *rafP*. The promoter is controlled by the RafR repressor binding to two operators, O_1 and O_2, which embrace the –35 sequence of *rafP* [13, 14]. RafR is encoded by the upstream *rafR* gene, which is transcribed from its own promoter P_R [15]. The only known inducer of the *raf* operon is a disaccharide, melibiose. The inducer may result from raffinose hydrolysis by *raf*D invertase [13]. Melibiose may be found also in complex polysaccharide sources, e.g., in malt and molasses. The *rafP* belongs to the Class I of CRP activated promoters [16].

In the natural environment, which usually contains mixture of carbohydrates, *raf* operon should be induced, but not activated [13, 14, 16]. In this work we have studied *rafP* sequence elements, which are required for the basal activity of α-Gal synthesis in the absence of RafR-dependent repression and CRP-dependent activation. Our study demonstrates a new, UP-related, sequence element in *rafP* and characterizes its impact on the promoter activity.

1. Materials and methods

2.1. Bacterial strains, plasmids and cultivation conditions

Bacterial strains and plasmids used in this study are listed in Table 1. The structure of wt *rafP* and its derivatives is outlined in the Fig. 1. *E. coli* strains were cultured in standard 2 x TY liquid medium or on LB agar plates supplemented with 100 μg/ml ampicillin [20].

2.2. Plasmid construction and site-directed mutagenesis

Basic plasmid DNA manipulations were done as described in [20]. Site-directed mutagenesis was done by Amersham Sculptor *in vitro* mutagenesis kit (Amersham Pharmacia Biotech Europe GmbH, Freiburg, Germany) or by PCR-based mutagenesis [21]. The oligonucleotide substitutions of the *rafP* region in 5'-direction from the position -28 were done by cloning of synthetic oligonucleotides (MWG Biotech, Ebersberg, Germany) (Fig. 1C) between the *Hin*dIII and *Bst*BI sites of the plasmid pRU1307 [14]. The oligonucleotide substitutions affected also the sequence of –35 element, which was optimized to match the consensus TTGACA. Oligonucleotides, which replace wt *rafP* sequence in pRU1369 contain 4 runs of $(T)_n(A)_n$ tracts, in pRU1371 – 4 runs of $(A)_n(T)_n$ tracts (n=4-5) in phase with the DNA helical turn. The A/T tracts of the oligonucleotides were shifted by ca. half a helical turn in respect to –35 element of *rafP* by filling-in of the resident (in pRU1369) or mutagenesis-restored (in pRU1371) *Mun*I sites by DNA polymerase I Klenow fragment. The ensuing plasmids are pRU1369+4 and pRU1371+5 respectively, where

four or five nucleotides are added between the oligonucleotide and –35 element sequences.

2.3. Polyacrylamide gel electrophoresis of DNA

Sequence dependent DNA curvature of *rafP* derivatives was characterized by the analysis of DNA fragment electrophoretic mobility in non-denaturing polyacrylamide gels (PAAG) as described in [16]. Intrinsically bent DNA fragments migrate in PAAG electrophoresis slower than straight molecules of identical sequence length. Coefficient K_R is calculated as the ratio of apparent fragment length estimated in PAAG electrophoresis versus its sequence length. K_R is inversely correlated to the fragment end-to-end length and it may be used to calculate the angle of DNA bending [22].

2.4. α-Gal assay

The efficiency of *rafP*–directed transcription was assessed using *rafA* as the natural reporter gene. The product of the *rafA* gene, α-Gal, was assayed as described in [23], by the liberation of the *p*-nitrophenol from *p*-nitrophenyl–α–D-galactopyranoside, and the resulting increase of absorption at 420 nm. The activity was measured twice during the middle logarithmic growth phase of liquid cultures. The activity was expressed as the change of OD_{420} per minute per one OD_{600} of the cells. At least three independent assays were performed for each experiment. Estimated error was ca. 20%.

3. Results and discussion

3.1. Localization of the rafP UP element

The transcription of *E. coli* raffinose operon in complex natural environments may occur at the basal level, because of the induction of repressor RafR by melibiose and exclusion of activator CRP by low level of cAMP. Under these conditions only the promoter DNA sequence and topology are responsible for the regulation of *rafP*. In order to determine the role of upstream promoter region in regulation of basal *rafP* activity, the mutagenesis and deletion analysis of the corresponding region was carried out (Table 1, Fig. 1).

The plasmids carrying *rafP* derivatives were tested in *rafR*$^-$, *crp*$^-$ and *cya*$^-$ genetic background to assess the basal expression level. The production of α-Gal by various promoter constructions is summarized in the Table 2. Enzyme activity was expressed in % of the activity produced by wt *rafP* in pRU1307, which was 0.34±0.07 arbitrary units.

The deletions upstream from the position –59 in respect to the mRNA start site (pRU1303) did not affect the basal activity of the promoter. Further consecutive deletions caused gradual decrease in *rafP* activity. pRU1302, with the deletion of *rafP* sequence upstream from position –52, displayed ca. 36% of the wt promoter activity. Deletion of the region upstream from position –45 in pRU1305 diminished α-Gal production 9-fold, and the deletion removing –35 region reduced the *rafP* activity ca. 50-fold (pRU1301). The deletion in *rafP*, which retained only 15 nucleotides upstream from the α-Gal mRNA start point, exhibited no detectable promoter activity [16].

