ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

DESIGN,SYNTHESIS AND CHARACTERIZATION OF DNA-BINDING PEPTIDES BASED ON THE HELIX-TURN-HELIX MOTIF

Thesis submitted for the degree
"Magister Philosophiae"

CANDIDATE

Piergiorgio Percipalle

SUPERVISOR

Prof. Sandor Pongor

October 1993

TRIESTE

SISSA - SCUOLA
INTERNAZIONALE
SUPERIORE
DI STUDI AVANZATI
TRIESTE
Strada Costiera 11

INDEX

Abstract

The helix-turn-helix (HTH) motif is common to many dimeric DNA-binding proteins, including phage repressors and bacterial regulatory proteins. The HTH motif is responsible for binding specificity, while the dimeric architecture of regulatory and repressor proteins ensures the correct spacing and orientation of two such motifs so as to recognize a cognate, symmetric DNA sequence in the operator of the regulated genes.

The study of the DNA-binding N-terminal domain of the λ 434 repressor protein was undertaken using computer modelling techniques based on homology alignments with other phage repressors, so as to determine those structural factors that are most important for both the specificity and affinity of the protein for its operator. It was found that apart from the HTH motif, a loop (L) that contains basic residues stabilizing binding to DNA and hydrophobic residues stabilizing the HTH (H2TH3 in the 5-helix subunit) motif, is important. Also, various means to mimic dimerization were evaluated, as this is known to be essential for binding.

This information was used to design a number of peptides which, in the minimum number of residues, might include essential binding features. The peptides were then synthesized and their binding activity with DNA studied using band-shift assays, footprinting and UV crosslinking techniques against oligonucleotides and plasmids containing either the operator sites of the λ phage 434 or the phage 16-3 of the nitrogen fixing bacterium Rhizobium meliloti which share mutually recognized operator sequences with common half sites.

CD studies indicate that the peptide is capable of folding into an a-helical conformation in an organic environment. Band-shift studies indicate that binding to the cognate DNA sequence is specific but of low affinity. A degree of cooperativity is shown. UV crosslinking studies confirm that the HTHLDIM

peptide (obtained by linking two H2TH3L units via a branched lysine) binds specifically to its operator site. Dimerization is, as expected, essential for binding, as the monomeric HTHL peptide does not bind to DNA. The stabilizing helix is also important as a shorter HLDIM (containing only the recognition helix) does not appear to bind to DNA.

This study shows that a combination of simple modelling procedures and automated peptide synthesis, allows the construction of relatively simple peptides with potential biological activity, and which may find useful applications as DNA probes, while bringing an increased understanding of how helix-turn-helix based DNA-binding proteins function.

This thesis work provides a general review of the main classes of DNA-binding proteins in chapter 1.0, through examples of known x-ray structures. The helix-turn-helix motif is described in details. Chapters 2.0 and 3.0 respectively present the results obtained by modelling and biochemical experiments. All the experimental procedures utilized are reported in chapter 4.0. The general conclusions obtained from the experimental data and discussion are included in chapter 5.0.

1.0 INTRODUCTION

1.1. DNA binding proteins

There are various ways in which a molecule can bind to DNA, including electrostatic binding to phosphates, intercalation and partial intercalation between DNA bases, and covalent binding occurring in the major and minor grooves of DNA.

Protein-DNA interactions are involved in many of the fundamental processes that occur in cells, including packaging, replication, recombination, restriction and transcription (1). It is important to understand the nature of these interactions, as they are intimately involved in the control of gene expression, cell division and differentiation. Of the proteins that interact with DNA, restriction enzymes and transcription factors need to discriminate DNA binding sites with specific sequences for correct functioning. Other proteins, such as nucleases and chromosomal proteins, must by definition be able to bind many different and unrelated DNA sequences. Such proteins posess a certain selectivity in the binding process even though they do not bind with equal affinity to different DNA sequences. The resulting protein-DNA complexes are often well defined as for example that of DNAseI, inducing a local bend in DNA of ca. 20-30 degrees (2) or that of the histone octamer, involved in DNA packaging and able to bind chromatin by wrapping 145 base-pairs of DNA through superhelical scaffolds (3).

Most sequence specific proteins recognize and bind cognate DNA through well defined structural domains which make sequence specific contacts with the DNA bases in the major groove. These contacts are via the H-bonding interactions of side chains, and to a lesser extent via main chain atoms from residues in the protein, with atoms in DNA bases, and are stabilized by hydrophobic and electrostatic

interactions. In contrast, structural studies of the HU class of proteins, DNAseI (3) and certain DNA binding drugs suggest that the minor groove and phosphate backbone may be primary recognition features for non sequence-specific DNA ligands (4,5,6). Most of the regulatory DNA binding proteins tend to bind the DNA major groove also because it is a relatively wide site of interaction and thus base pairs are more easily accessible.

At present there are at least four structural classes of major groove DNA binding proteins that have been characterized in some structural detail: (i) the Helix-turn-helix (HTH) proteins, (ii) the Zinc-finger proteins, (iii) the Leucine zipper proteins and (iv) the Beta proteins. A list of some well known DNA-binding proteins is shown in Table 1(1).

The motifs involved in DNA-binding can be highly conserved in sequence or secondary structure. However it appears that they are by themselves insufficient to impart DNA-binding properties to a protein as it has been shown that in some cases the DNA-binding domain can extend well beyond such structural motifs. In the case of the λCI and trp repressor proteins, for example, flexible terminal extensions are present, while a general role of the electrostatic complementarity is important for CAP (7). Furthermore, the DNA molecule often undergoes conformational changes on binding. Thus EcoRI restriction enzyme causes partial unwinding of the DNA double helix, while the CAP protein and the trp and 434 repressor proteins cause substantial bending and kinking of the DNA molecule (9). It is interesting to note that those proteins which cause conformational changes to the DNA molecule do not make use of flexible extensions to create further contacts, whereas proteins that appear to have little effect on DNA conformation in contrast do posess such flexible extensions so as to increase the interaction surface (1).

In this thesis the α-helix-turn-α-helix structural motif is dealt with in detail, and section 1.2 describes several regulatory proteins that use this motif to provide

TABLE 1. Classes of DNA-binding proteins

Class	Example	Subunit (KDa)	Allosteric Effector	Target DNA
HTH	λCI	α2(26)	No	TATCACCGCCAGTGGTA
	λcro	α2(7.5)	No	TATCACCGCCAGTGGTA
	P434	α2(7.5)	No	TACAAGAAAGTTTGTT
	P434cro	α2(7.5)	No	TACAAGAAAGTTTGTT
	trp	α2(12.5)	L-Trp	CGAACTAGTTAACTAGTACG
	CAP	α2(22.5)	c-AMP	TGTGANNNNNNTCACT
	lac	α4(35)	allo-Lac	GAATTGTGAGCGCTCACAATT
	ant	α (7)	No	ATTA
	eng	α (7)	No	TAAT
	FIS	α2(10)	No	Not Applicable
Zn-finger (C2H2)	mKR2	α (3)	No	Not Known
	ADR1	α (3)	No	Not Applicable
	SW15	α (3)	No	CCAGCATGCTATAATGC
	HEB	α (3)	No	GGGGAATCCCC
	Xfin	α (3)	No	Not Known
Zn-finger (C4)	glc	α2(7.6)	glucorticoid	AGAACANNNTGTTCT
	est	α2(7.6)	oestrogen	AGGTCANNNTGACCT
L-zipper	fos	α2(8)	No	GTGACTCAG
	jun	α2(8)	No	GTGACTCAG
	GCN4	α2(8)	No	ATGACTCTT
β-Sheet	arc	α2(8.5)	No	CATGATAGAAGCACTCTACTA
	metJ	α2(12)	S-Adenosyl methionine	AGACGTCGT
	HU	α2(9.5)	No	Not Applicable
Homeobox	MATα1	α	No	ATTA
Other	TFIID	α((	No	TATA
	EcoRI	α2(31)	No	GAATTC
	EcoRV	α2(28.5)	No	GATATC
	DNAseI	α (30.5)	No	Not Applicable
	DNA pol I	α (103)	No	Not Applicable

sequence-specific binding to DNA. The *in vivo* mechanism of action of such proteins is taken into account in order to pinpoint structural and functional features deriving from the HTH motif that are in common among them. It is however, useful first to briefly consider other DNA-binding protein systems:

1) Zinc Fingers. Transcription factors are proteins that bind DNA and regulate the transcription of mRNA and therefore gene expression. They are known to contain a DNA-binding domain through which they bind in a sequence specific manner (10). Such highly conserved DNA-binding and dimerization domains have been characterized both by structural studies and by sequence comparisons and each structural motif may be considered to represent the evolution of different solutions to the design of proteins that need to be fully compatible to their target DNA, both from a structural and chemical stand point (11, 12).