The promoter sequence in the plasmids pRU1302 and pRU1305 was truncated using the restriction enzyme sites, which were introduced by site-directed mutagenesis in the plasmids pRU1307E and pRU1307M, respectively. These mutations themselves had detrimental effect upon basal *rafP* activity. AA to GG replacement in the pRU1307E at the positions –51 and –50 reduced the *rafP* activity ca. 2-fold, whereas AA to CG replacement in the pRU1307M at the positions –45 and –44 reduced the *rafP* activity 4.5-fold. Non-adjacent mutations A to G and T to G in the pRU1340 at the positions –46 and –40 reduced the *rafP* activity 6-fold. Mutations introduced in the –35 box of the *rafP* (pRU1307B and pRU1307C) diminished the wt *rafP* activity 100- and 50-fold, respectively. This was comparable with the activity of promoter, where –35 region was deleted (pRU1301).

Mutagenesis of the positions –40 and –45 of *rafP* demonstrated that any changes in wt sequence were detrimental to the *rafP* activity (Table 2). Substitution of T at the position –40 by C or G reduced *rafP* activity ca. 5-fold, while substitution by A caused only 3-fold decrease. Substitution of A at the position –45 by C or G reduced activity ca. 2- to 2.5-fold and by T – less than 2-fold. Replacement of wt *rafP* sequence by unrelated oligonucleotide in plasmid pRU1372 resulted in more than 50-fold decrease of α-Gal activity, although the –35 element of the promoter was not impaired.

The deletion and mutation analysis between the positions –52 and –39 in respect to mRNA start site in *rafP* disclosed a novel sequence element. Alterations in the structure of this element diminished the basal activity of *rafP in vivo* up to 50-fold. This element and –35 sequence were equally important for the CRP non-activated expression of the promoter, their defects could not be mutually complemented. Within this element individual point mutations displayed cumulative

effect and promoter - proximal sequence changes were more deleterious. Thus, -52/–39 sequence in *rafP* show functional features of the UP elements, which are described in a number of other promoters [8].

3.2. Comparative analysis of the rafP UP element structure

The comparison of the wt *rafP* –52/-39 region with the consensus sequence of UP element [8] revealed sparse homology, only four nucleotide matches within 15 defined positions (Fig. 2). Eight nucleotides in the UP element of *rafP* are G or C, while UP consensus includes only A or T bases. In plasmids pRU1369 and pRU1371 we replaced the wt *rafP* –52/-39 region with the synthetic oligonucleotides containing $(T)_n(A)_n$ or $(A)_n(T)_n$ tracts. The number of A and T residues within the UP element region of these plasmids was increased to 15 and 12, respectively, although the direct match with the consensus sequence remained low (Fig. 2). Neither pRU1369, nor pRU1371 produced significant α-Gal activity (Table 2). Further mutagenesis and insertion of four and five nucleotides between the cloned synthetic sequence and –35 element of *rafP* generated plasmids pRU1369+4 and pRU1371+5. Here the homology with the UP element consensus sequence was improved (10 matches from 15 in both plasmids), but the number of A or T nucleotides within UP element region was reduced (Fig. 2). pRU1369+4 still produced hardly detectable α-Gal activity, but in pRU1371+5 *rafA* expression was increased more than 25-fold in comparison to pRU1371.

These results should be interpreted in context of the promoter topology. According to the preliminary data and theoretical prediction [25], the oligonucleotide sequence in pRU1369 and 1369+4 should not contain significant sequence-directed

DNA curvature, while the oligonucleotide in pRU1371 and pRU1371+5 should be statically bent at 80 – 90°. The electrophoretic mobility analysis of 215 bp restriction fragment spanning *rafP* region from these plasmids in non-denaturing PAAG confirmed the predictions. K_R value for pRU1369 and pRU1369+4 fragments was 1.02; for pRU1371 and pRU1371+5 – 1.45 (results not shown).

Our data (Table 2, Fig. 2) indicate that the abundance of A and T nucleotides alone and even high homology to the consensus sequence is not sufficient for the proficiency of UP element in *rafP*. The spatial structure of UP element is as important for the recruitment of RNAP as are the sequence determinants. The activity of UP element in *rafP* activity can be mimicked by properly phased segment of intrinsically curved DNA fragment (pRU1371 and pRU1371+5, Table 2). Similarly, promoter activation by CRP may be partially replaced by statically curved DNA fragments [26]. DNA flexibility, which is increased by G or C nucleotide patches within monotonous and stiff A/T tracts [27] may also contribute to UP element functionality.

In *rafP*, UP element may stabilize RNAP binding to –35 region by providing additional contacts with the C-terminal domain of α subunit. This assumption is supported by the fact that CRP activation of the promoter largely abolishes the impact of UP element. Point mutations within the *rafP* UP sequence at the positions –40 and –45 decreased the promoter activity in crp^-, but not in crp^+ strain (results not shown).