GAL4 is a yeast transcriptional activator whose DNA-binding specificity and ability to dimerize is retained in the 65 residue long amino-terminal domain of the protein (881 aminoacids). Structural studies, both by x-ray crystallography and NMR (13, 14, 15) show that dimerization is affected by the carboxy end of the 65 residue terminus, in particular by a two-stranded coiled-coil structure formed by weak hydrophobic interactions between two aphipathic α-helices from each monomer. The core of the DNA-binding domain (residues 8-40 in the N-terminal domain) forms a compact globular structure folded around two Zn ions which are tetrahedrically coordinated to six cystein residues in a binuclear-type cluster. The metal-binding domain probably mediates the sequence-specific contacts with the bases via two α-helices, one per each monomeric subunit. The dimer binds a 19bp DNA sequence by inserting side chains from residues in these appropriately positioned α-helices into the major groove on opposite faces of DNA (Fig. 1).

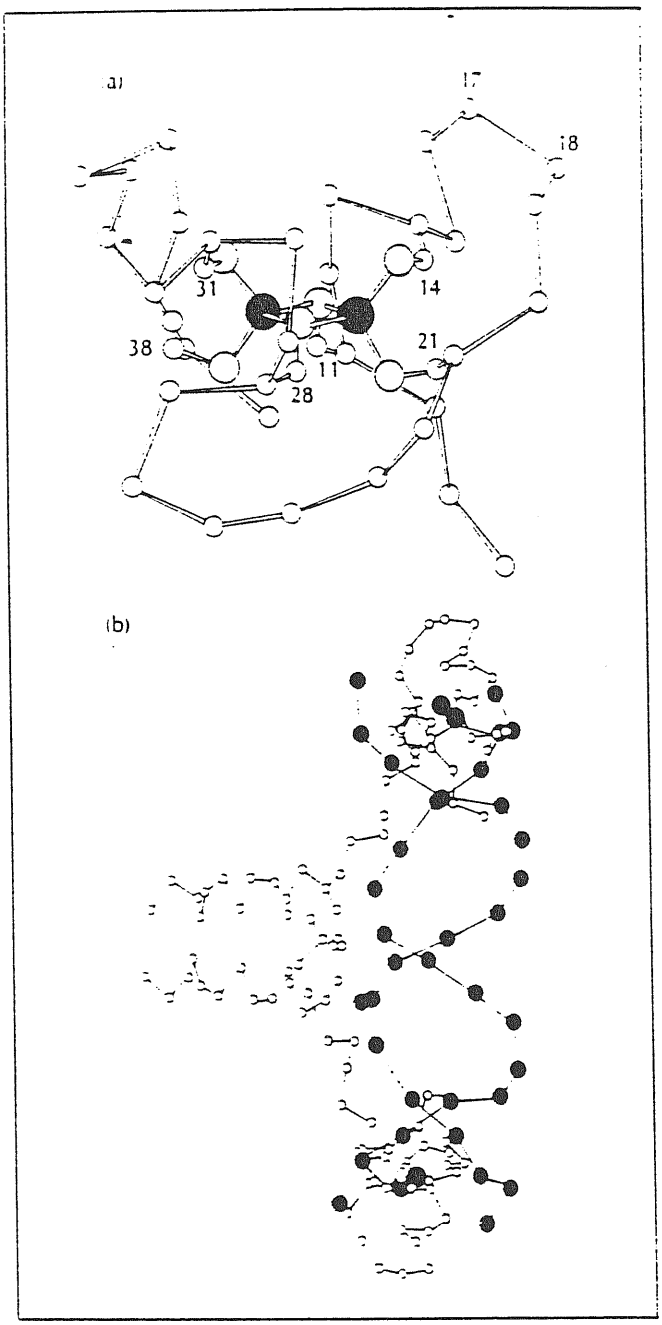


Figure 1. Structure of the GAL4/DNA complex. (A) The metal binding domain showing backbone carbons as white spheres, cystein sulphurs as shaded spheres and zinc atoms as black spheres. (B) View of the entire complex

GAL4 is thus characterized by a binuclear cluster and represents a new type of zinc-finger with a generalized Cys-X-X-Cys sequence (X may be any aminoacid) at the beginning of the DNA-binding α-helices. This is also observed in the case of the steroid receptor class of zinc-fingers. More generally, zinc-finger containing DNA-binding proteins have been divided into two main classes according to the aminoacid residues that are involved in metal binding; i.e., (i) Cys2His2 zinc fingers and (ii) Cys3His domains.

TFIIIA is the first member discovered for the Cys2His2 class of zinc finger proteins. It has been shown that it contains nine zinc fingers but it is still not known precisely how it interacts with DNA, considering that it is probably able to bind both DNA and RNA with high specificity and affinity. The role of each zinc finger (16, 17) has been better elucidated for Zif268, for which the Zif268-DNA complex crystal structure has been determined (18), and that of SP1 (16, 19) in which case the middle zinc finger domain out of three is involved in interacting with the preferred DNA-binding site.

Retroviral nucleocapsid proteins, on the other hand, belong to the Cis3His class of zinc finger domains. In fact it has been shown that large quantities of Zn copurify with retroviruses such as HIV-I and HTLV-I (20). Extensive use of synthetic peptides coupled with x-ray absorption fine structure studies (21) have shown that the sequence Cys-X-X-Cys-X-X-X-X-His-X-X-X-X-Cys is indeed responsible for binding Zn (II) ions in the case of the HIV-I nucleocapsid protein.

2) TATA-box binding protein The TFIID TATA-box binding protein is one of the factors that plays a role in initiating transcription of mRNA in eukaryotes. In particular it is a site-specific DNA-binding protein that is required for assembly of the pre-initiation complex and moreover interacts with a set of proteins called TATA-associated factors (18, 19). The structure of a 200 residue long polypeptide

corresponding to its carboxy-terminal domain has recently been determined by x-ray crystallography (21) and shown to be built up from two 88 residue long domains related to each other by an intramolecular pseudo-dyad. Each domain is made up of five β-strands (S1, S2, S3, S4, S5) and two α-helices (H1, H2) which are arranged in the order S1-H1-S2-S3-S4-S5-H2. A linker connects the C-terminal group of H2 with the N terminus S1′ on the other monomer (as indicated by the "prime" index). Interestingly, in such a saddle-like structure H2 and H2′ are facing DNA in a concave-type fashion whereas H1 and H1′ are respectively perpendicular to H2 and H2′. TFIID is a rare example of a transcription factor that binds DNA in a sequence-specific manner but interacts with the minor groove (22, 23). More insights on the residues involved in direct binding to the TATA-box will be provided by the crystal structure of TFIID with its cognate DNA.

3) Leucine zippers. The yeast transcriptional activator GCN4 contains a basic region leucine-zipper (bZIP), a conserved element involved in dimerization and DNA-binding which has been found in a number of eukaryotic transcription factors. The recently determined 2.9 Å resolution crystal structure (24) of the bZIP with its cognate DNA AP1 site, as well as NMR structures in solution (27) show that the protein binds DNA as a Y shaped dimer that recognizes binding sites on DNA of directly abutted, dyad-symmetric half sites (Fig. 2). The stem of such a Y-shaped dimer is a coiled pair of amphipathic alpha helices adhering together through interchain hydrophobic Leu-Leu interactions occurring every seven repeats i.e. the leucine zipper. Theoretical observations coupled to experimental evidence are consistent with the alpha helical arrangement of the leucine zipper. In fact (i) no alpha helix destabilizing residues (Pro and Gly) are present in the leucine zipper and (ii) there is a high occurrence of oppositely charged residues configured three or four aminoacids apart, thus stabilizing the helix. From the experimental standpoint,

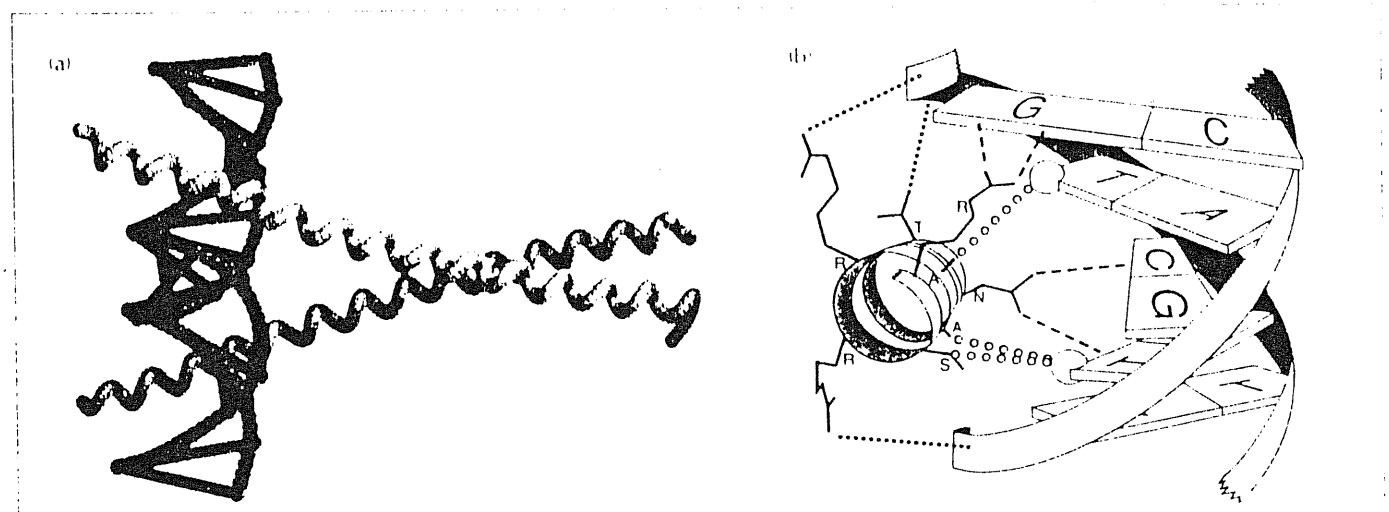


Figure 2. Structure of the GCN4/DNA complex. (A) Cα backbone structure of GCN4 bound to DNA. (B) Contacts formed in a half-site between residues in a GCN4 monomer basic helix and DNA bases and backbone. Dashed lines indicates H-bonding to bases, dotted lines show H-bonds to phosphate groups, open circles indicate VanderWaals interactions (21b).

disulfide crosslinking experiments (28), spectroscopic studies (29) and mutational studies substantiate the hydrophobic spine model punctuated every seven residues by a Leu side chain, where alpha helices adhere to one another in parallel forming a coiled coil. The bifurcating arms are a linked set of DNA contact surfaces in the form of α-helices, each arm being positioned along each half of the recognition site in opposite directions (see Fig. 2) i.e. the basic motifs. The bifurcation point is located on the rotational axis of the dimer and it contacts the major groove at the very centre of the dyad symmetric binding site. In vitro binding studies with chimeric proteins (30) demonstrate that the region directly responsible for DNA binding is immediately N-terminal to the leucine-zipper and is rich in basic residues. The GCN4 basic helices lie nearly parallel to the major groove, one on each side of the DNA duplex and this mode of binding thus supports the "induced helical fork" model (31). An alternative model, the "scissors grip", predicts a kinking of the α-helices and pronounced bending of cognate DNA (32). Sequence alignment of proteins containing the leucine zipper basic motif domain show a high level of conservation of this motif (33).

In vivo, GCN4 is functional only when it forms a homodimer. On the other hand other regulatory proteins that carry out their roles through the typical bZIP domain are functional only upon heterodimer formation, one example given by the interaction between Fos and Jun which produce a functional dimer only when they are bound to each other. Direct interaction of Fos and Jun nuclear oncoproteins has been shown to be via the leucine zipper domain. Sassone-Corsi et al. (34) report that the Fos protein directly modulates Jun function in this manner. The Fos leucine zipper domain is necessary for the DNA binding of the heterodimer whereas a distinct domain, localized in the C-terminal region of the Fos protein, is responsible for transcriptional regulation. This heterodimeric model is also supported by the evidence that the Fos-Jun complex binds the AP-1 site with an approximately 30 times greater affinity than Jun alone, primarily because of stabilization of the protein-

DNA interaction. A mutagenesis analysis (35) shows that Fos and Jun do indeed dimerize via their respective leucine zipper domains and the association appears to occur in a parallel rather than an antiparallel configuration.

4) Beta Proteins. An entirely new direction in DNA binding proteins has been recently provided by the transcriptional activator E2 from bovine papilloma virus. Its crystal structure has been determined to a resolution of 1.7Å (36) and it shows highly conserved features among analogous proteins in papilloma viruses. Like the vast majority of transcription factors it utilizes α-helices to provide the sequence-specific binding to DNA, but it presents these helices to the cognate DNA using a novel protein scaffold (Fig. 3). This scaffold is built up from eight β-strands which pack together forming a typical β-barrel type of a structure. Each half barrel is made of four strands, in which strand 1 and strand 2 are connected by a long α-helix whereas a short α-helix followed by a β-strand makes crossover connections between strands 3 and 5. Intersubunit hydrogen bonds and hydrophobic interactions contribute to stabilizing the entire β-barrel which can be disrupted only by means of denaturation. The longest α-helix in each monomer provides the direct specific interactions with the cognate palindromic DNA sequence. These contacts are facilitated by a slight bending of the DNA molecule which wraps around the β-barrel thus allowing penetration of side chains from residues in the two binding α-helices and at the same time causing a narrowing of both major and minor groove in the concave surface.

5) Homeobox binding proteins. Another example of a structural motif used by transcriptional activators in binding to DNA is given by the yeast MATα2 homeodomain, which is a DNA binding unit made up of three α-helices (ca. 60 residues, including the recognition helix) and an amino terminal arm. Such motifs are found in eukaryotic organisms only but they are structurally and functionally related

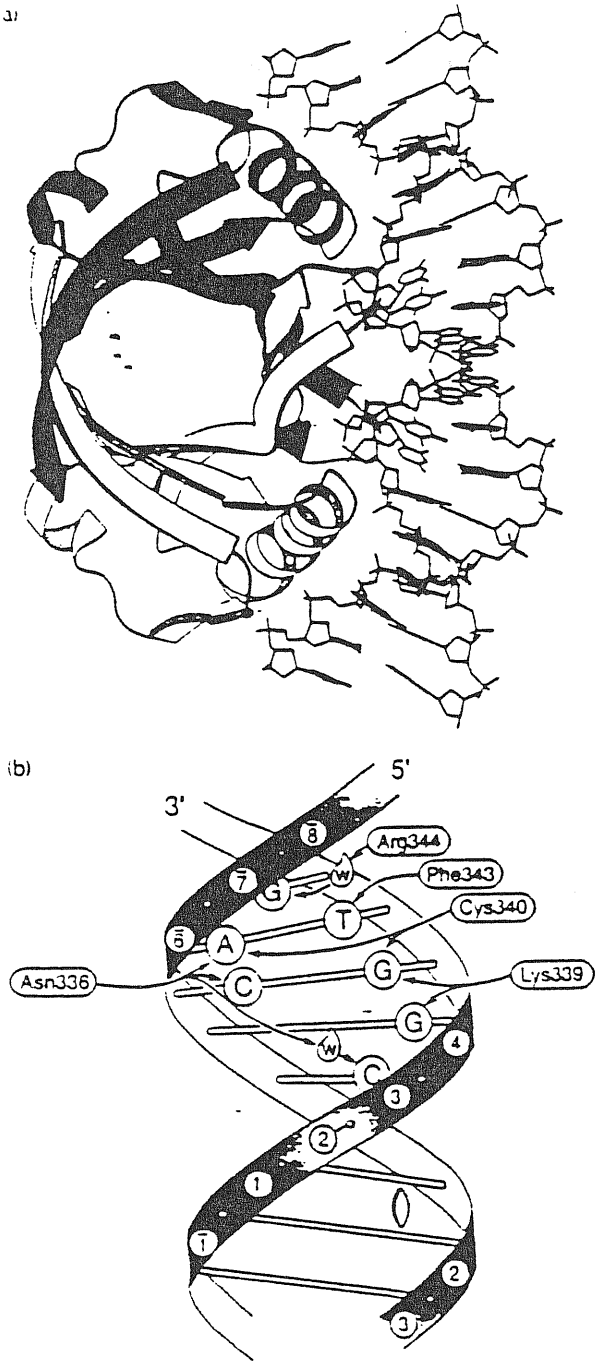


Figure 3. Structure of the Papilloma virus E2 transcription factor bound to DNA. (A) View down the axis of the b-barrel, perpendicular to the dyad of the complex, (B) Schematic diagram of contacts formed between E2 residue side chains and DNA base-pairs. Water molecules are represented as droplets labelled W

to the α-helix-turn-α-helix motif (see section 1.2) exclusively found in bacterial proteins. Studies on the Antennapedia homeodomain (37) show that the long α-helix 3 is used for direct binding to the major groove of the cognate DNA sequence by accomodating it into the major groove whereas the long, random-like, N-terminal arm wraps around the minor groove. The determination of the crystal structure of MATα2 at 2.7Å resolution (38) has shown that four side chains in the homeodomain, three of them belonging to residues in the recognition helix, provide specific contacts with DNA bases, whereas some eight additional side chains form contacts with the DNA backbone. Sequence comparison of MATα2 with the Drosophila engrailed homeodomain, and structural comparisons of the respective DNA complexes determined by x-ray crystallography (39), show a high level of homology (rating up to 72%) and a strong structural similarity since the respective recognition helices are spatially positioned in an almost identical manner.