Acknowledgements

This work was supported by grants from the Deutsche Forschungsgemeinschaft to R.S. (Schm 68/24-3) and from the Latvian Council of Science to I.M. (grant 96-923). N.R. was supported by a doctoral fellowship from the Latvian Council of Science and a travel grant from the European Community (TEMPUS Project S_JEP-09273-95).

References

[1] Record, M.T., Jr., Reznikoff, W.S., Craig, M.L., McQuade, K.L. and Schlax, P.J. (1996) *Escherichia coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcription initiation. In: *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology, 2nd edn. (Neidhardt, F.C., Ed. in Chief), pp. 792-821. ASM Press, Washington, D.C.

[2] Perez-Martin, J., Rojo, F. and De Lorenzo, V. (1994) Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiol. Rev. 58, 268-290.

[3] Rao, L., Ross, W., Appleman, J.A., Gaal, T., Leirmo, S., Schlax, P.J., Record, M.T., Jr. and Gourse, R.L. (1994) Factor independent activation of *rrnB* P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. J. Mol. Biol. 235, 1421-1435.

[4] Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. and Gourse, R.L. (1993) A third recognition element in bacterial

promoters: DNA binding by the α subunit of RNA polymerase. Science 262, 1407-1413.

[5] Ross, W., Aiyar, S.E., Salomon, J. and Gourse, R.L. (1998) *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. J. Bacteriol. 180, 5375-5383.

[6] Negre, D., Bonod-Bidaud, C., Oudot, C., Prost, J.F., Kolb, A., Ishihama, A., Cozzone, A.J. and Cortay, J.C. (1997) DNA flexibility of the UP element is a major determinant for transcriptional activation at the *Escherichia coli* acetate promoter. Nucleic. Acids. Res. 25, 713-718.

[7] van Ulsen, P., Hillebrand, M., Kainz, M., Collard, R., Zulianello, L., van de Putte, P., Gourse, R.L. and Goosen, N. (1997) Function of the C-terminal domain of the α subunit of *Escherichia coli* RNA polymerase in basal expression and integration host factor- mediated activation of the early promoter of bacteriophage Mu. J. Bacteriol. 179, 530-537.

[8] Estrem, S.T., Gaal, T., Ross, W. and Gourse, R.L. (1998) Identification of an UP element consensus sequence for bacterial promoters. Proc. Natl. Acad. Sci. U. S. A. 95, 9761-9766.

[9] Czarniecki, D., Noel, R.J., Jr. and Reznikoff, W.S. (1997) The -45 region of the *Escherichia coli lac* promoter: CAP-dependent and CAP-independent transcription. J. Bacteriol. 179, 423-429.

[10] Noel, R.J., Jr. and Reznikoff, W.S. (1998) CAP, the -45 region, and RNA polymerase: three partners in transcription initiation at *lacP1* in *Escherichia coli*. J. Mol. Biol. 282, 495-504.

[11] Murakami, K., Fujita, N. and Ishihama, A. (1996) Transcription factor recognition surface on the RNA polymerase α subunit is involved in contact with the DNA enhancer element. EMBO. J. 15, 4358-4367.

[12] Aslanidis, C., Schmid, K. and Schmitt, R. (1989) Nucleotide sequences and operon structure of plasmid-borne genes mediating uptake and utilization of raffinose in *Escherichia coli*. J. Bacteriol. 171, 6753-6763.

[13] Aslanidis, C., Muiznieks, I. and Schmitt, R. (1990) Successive binding of *raf* repressor to adjacent *raf* operator sites *in vitro*. Mol. Gen. Genet. 223, 297-304.

[14] Muiznieks, I. and Schmitt, R. (1994) Role of two operators in regulating the plasmid-borne *raf* operon of *Escherichia coli*. Mol. Gen. Genet. 242, 90-99.

[15] Aslanidis, C. and Schmitt, R. (1990) Regulatory elements of the raffinose operon: nucleotide sequences of operator and repressor genes. J. Bacteriol. 172, 2178-2180.

[16] Muiznieks, I., Rostoks, N. and Schmitt, R. (1999) Efficient control of *raf* gene expression by CAP and two Raf repressors that bend DNA in opposite directions. Biol. Chem. 380, 19-29.

[17] Sabourin, D. and Beckwith, J. (1975) Deletion of the *Escherichia coli crp* gene. J. Bacteriol 122, 338-340.

[18] Breul, A., Assmann, H., Golz, R., von Wilcken-Bergmann, B. and Muller-Hill, B. (1993) Mutants with substitutions for Glu171 in the catabolite activator protein (CAP) of *Escherichia coli* activate transcription from the *lac* promoter. Mol. Gen. Genet. 238, 155-160.

[19] Bullock, W.O., Fernandez, J.M. and Short, J.M. (1987) XL1-Blue: a high efficiency plasmid transforming *rec*A *Escherichia coli* strain with β-galactosidase selection. Biotechniques 5, 376-378.

[20] Sambrook, J., Fritsch, E. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Vol. 1-3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[21] Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77, 51-59.

[22] Thompson, J.F. and Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. Nucleic Acids Res. 16, 9687-9705.

[23] Schmid, K. and Schmitt, R. (1976) Raffinose metabolism in *Escherichia coli* K12. Purification and properties of a new α-galactosidase specified by a transmissible plasmid. Eur. J. Biochem. 67, 95-104.

[24] Josaitis, C.A., Gaal, T., Ross, W. and Gourse, R.L. (1990) Sequences upstream of the -35 hexamer of *rrnB* P1 affect promoter strength and upstream activation. Biochim. Biophys. Acta 1050, 307-311.

[25] Ulanovsky, L.E. and Trifonov, E.N. (1987). Estimation of wedge components in curved DNA. Nature 326, 720-722.

[26] Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. & Buc, H. (1989). Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli.* EMBO J. 8, 4289-4296.

[27] Munteanu, M.G., Vlahovicek, K., Parthasarathy, S., Simon, I. and Pongor, S. (1998) Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. Trends. Biochem. Sci. 23, 341-347.

Table 1

Strains and plasmids

Strain/Plasmid	Purpose/relevant genotype	Reference
Strains		
CA8000	α-Gal activity measurements / *crp*$^+$, *cya*$^+$	[17]
CA8306	α-Gal activity measurements / *crp*$^+$, *cya*$^-$	"
XA3D1	α-Gal activity measurements / *crp*$^-$, *cya*$^+$	[18]
XL1-Blue	General cloning procedures	[19]
Plasmids		
	wt *rafP*	
pRU1307		[14]
	Promoter deletions (Fig. 1B)	
pRU1301, pRU1302, pRU1303		[16]
pRU1305		This study
	Promoter mutations (Fig. 1B)	
pRU1307E, pRU1307M		[16]
pRU1307B, pRU1307C, pRU1340, pRU1350-1352, pRU1360-1362		This study
	Oligonucleotide substitutions (Fig. 1C)	
pRU1369, pRU1369+4, pRU1371, pRU1371+5; pRU1372		This study