The five examples briefly described above show that transcription factors use completely different structural motifs in order to present a set of side chains to interact with the DNA bases and the sugar-phosphate backbone of DNA. It is important to underline that in order to achieve a full understanding of the protein-DNA recognition, it is necessary to solve the structure of a certain number of protein-DNA complexes for each protein, also including non-specific complexes. There are in fact examples of proteins that are able to make both specific and non-specific interactions with the DNA molecule, as is the case of the glucocorticoid receptor-DNA complex.

In this respect, NMR spectroscopy is an extremely powerful tool for elucidating the structure of the DNA binding protein, and is also extensively used in determining the structure of the protein-DNA complex. A reason for which NMR is becoming a more and more widely spread technique in investigating the structural features of DNA-binding proteins is essentially due to the fact that most gene

regulatory proteins have a modular architecture. This means that they are made of distinct domains which take up different functions such as DNA binding, dimerization, ligand binding and so forth. In most cases these retain their biological activity even when they are chopped off the rest of the protein, facilitating NMR investigation. For instance, in the case of DNA binding domains, which are usually rather small, one can carry out NMR investigations to obtain information on most stable structures of the given polypeptide in solution. An important point is that experimental data from NMR may resemble the in vivo features of the protein more than that coming from x-ray crystallography of a rigid structure. However, NMR and crystallographic studies mostly complement each other (40).

For several families of DNA binding proteins the first information came from NMR studies, in particular for the zinc finger DNA binding proteins such as TFIIIA (41) and ADR1 (42). The crystallographic study of the Zif268-DNA complex extended the NMR investigations supporting the theory according to which zinc fingers take part in DNA recognition. Similarly, NMR analyses have been carried out on the homeodomain found in several regulatory proteins and on the complex with the target DNA. (40). NMR structures have been determined also for the glucocorticoid receptor (43) and estrogen receptor (44).

In conclusion the process of protein-DNA binding is extremely complex and it is clear from the copious literature on the subject, that it requires a multidisciplinary approach for a full understanding of its nature. The protein-DNA complex should not be described through a rigid "lock and key" model, but all the conformational changes that occur both on the DNA side (eg. protein-induced DNA bending and kinking) and on the protein side (eg. transitions from flexible random coils to highly organized structures) upon reciprocal binding have to be taken into account. In this regard biochemical techniques can be coupled to physico-chemical ones, such as CD and NMR so as to gather information not only on structure, affinity

and specificity of binding, but also on the dynamic behaviour of the biomolecules involved in binding. This is particularly important when an attempt is made at the design of de novo and engineered of DNA-binding proteins and peptides.

In the case of sequence-specific DNA-binding proteins i.e. those proteins with mainly regulatory roles, dimerization is a characteristic property. Basically one could highlight two reasons for which they work as dimers. The first is related to the affinity of interacting molecules and the free energy of binding. The two parameters are exponentially related to each other and thus dimer formation doubles the protein contact area consequently squaring the affinity constant. The second comes from the dyad simmetry found in the DNA sugar phosphate backbone at many recognition sites. This demands that two contact surfaces be assembled in the protein so that they may bind the operator site simultaneously.

Several devices are used by proteins in order to bind DNA in the form of dimers. For example in the case of repressor proteins in prokaryotic organisms one can distinguish two subunits, an N-terminus and a C-terminus, the latter being responsible for carrying out dimerization between two molecules through non-covalent interactions. In the case of transcriptional activators especially found in eukaryotic organisms, the dimerization strategy is completely different, using the coiled-coil leucine-zipper motif, described above.

1.2. The Helix-Turn-Helix Motif

The helix-turn-helix (HTH) motif has been found in proteins that are directly involved in the regulation of gene expression and also in proteins that serve structural and catalytic roles in other cellular processes (45). Many prokaryotic DNA-binding

domains contain this motif whereas in eukaryotic organisms other DNA-binding motifs, such as the Zinc Finger and the Leucine Żipper, frequently occur (see section 1.1).

An α-helical motif is often used by proteins for DNA-binding. Such a structural feature is present in completely different and unrelated DNA binding proteins and may vary from protein to protein in terms of aminoacid composition which may influence recognition of a particular local structure or global configuration of DNA. A particular 3D arrangement of alpha helices gives rise to the helix-turn-helix family of DNA binding proteins eg. bacterial repressors, in which the two helices are roughly perpendicular to each other and are separated by a beta turn. Dimerization results in a rotationally symmetric molecule, in which a HTH motif is contained in each N-terminal domain, whereas the C-terminal domain serves for dimerization (45).

The most extensively studied prokaryotic regulatory proteins belong to bacteriophage lambda and related phages (i.e. phage 434 and p22) which code for two regulatory proteins (46), namely the repressor itself and cro. Such proteins provide a molecular mechanism for gene control in the form of a genetic on-off switch between the lytic and lysogenic life cycles. The genetic switch is contained in a small portion of the genome comprising the two structural genes that code for the two proteins, and the operator region on which they act. The genes are transcribed in opposite directions and Cro is switched on when RNA polimerase is bound to the right-hand promoter in turn switching on early lytic genes (Fig. 4). *Vice versa* the λ CI repressor is switched on when RNA polimerase interacts with the left-hand promoter, leading to the development of lysogeny during which lytic genes are repressed. Therefore the lytic-lysogeny decision is taken according to which of the two promoters within the operator is bound by RNA polimerase, this being dependent upon the binding of the Cro or repressor protein to three binding sites OR1, OR2 and

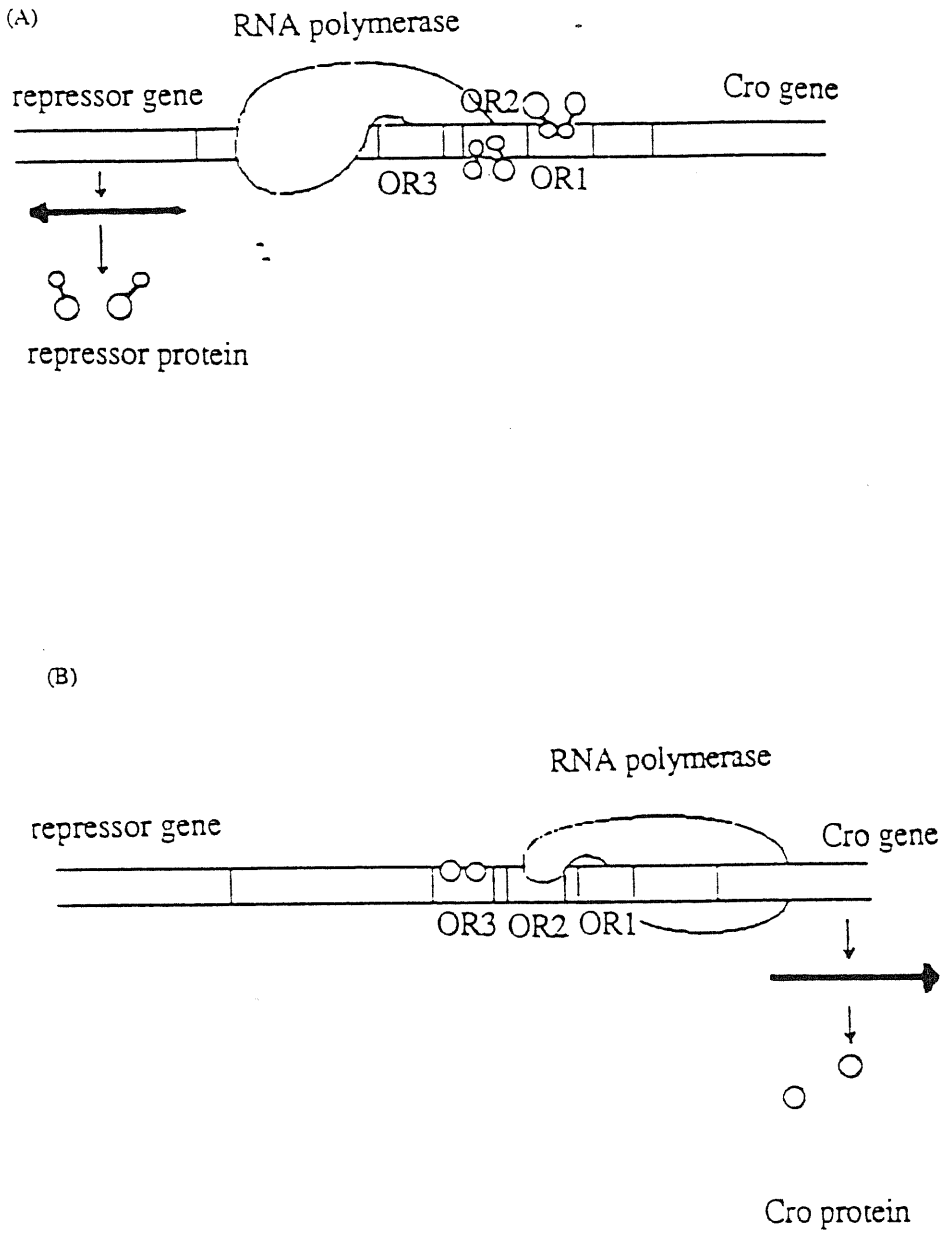


Figure 4. Schematic representation of the lytic to lysogeny switch. (A) Repressor proteins are shown as double spheres. They are bound so as to prevent transcription in the direction of the Cro gene, (B) Cro proteins (single spheres) are bound so as to prevent transcription of the repressor gene