Table 2

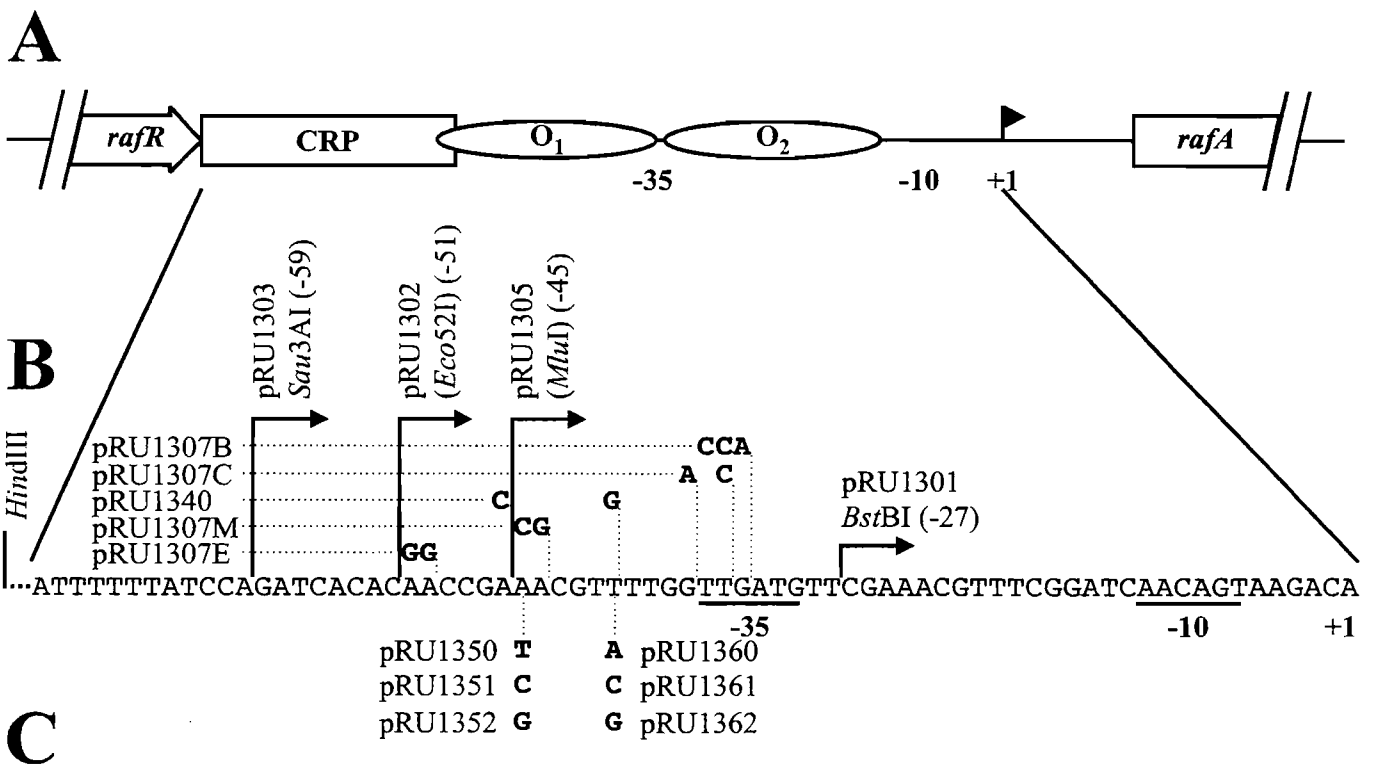
α-Gal activity of *rafP* derivatives

		Plasmid	Promoter properties	Activity
Deletion derivatives		pRU1307	wt	100
		pRU1303	upstream from –59	100
		pRU1302	upstream from –52	36
		pRU1305	upstream from –45	11
		pRU1301	upstream from –28	2
	Promoter mutants	pRU1307E	$A_{-51}A_{-50}$ to GG	55
		pRU1307M	$A_{-45}A_{-44}$ to CG	22
		pRU1307B	$T_{-35}T_{-34}G_{-33}$ to CCA	1
		pRU1307C	G_{-36} to A, T_{-34} to C	2
		pRU1340	A_{-46} to G, T_{-40} to G	16
		pRU1350	A_{-45} to T	57
		pRU1351	A_{-45} to C	49
		pRU1352	A_{-45} to G	38
		pRU1360	T_{-40} to A	37
		pRU1361	T_{-40} to C	22
		pRU1362	T_{-40} to G	18
Oligonucleotide substitutions		pRU1369	See Fig. 1. for sequence	3
		pRU1369+4	"	2
		pRU1371	"	2
		pRU1371+5	"	54
		pRU1372	"	2

Figure legends

Fig. 1. Schematic representation of the raffinose operon promoter *rafP*. A: Genetic map of the *rafR – rafA* intergenic region containing *rafP*. Localization of the protein binding sites involved in regulation of *rafP* is marked as follows: CRP – CRP binding site; O_1 and O_2 – *raf* repressor operator sites. Transcription initiation site is marked by a pennant and basic promoter elements are underlined. B: Sequence of the wt *rafP* region. Mutations introduced in the *rafP* are indicated above and below the promoter sequence along with the respective plasmid names. The deletions used to delimit boundaries of upstream element are marked by arrows along with the relevant plasmid names and restriction enzyme sites. The *Hin*dIII site from the plasmid vector polylinker region, which was used to introduce deletions, is marked on the left. C: Sequences of the oligonucleotide substitutions of the *rafP* upstream region from position –133 till -28. Four and 5 nucleotide insertions in pRU1369+4 and pRU1371+5, respectively, which were created with purpose of phasing of A/T rich region in respect to basic promoter elements, are marked by asterisks.

Fig. 2. Comparison of the UP element consensus sequence with the corresponding regions from wt *rafP* and oligonucleotide – substituted derivatives (panel A) and the correlation of sequence homology with nucleotide composition and activities of constructions (panel B). A: Nucleotide matches with the consensus sequence are marked with asterisks. B: Only the exact and W (A or T) matches were counted in the case of consensus sequence.



pRU1369

AGCTTAAAATATTTTTAAAATATTTTAAAATATTTTAAAAATATTTTAAGCAATTGACATT

-88 -35

pRU1369+4

****

AGCTTAAAATATTTTTAAAATATTTTAAAATATTTTAAAAATATTTTAAGCAATTAATTGACATT

-92 -35

pRU1371

AGCTTAAGCAATTTTCGAAAATTTTTCGAAAATTTTCGAAAAATTTTCGAAAATTGACATT

-88 -35

pRU1371+5

* ****

AGCTTAAGCAATTTTCGAAAATTTTTCGAAAATTTTCGAAAAATTTTCGAGCAATTAATTGACATT

-93 -35

pRU1372

AGCTTGGGCCCTTAAGGGCCCCTTAAGGGCCCTTAAGGGGCCTTAAGGGCCCAATTGACATT

-89 -35

A

UP element consensus [8]	59-NNAAAWWTWTTTTNNAAAANNN-38
wt *rafP* [15]	59-GATCACACAACCGAAACGTTTT-38
pRU1369 [this study]	59-ATATTTTAAAAATATTTTAAGC-38
pRU1371 [this study]	59-AAATTTTCGAAAAATTTTCGAA-38
pRU1369+4 [this study]	59-TTTAAAAATATTTTAAGCAATT-38
pRU1371+5 [this study]	59-TTCGAAAAATTTTCGAGCAATT-38

B

	pRU1307	**pRU1369**	**pRU1371**	**pRU1369+4**	**pRU1371+5**	**pRU1372**
Matches to consensus	4	6	3	10	10	4
A/T content	7	15	12	13	11	6
Activity	100	3	2	2	54	2

4

Jūrmala,
Republic of Latvia

25–29 September,
1995

ABSTRACTS
2nd Baltic Genetical Congress

TABLE OF CONTENTS

PROCEEDINGS OF THE LATVIAN ACADEMY OF SCIENCES, Section B, 1995, no. 5/6 (574/575), pp. A0–A54

CLONING OF SARCOSINE OXIDASE FROM *Arthrobacter* sp. 1-IN

R. Meškys, R. Rudomanskis
Institute of Biochemistry, Vilnius 2600, Lithuania

The heterotetrameric sarcosine oxidase (SOX) producing *Arthrobacter* sp. 1-JN strain was earlier isolated in our laboratory. The genes encoding the four subunits of SOX were cloned and expressed in *E. coli* DH5α. The genomic library was constructed by partial digestion of chromosomal DNA from *Arthrobacter* sp. 1-IN with PstI and ligation into pUC19 linearized with PstI. The positive clone (DH5α/pROX1) that generates hydrogen peroxide in a sarcosine-dependent reaction was isolated by using indicator plates (containing sarcosine, horse radish peroxidase and o-dianisidine) to screen the genomic library. The fragment of DNA inserted into pROX1 was approximately 12.5 kb. The genes that confer SOX activity were further localized by deletion analysis. It was found that an, about 7 kb, subfragment contained the SOX operon. Crude cell lysate from *E. coli* DH5α/pROX1 cells grown on LB media was prepared and subjected to native gel electrophoresis. SOX activity was detected as a violet band when the gel was stained using nitro blue tetrazolium as a redox dye. The band from DH5α/pROX1 lysate comigrated with authentic SOX purified from *Arthrobacter* sp. 1-IN. The results suggested that DH5α/pROX1 cells express a sarcosine oxidase similar to the enzyme isolated from *Arthrobacter* sp. 1-IN and prompted further studies to characterize the recombinant enzyme.

POSTTRANSCRIPTIONAL REGULATION OF GENE EXPRESSION IN BACTERIOPHAGE T4

R. Nivinskas, N. Malys, A. Zajančkauskaitė, V. Klausa, R. Vaiškunaitė
Institute of Biochemistry, Vilnius 2600, Lithuania

Posttranscriptional mechanisms modulate expression of T4 genes in a variety of ways which include the regulation of translational initiation via potential RNA secondary structures. It is well established that translational initiation can be prevented by sequestering the ribosome-binding site in double-stranded RNA. T4 uses this mechanism to prevent or reduce the early translation of at least three late genes, *soc*, *e*, and 49. On the other hand, in the case of T4 gene 38, there exists a secondary structure that brings the Shine and Dalgarno sequence and initiation codon into close proximity. Such structures that enhance or facilitate initiation on T4 mRNAs are rare. Earlier work of our laboratory suggested that such a secondary structure may exist in the translational initiation region of T4 gene 25. Now we demonstrate that the expression of gene 25 *in vivo* is regulated via the proposed RNA secondary structure. Our data also indicate the existence of RNA secondary structure just upstream of the gene 30.3', which is completely embedded within gene 30.3, but in a different reading frame. We suppose that this structure may be important for optimal gene 30.3' expression. We also studied the regulation of expression of gene 26', which starts with the rarest initiation codon, AUU, and in frame overlaps with 3'-end of gene 26.

LONG DISTANCE REGULATION THROUGH REPLICON TOPOLOGY IN PLASMID GENE EXPRESSION

I. Muižnieks, N. Matjuškova, N. Rostoks, G. Makarenkova, G. Dumpe
Faculty of Biology, University of Latvia, Rīga LV-1586, Latvia

The gene expression in bacteria is regulated primarily through the efficiency of the transcription of relevant mRNAs. The data accumulated in our laboratory over the years of experience with various recombinant *Escherichia coli* strains, indicate that the efficiency of gene expression can be influenced not only by the adjacent sequence elements involved in transcription regulation, e.g. — promoters, terminators, etc. The transcription process that takes place within a relatively small, supercoiled circular replicon is influenced also by the remote sequence elements which modulate the topology constraints of the plasmid. We demonstrate the phenomenon of long range regulation on the following model systems.

1. Recombinant plasmid maintenance stability and copy number per cell is enhanced by insertion of IS elements into the *tet*-gene domain, which is opposite the RNAII transcription regulating elements on the circular map. The cloning of IS5 DNA fragment, which carries sharp sequence-directed bent, into the *tet*-gene 5' part, is sufficient to produce the effect of plasmid maintenance stabilisation.
2. The expression of the *raf*R gene promoter from the bacterial raffinose degradation operon is modulated by the CAP protein binding sites located up- and downstream the α-galactosidase reporter gene. The CAP binding on these sites can not influence the test promoter activity directly. The CAP-induced effects can be mimicked through replacement of the protein binding sites by intrinsically bent DNA fragments.
3. A cloned human α-interferon gene, which was perfectly silent under the control of strong bacterial promoter, was induced for high-level expression by the deletion of the vector plasmids part more than 200 b.p. downstream the 3' end of the cloned gene. The deleted DNA is a portion of *tet*-gene, which is known to participate in regulation of plasmid supercoiling properties.