OR3 (each 14 base-pairs long) in the operator. Such binding sites are palindromic sequences with a twofold axis of symmetry prôviding a dimeric repressor protein with two identical binding sites, one for each subunit. In addition to the operator regions (OR) the repressor also controls a second set of operators (OL) that regulate a different set of phage genes. It is significant that OL also has three binding sites (OL1, OL2 and OL3) each also 14 base-pairs long (46).

It has been postulated, and partly proven, that since the three operator sites (OR1, OR2, OR3) are separated from each other only by 6-7bp, both dimeric Cro and dimeric repressor proteins upon binding to one OR site somehow affect the binding of another dimer to the adjacent operator sites (47). Thus when repressor binds OR1, OR2 is bound more avidly by a second repressor molecule. This cooperation in the binding however seems to occur only in the case of the entire protein (48) and does not unequivocally appear in the case of the single, DNA-binding, N-terminal domain. For this reason one is more prone to think that cooperativity is mainly due to non covalent interactions involving residues in the C-terminal domain of the entire protein. It has also been suggested by experimentally inactivating OR1 through a mutation, that one repressor molecule bound at OR2 can directly interact with another molecule in OR3 and that in any case one bound dimer is able to make contacts only with one other dimer regardless of the operator. These observations can be extended both to the case of the right-handed operator region (OR) and to that of the left-handed one (OL).(46).

The crystal structure of λ Cro (50) shows a peculiar arrangement of alpha helices similar to that found the lambda repressor (45, 49, 51). The catabolite gene-activating protein CAP also has a similar arrangement (45). Helix 2 and helix 3 of the λ Cro N-terminal domain, form a helix-turn-helix motif. The domain responsible for dimerization is located at the C-terminus of the protein and is separate from the

DNA-binding region found in the N-terminal portion of the protein, from which it can be removed by proteolytic cleavage.

The helix-turn-helix motif is used to bind specific sites on DNA and utilizes one α-helix to contact the major groove along one face of the double helix. In Cro, each binding or recognition alpha helix in the HTH motifs of the two subunits are at opposite ends of the elongated dimer, and are separated by a distance of 34 Å (Fig. 5). This distance corresponds to one turn of a B-DNA double helix. Thus, the recognition helices bind into the major grooves separated by one turn of the DNA molecule. The second helix in each helix-turn-helix motif, separated from the recognition helix by a beta turn, lies across the major groove, making non specific contacts with DNA.

Thus four features of the Cro structure are important for specific binding to the DNA operator, (i) the presence of a HTH motif providing a correctly oriented recognition helix, (ii) the specific aminoacid sequence of this alpha helix, (iii) the subunit interactions that provide the correct distance and relative orientation for the two recognition helices of the dimer, (iv) non-specific interactions between the DNA and protein surfaces, which increase the binding affinity. In some cases this requires bending of the DNA to maximize surface interactions. The alpha helices forming the DNA binding motif can in fact be rich in basic residues such as lysine and arginine which, being positively charged, can directly interact with the negative phosphates of DNA.

Genetic studies carried out by Ptashne et al (46) further show that the proposed structural model for DNA binding by λ Cro and repressor is essentially correct and can be applicable to the Cro proteins and repressors of phage 434 and phage P22. The N-terminal domain of the repressor from phage 434, and its complex with the operator, have been crystallized and the structure determined (Fig. 6)(52). It comprises a weak dimer of 69 aminoacid subunits complexed with a 14 base-pair

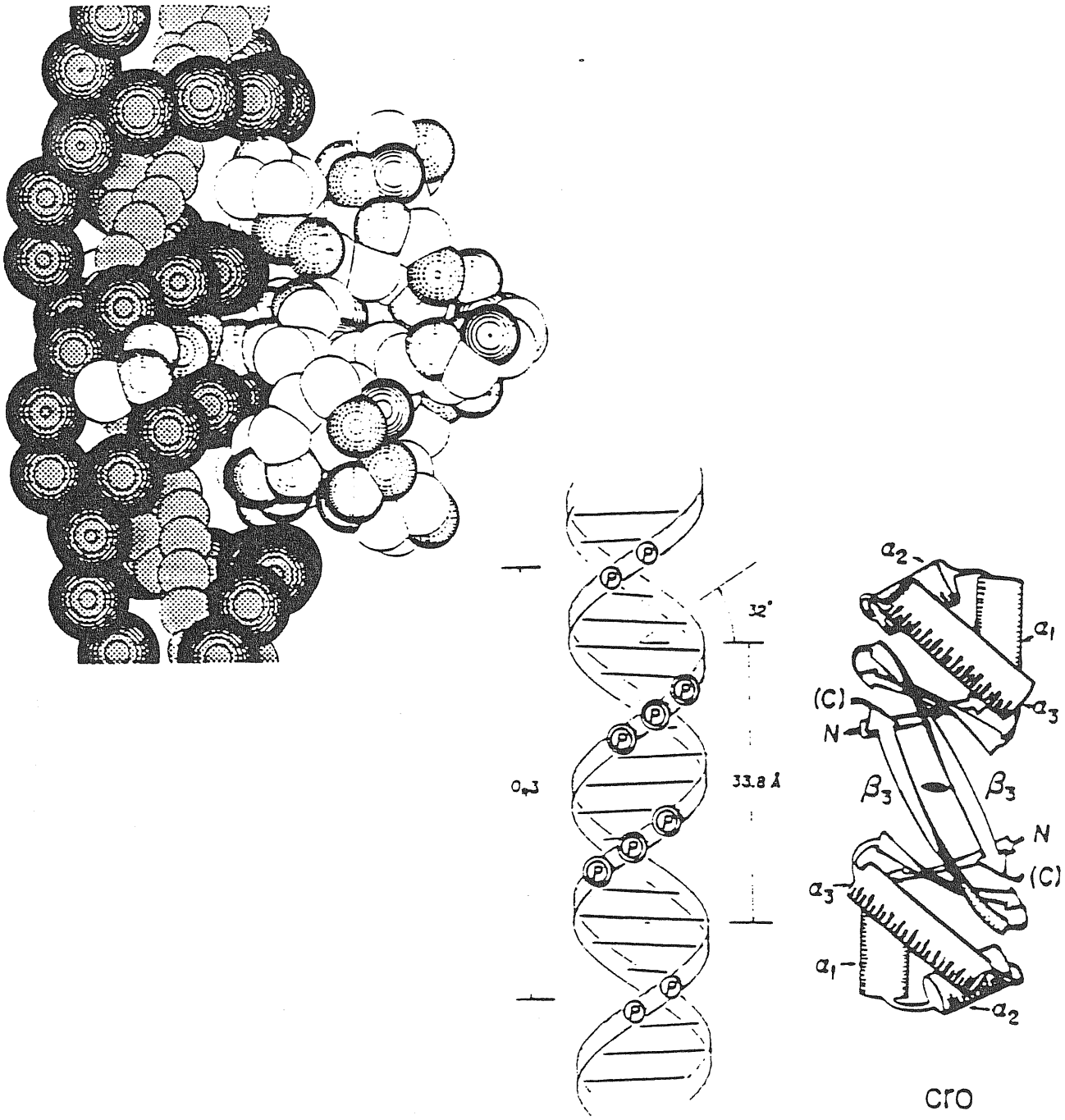


Figure 5. The λ Cro protein. (A) Space-filling model of Cro bound to its cognate DNA. , (B) schematic representation of the Cro protein aligned to DNA (43b).