SPECIFIC GENOMIC FINGERPRINTS OF SOIL PSEUDOMONADS GENERATED WITH REPETITIVE SEQUENCES AND PCR

A. Nurk, M. Peters
Institute of Molecular and Cell Biology, Tartu University, Tartu EE2400, Estonia

The understanding of microbial gene transfer, including how bacteria acquire and disseminate genes in natural environments, will provide data on the role of horizontal transfer in evolution. This prospect has increased interest in the determination of the fate and persistence of bacteria injected into aquifers as well as the detection of genetic interactions between bacteria introduced into aquifer (pro soil) and the indigenous microflora. Many known strain identification methods, which are usable for cells with defined life conditions in the laboratory, cannot be applied for cells in the long term after release. The genetic polymorphism increases much more quickly after release of bacteria than in laboratory conditions. During the last year, we succeeded in developing a specific chromosomal fingerprinting method for bacteria on the basis of rep-PCR. This new method is bacterial strain-specific and, using this approach, it is possible to exactly answer: a) whether the reisolated bacterium is the same as that released (despite differences in phenotypic and plasmid-mediated genotypic characters); b) which reisolated bacteria could be siblings (Ka et al., 1994. *Appl. Environ. Microbiol.* **60** 1106-1115). Our experience supports the recent finding that the rep-PCR technique appears to be a rapid, simple, and reproducible method to identify specific strains (Louws et al., 1994. *Appl. Environ. Microbiol.* **60** 2286-2295). Families of repetitive DNA sequences are dispersed throughout the genome of diverse bacterial species. These elements have the potential to form stem-loop structures and may play an iportant role in the organization of the bacterial genome. Genome organisation is thought to be shaped by evolution and appears sufficiently conservative, but at the same time also strongly strain-specific character. The advantage of rep-PCR before random amplifed polymorphic DNA (RAPD) analysis lies in the length of the primers used and the corresponding PCR conditions. The extended primers for rep-PCR allow the use of more stringent PCR conditions, which in turn may reduce experimental variation and PCR artifacts.

5

ISSN 0235—7232

LIETUVOS MOKSLŲ AKADEMIJA
ЛИТОВСКАЯ АКАДЕМИЯ НАУК
LITHUANIAN ACADEMY OF SCIENCES

EKSPERIMENTINĖ BIOLOGIJA
ЭКСПЕРИМЕНТАЛЬНАЯ БИОЛОГИЯ
EXPERIMENTAL BIOLOGY

1992 Nr. 3—4

PROCEEDINGS OF THE 1ST GENETICAL CONGRESS OF THE BALTIC STATES (ESTONIA, LATVIA, LITHUANIA)

Vilnius "Academia" 1993

CLONING OF HUMAN GENES REGULATING THE IMMUNE RESPONSE

G. Makarenkova, E. Jankevics[1]

Latvian University, [1]Institute of Molecular Biology of the Latvian Academy of Sciences, Riga, Latvia

Different methods for constructing genomic library in bacteriophage λ vector Charon 4A and stability of the recombinant DNA during reproduction in *E. coli* cells were estimated. The maximal number of recombinant phages was obtained by the in vitro packaging method involving freshly freeze - thaw lysate from BHB 2688 and sonicated extract from BHB 2690 preparation (N. Sternberg) .

A Charon 4A human gene library consisting of 3.6×10^6 independently derived clones was prepared. Recombinant phage clones carried about 15-20 kb-size DNA fragments from the partially EcoR1 - digested human blood leukocytes DNA. Investigation of structural stability of both vector phage Charon 4A and its derivatives carrying DNA inserts of bovine leukemia and mouse sarcoma viruses revealed considerable heterogenity of phage populations after 1-2 cycles of reproduction in *E. coli* strains K 802, K 803 and QD 5003 judged by appearance of new restriction sites in inserted DNA. Taking into account the instability of recombinant phage populations during propagation, only unamplified genomic library was used for isolation of specific sequences.

Human genomic library was screened by the in situ bacteriophage plaques hybridization techniques using separately nicktranslated human interleukin 2 (IL-2) cDNA, human αN-interferon (IFN-α) and mouse immunoglobulin G heavy chain constant region (CH) probes. One of the recombinant bacteriophages containing a whole human IL-2 gene and its 5'-flanking region as well as 3 other clones containing IFN-α5, IFN-α8 and IFN-α13 genes with extensive 3'- and 5'-flanking regions were isolated and physical maps of gene-containing DNA segments were obtained by digestion with restiction enzymes. Seven plaques which gave a positive signals with mouse CH DNA probe in relaxed strigency hybridization conditions were isolated and are under identification.

IL-2 gene fragments and IFN-α genes were further subcloned into plasmid vectors, their nucleotide sequences were determined and compared with the known genes. The coding parts of cloned genes contain only few silent nucleotide exchanges in comparison with the previously published sequences, while marked differences could be detected in regulatory 5'-flanking regions. Those are promising for detection of DNA elements controlling gene expression neihbouring and distant from the coding part as well as for the studies of the interactions between such elements. Cloning of the IFN-α genes can be used for constructing of new bacterial strains capable to produce human interferon.

SPONTANEOUS INSERTION OF TRANSPOSON Tn 1000 IN raf A GENE OF PLASMID pRU 1307

N. Matyushkova, N. Rostoks

Latvian University, Riga, Latvia

Recombinant plasmid pRU 1307 (constructed by Dr. I. Muiznieks) is a pUC 8 derivative which carries raf A gene with its own promoter from *E. coli* plasmid-born raf-operon

```
                10          20          30          A 40
5' CACGGTTCAT  TCTCCTTCCT  TCTGATTTTT  TATCCAGATC
               50          60          70          80
   ACACAACCGA  AACGTTTTGG  TTGATGTTCG  AAACGTTTCG
               90          100         110         120
   GATCAACAGT  AAGACATACC  TGAAAGCGGA  GATGTCTTAT
              130         140         150         160
   GATTTCAAAG  TAGGGGTTTG  AGGGCCAATG  GAACGAAAAC
              170         180         190         200
   GTACGTTAAG  GAGATAATTC  GTTGTTTATA  TTTAAATTTA
              210         220         230         240
   GAGCTCTCAG  TTCCCCTTTT  AAAATATCCT  CTGGCAACGT
              250         260         270         280
   GAATGTATAA  GGCCCAACAT  ATTGATATCC  CGTGCATCAG
          290         300         310
   GGGAGATAGC  CGAGCGATAT  CTTCATCTAT  A          3'
```