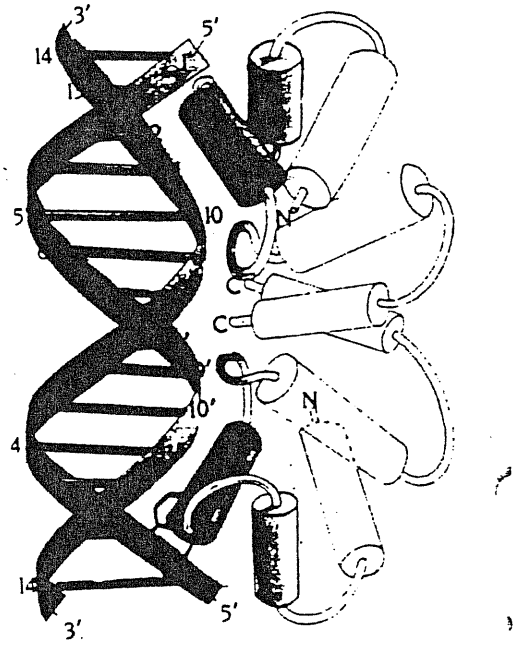


Figure 6. Structure of the CI 434 repressor N-terminal domain bound to DNA

oligonucleotide containing a palindromic sequence, and thus two identical half sites. Comparison of the N-terminal domains of 434 Cro (71 aminoacids) and repressor (69 aminoacids) shows a homology rating of 48% in terms of aminoacid composition (45). The homology is higher in the case of the HTH DNA-binding motifs. A high degree of homology also occurs when comparing the aminoacid composition of the DNA-binding motif of 434 Cro and repressor with the corresponding λ phage proteins. Like its λ counterpart the N-terminal subunit of 434 repressor consists of five alpha helices with H2 and H3 in the DNA-binding HTH motif, H3 being the recognition helix. While the degree of homology between λ and 434 repressors is only moderate the structures are very similar. The backbones of the two repressors are in fact highly superimposable (45, 52). Two HTH motifs are at either ends of the dimer in the N-terminal domain while the C-terminal domain is responsible for dimerization. Interestingly it was shown that both 434 Cro and 434 repressor are monomers in solution whereas they form dimers only when bound to the cognate DNA (45).

It has further been shown that both 434 Cro and repressor proteins induce DNA distortions which are related to the local twist, since DNA is overwound at the palindromic recognition sites and underwound at its ends, causing a narrowing of the minor groove at the center and a widening at the ends (45).

The helix-turn-helix motif has also been found in other proteins involved in gene regulation. Interestingly similar motifs have been discovered in proteins belonging to completely different and functionally unrelated organisms. Dallmann et al. (54) found that the repressor from the 16-3 phage that infects *Rhizobium meliloti* is able to cross bind the 434 operator site from the repressor of phage 434 which infects *Escherichia coli* and vice versa. The 16-3 and 434 operator sites consist of identical half sites separated by six non conserved nucleotides thus showing a related repressor specificity for unrelated phages. The 16-3 repressor shows only a moderate

homology in its N-terminal region to the 434 repressor, except for the HTH motif where it is high .

The tryptophane repressor (trp) and CAP are two DNA-binding proteins whose interaction with DNA is regulated by allosteric control i.e. small molecules whose binding causes conformational changes altering and thus regulating the DNA binding site and its affinity for DNA. The trp repressor controls the operon responsible for the biosynthesis of tryptophane which is regulated through negative feed-back. In fact in absence of L-tryptophane the repressor is inactive and the operon is switched on; in the presence of tryptophane the repressor is switched on and the operon is turned off. As the concentration of tryptophane increases it binds the repressor and induces a conformational change to the active form. The trp repressor has been crystallized (55) and is similar to the lambda repressor, apart from the arrangement of six N-terminal helices (45). The active form is a dimer in which the two monomers pack their helices so as to form a distinct overall strucure in which one can distinguish two heads and one core. The two HTH motifs responsible for the binding to the DNA molecule reside in the two heads of the dimeric molecule. A conformational change occurring upon binding of two L-tryptophane molecules provides the molecular mechanism of the functional switch, the orientations of the recognition helices being altered into positions appropriate for binding (Fig. 7).

CAP works as a positive control element that binds DNA and assists RNA polimerase in binding to certain promoters (45). It becomes a specific DNA-binding protein upon binding of cyclic AMP. CAP controls a set of operons among which the lac operon, responsible for sugar breakdown. When the level of lactose breakdown is too low, the level of c-AMP is increased, activing CAP. Structurally speaking, CAP consists of an eight-stranded jelly roll beta barrel containing a pocket which supposedly acts as the binding site for c-AMP. The barrel-like structure is linked to the DNA-binding region containing the HTH motif. A dimer is formed in such a way

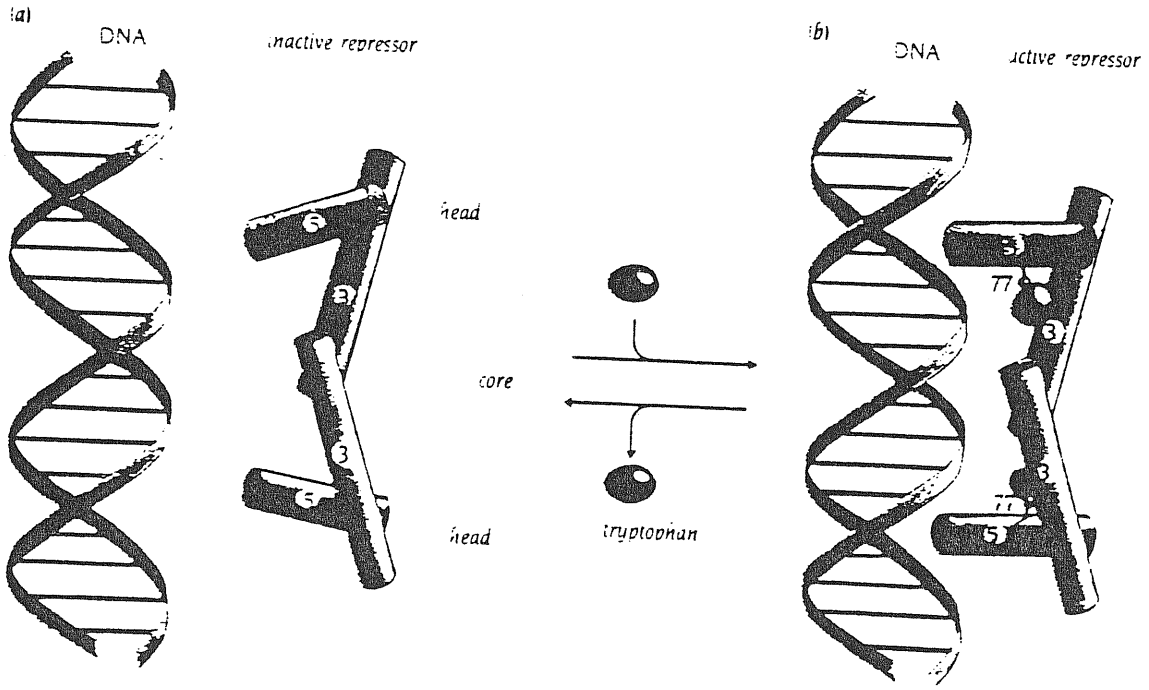


Figure 7. Schematic representation of the functioning of the of the trp repressor domain bound to DNA

that the two recognition helices of both HTH motifs are 34Å apart from each other, in a similar arrangment to that observed for lambda Cro.

Crystallographic data has been collected for a number of HTH type DNA-binding proteins (Table 2). NMR data concerning solution structures is also available. For instance, the 434 repressor protein has been determined by NMR methods (57) and compared to its crystal structure (58). Both data are found to be in relatively good agreement especially in the case of the packing of hydrophobic residues in the interior of the protein. Interestingly NMR carried out on urea denatured 434 repressor protein revealed a region from residue 53 to residue 60 rich in hydrophobic aminoacids whose conformation is different to that in the folded protein and may thus serve as a nucleation site in the folding pathway of the protein (59).

Another example of the usefulness of NMR in structure determination of a typical helix-turn-helix DNA binding protein is given by trp repressor. The crystal structure (55) surprisingly shows that there is no evidence of direct specific interactions with DNA bases. Several NMR studies were carried out on the solution structure of the trp repressor (60, 61), showing that helices D and E (which build up the HTH DNA-binding motif) of the five helix bundle are somehow more flexible than is the core region of the protein, in accordance with the crystallographic data, and a model was proposed in which trp repressor would scan the DNA molecule by means of its flexible reading head until it reaches its operator site.

In conclusion, the HTH motif is a well-defined supersecondary structural element that has convenient geometric features for DNA recognition. However the remainder of the protein in which the motif is located can vary significantly among different repressors thus implying that dimerization interfaces can be very different from each other even among functionally related proteins. Dimerization is carried out in a number of ways, including nearly parallel α-helices (CAP), antiparallel helices (

λCI), antiparallel β-sheet (λ cro repressor) or an intertwined arrangement of helices (trp repressor) (62, 63).

TABLE 2. Existing x-ray cristal structures for DNA-binding proteins containing the HTH motif.

Protein	Resolution (A)	R factor (%)[a]	Co-crystallized DNA	Reference
λCI repressor	2.5	24.2	TATATCACCAGTGGTAT TATAGTGGTCACCAT	(51)
λcro repressor	3.9	-	TATCACCGCGGGTGATA ATAGTGGCGCCCACTAT	(50)
P434 repressor	2.5	17.9	TATACAAGAAAGTTTGTACT NTATGTTCTTTCAAACATGA	(53)
P434cro repressor	3.2	36.6	ACAATATATATTGT TGTTATATATAACA	(26)
trp repressor	2.4	24.9	TGTACTAGTTAACTAGTCN NCATGATCAATTGATCAGT	(55)
CAP	3.0	-	GCGAAAAGTGTGACATT* NGCTTTTCACACTG	(8)

a) The R factor is defined as $R=\Sigma|Fo-Fc|$ where Fo and Fc are the observed and calculated structure factors respectively. The nucleotides responsible for making specific contacts are in bold charachters.

2.0 DEVELOPING MODEL PEPTIDES BASED ON THE HTH MOTIF

The development of model DNA-binding polypeptides based on the H-T-H type of DNA-binding protein can be approached in two ways. One is to take a protein subunit, say the N-terminal domain of the CI434 repressor, and devise a way of synthesizing it with modifications that will allow features such as a permanent dimerization, so as to obtain an improved synthetic protein. The other is a minimalist approach, in which the minimum ammount of structure that may give specific binding, as determined via computer modelling, is abstracted from the subunit and the appropriate peptide synthesized and tested. This requires a knowledge of all the factors that affect the interaction of the subunit with DNA and a decision as to which can be expended with, and which are instead fundamental. Thus one must not only consider those residues that are directly involved with binding to DNA bases (which give specificity to the binding), but also those that are involved in non-specific contacts with the DNA bases and backbone (which increase the affinity for the DNA), those which are involved in providing internal stabilization and hence the correct folding of the entire N-terminal domain of the 434 repressor protein under study, and those which provide for dimerization.

An idea as to which intramolecular interactions are structurally important for the subunit can be gained from three simple approaches: (i) observing the computer model so as to visually detect significant side chain interactions, (ii) determining the distance matrix for all residues in the subunit, as sequentially distant residues which are spatially close may be involved in stabilizing interactions (iii) determining the degree of conservation of residues among DNA-binding domains from structurally similar but sequentially different phage repressors, as this may indicate residues in structurally important positions.

The first approach was carried out by using the Biosym Insight programme and x-ray crystallographic structures abstracted from the Brookhaven database. The second was effected by using the Discover programme from the same package. This is normally used for energy minimization and molecular dynamics, but as a part of its iteration cycles it measures the interatomic distance for all atoms in the structure with all others. The third approach uses database searching programmes (such as those in the Intelligenetics or GCG suites) to abstract the sequences of all known bacterial repressors and cro proteins from databases such as Swissprot, and other programmes to carry out multiple allignments.

The two best candidates for modelling were the DNA-binding domains of the λ repressor or of the CI434 repressor, as detailed crystal structures for these complexed to DNA are available, and they do not require allosteric activators, which would be an added complicating factor. The 434 repressor cross-recognition of the 16-3 operator site added a particular interest to this repressor.

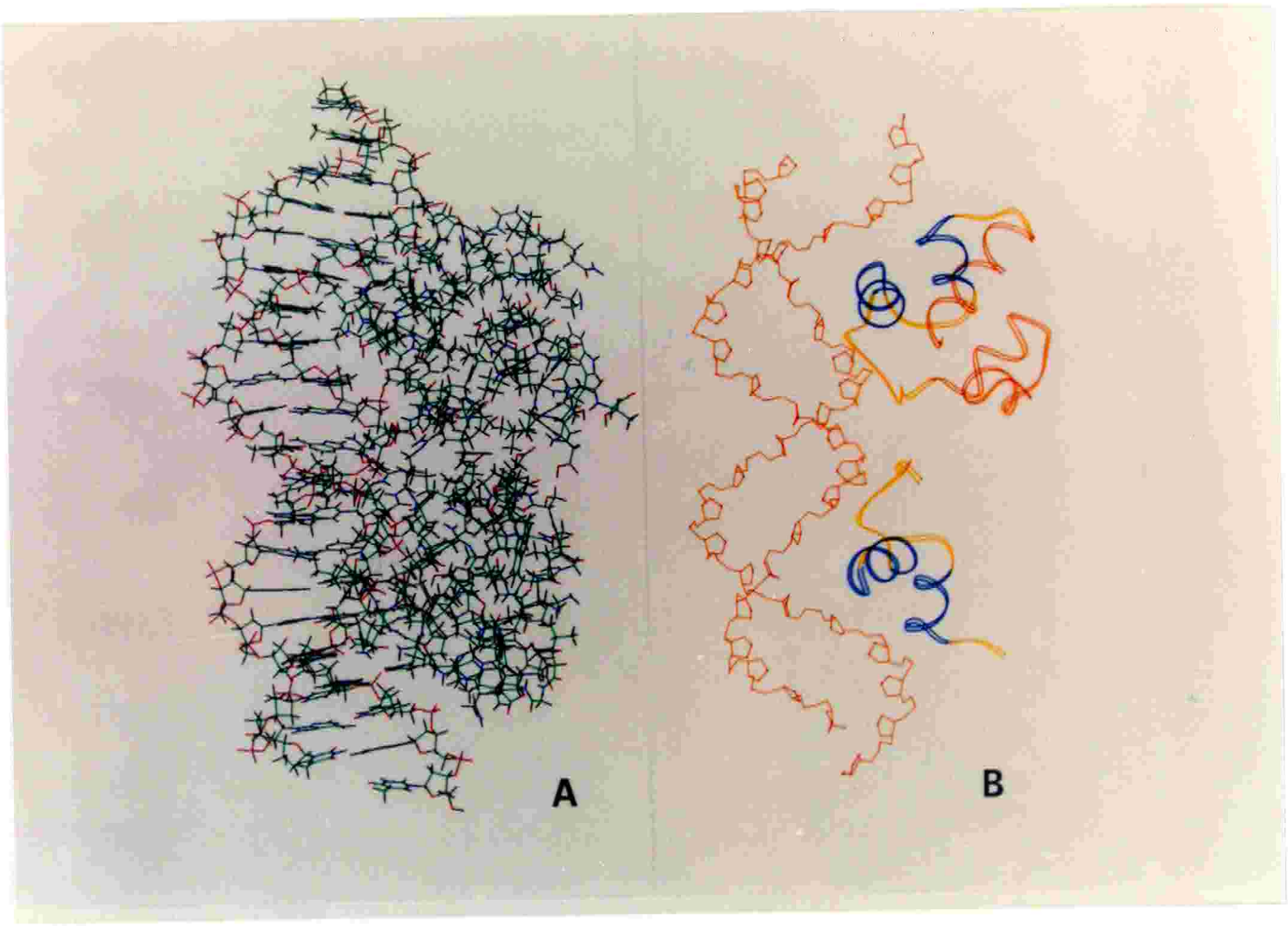
2.1. Computer Modelling

Complexes between both the lambda repressor N-terminal domain and the CI434 repressor N-terminal domain and their cognate DNAs were available for modelling. Sequence comparison of 434 repressor and lambda repressor highlighted a reasonable level of homology, rating up to 48% in the case of those residues located within the helix-turn-helix binding motif. Structural similarity between the lambda repressor and the 434 repressor was tested by superimposition and found to be very high (results not shown). This finding is important (results not shown), as it indicates that those residues which are conserved between the two repressors in their N-terminal subunits may play a structurally important role. The X-ray structure of the CI 434 repressor N-terminal domain bound to DNA is shown in Figure 8A. Visual inspection of the molecular model was carried out

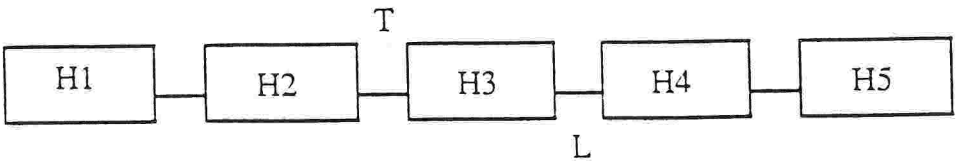
with the Insight II programme in the Biosym package, on a Silicon Graphics Power Plus workstation. This programme allows one to reduce the protein to a simplified form, such as the ribbon model shown in Fig. 8B, and then introduce the side chains of each residue in the protein, one by one, so as to directly observe their interaction with DNA or other residues. A schematic representation of the N-terminal domain and its sequence are shown in Fig. 8D and E.