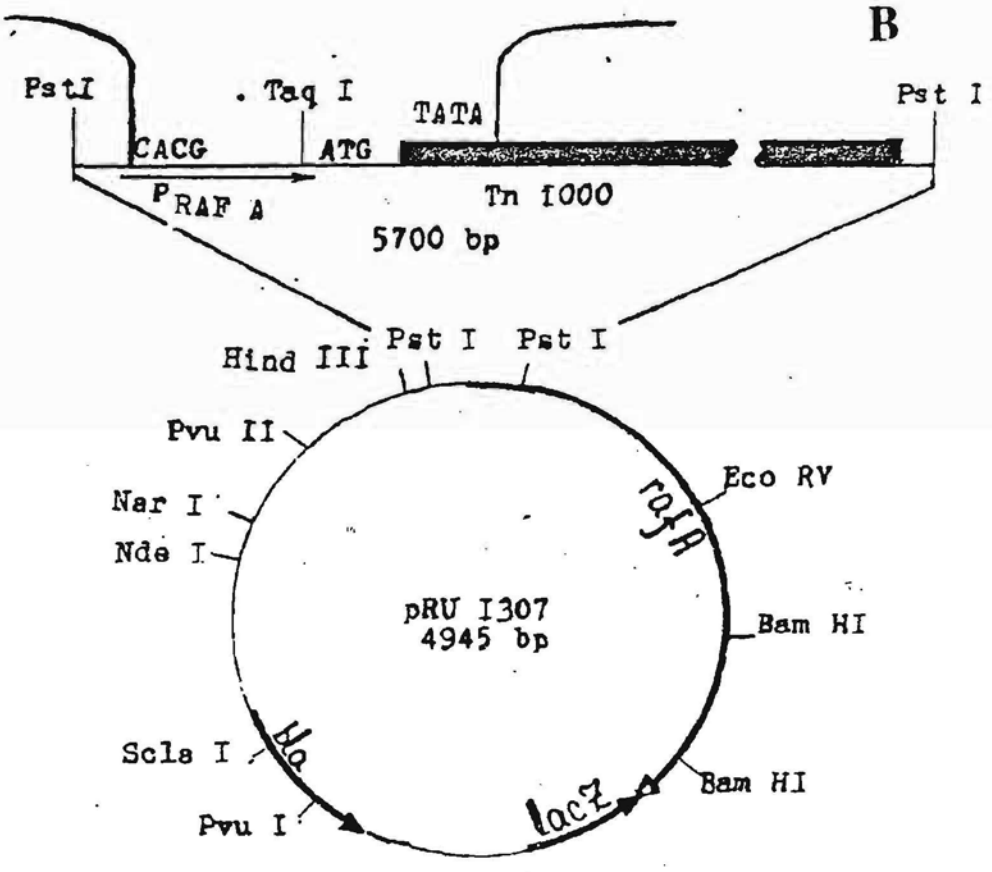


Fig. 1. The sequence of Tn 1000:: pRU 1307 junction (*A*) and schematic location of the transposon in the structure of plasmid pRU 1307 (*B*). The first translated ATG codon of raf A gene is boxed. The begining of inserted Tn 1000 sequence is marked with an arrow

(Aslanidis et.al., 1989). *E.coli* strains containing pRU 1307 are ampicillin-resistant (Ap^R) and express high levels of α- galactosidase (αgal$^+$).

Studies of pRU 1307 maintenance in *E.coli* TG I liquid batch cultures readily revealed population heterogenity, where alongside with the cells, which have lost plasmid ($Ap^S\alpha Gal^-$ -phenotype), $Ap^R\alpha Gal^-$-segregants could be detected. The accumulation of $Ap^R\alpha Gal^-$-cells was faciliated in cultures kept mainly in late-log-stationary phase of growth at low initial concentrations of Ap (25 mkg/ml).

Electrophoretic analysis of plasmid DNA from $Ap^R\alpha Gal^-$- segregants showed a replicon ca. twice as big as pRU 1307. Further restriction and sequence analysis allowed us to identify the inserted DNA in pRU 1307 as the transposon Tn 1000 (known also as γ δ) in its direct (γδ) orientation (Fig.1). The junction point of plasmid and Tn 1000 integration was sequenced in several independent isolates and appeared to be always the same. So the 5' end of raf A sequence is a Tn 1000 integrations hotspot.

pRU 1307::Tn 1000 was maintained in TG I strain in higher copy numbers and significantly more stable than the parent pRU 1307. The elimination of metabolic burden caused by α-Gal-protein synthesis appears to be crucial for improving the maintenance characteristics of the plasmid, not looking to the expenses of DNA-length increment.