The following aspects of the interactions of the CI434 N-terminal domain with its cognate DNA were analysed: i) the interaction of the binding helix H3 with the DNA bases in the binding site, ii) the effect of the stabilizing helix H2, iii) interactions of residues outside the HTH motif with the DNA (including the formation of a complementary surface between the DNA and the protein, which requires conformational changes in both), iv) interactions between residues in the N-terminal domain and v) interactions between the two N-terminal subunits leading to dimerization. The review by Harrison and Aggarwal (62) was of considerable help in this respect.

1) The recognition helix H3. The recognition helix consists of residues 28 to 36. Some of these residues project their side-chains into the major groove of the cognate DNA in order to provide non-covalent bonding with the edges of base-pairs (Fig. 9). There are six distinct binding sites for the repressor, the three operators OR1, OR2 and OR3 that control the lysogeny/lytic decision (Fig. 10) and three operator sites involved in the regulation of other phage genes (OL1 to OL3). Residues Gln 28 and Gln 29 are strongly involved in specific hydrogen bond formation. In particular, if the palyndrome is given by the sequence 5′ ACAANNNNNNTTGT on strand 1, Gln 28 interacts with N6 and N7 of the first adenine located at the 5′ end whereas Gln 29 forms hydrogen bonds with the O6 and N7 of the second guanine residue located at the 3′ end of the sequence 3′-TGTTNNNNNNAACA on strand 2. Interestingly a sort of hydrophobic pocket is formed by the methyl groups of the side chains of Thr 27 and Gln 29 which is



D



E

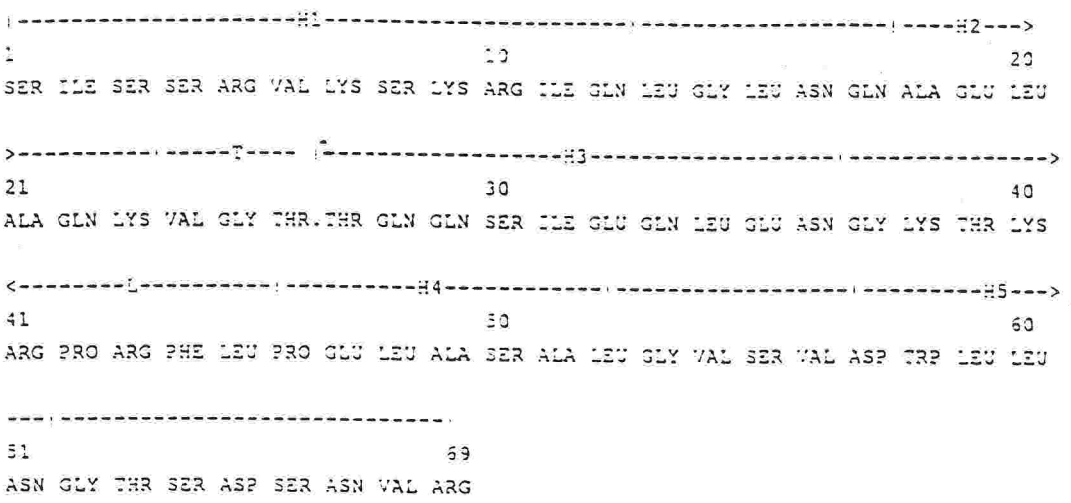


Figure 8. (a) Crystal structure of the CI434 N-terminal domain bound to its cognate DNA (B) Ribbon type representation of the N-terminal domain, the binding helix H3 and stabilizing helix H2 are in blue. The turn T and loop L are in yello, the other helices (H1,H4 and H5) are in orange. (C) Ribbon type representation of the HTH motif. (D) Schematic representation of the N-terminal domain showing the order of the helices H1 to H5. (E) Sequence of the CI434 N-terminal Domain.

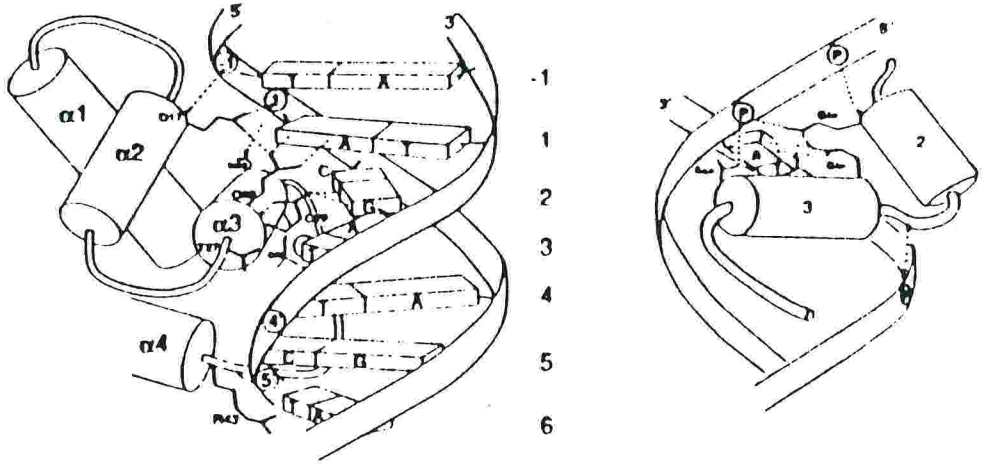


Figure 9. Schematic representation of the interactions of sidechains from residues in CI434 repressor with DNA bases and backbone. (A) Interactions from residues in helices H1, H2, H3 and H4 with DNA. (B) interactions from residues in the HTH motif seen from another angle.

Operator sites

```
OR1     5' A C A A G A A A G T T T G T 3'
        3' T G T T C T T T C A A A C A 5'
OR2     5' A C A A G A T A C A T T G T 3'
        3' T G T T C T A T G T A A C A 5'
OR3     5' A C A A G A A A A A C T G T 3'
        3' T G T T C T T T T T G A C A 5'

OL1     5' A C A A G G A A G A T T G T 3'
        3' T G T T C C T T C T A A C A 5'
OL2     5' A C A A T A A A T A T T G T 3'
        3' T G T T A T T T A T A A C A 5'
OL3     5' A C A A T G G A G T T T G T 3'
        3' T G T T A C C T C A A A C A 5'
```

Synthetic operator

```
                1       5       10      15
                |       |       |       |-->strand1
Syn-O   5' T A T A C A A G A A A G T T T G T A C T
      3' T A T G T T C T T T C A A A C A T G A A
                15      10      5       1
                |       |       |       |<--strand2
```

Figure.10 The six operator regions which bind the CI434 repressor.
(Also shown is a synthetic oligonucleotide, based on the operator regions, used in determining the repressor/DNA complex crystal structure, as shown in Figure 8).

probably important because it accomodates the methyl group of the third thymine on strand 2 (45).

It is important to observe that the first three residues on the three operators (OR1, OR2 and OR3) recognized by the 434 repressor protein, are identical as one can see in Figure 10. This is evidence that interactions between Gln 28 and Gln 29 with these first three base-pairs cannot be the device that contributes to the discrimination among the various binding sites on the DNA molecule. Rather, one could think of them as providing the general recognition of λ434 for all the operator regions. The pattern of hydrogen bonds and hydrophobic interactions therefore accounts for the recognition of the operator region both by the 434 Cro and the 434 repressor proteins. The role of Gln 29 results of particular importance because it interacts with two different base-pairs, namely C-G and A-T, in two different ways, by hydrogen bond formation and through hydrophobic interactions respectively (45). Interestingly it has been experimentally proven that replacement of glutamines 28 and 29 produces mutant phages which are no longer viable (45). Other residues in H3 which on visual inspection come close to the DNA bases and may be involved in binding are Gln33 and Asn 36.

2) The stabilizing helix H2. The helix H2 lies orthogonally across and behind the binding helix H3 (with respect to DNA, see Fig. 8B) and is thought to stabilize it. From the molecular model it is apparent that residues Gln17 comes into contact with residues Glu32 and Glu35 in H3, and residue Leu 20 comes into contact with Ilè 31 in H3, as well as hydrophobic residues in the loop L and helix H4, forming a hydrophobic pocket (Fig. 11). Also apparent are interactions between Asn 16 and G1 and T2 in strand two of the DNA operator site (see Fig. 10 for the sequence of the DNA) and Gln17 with T2 and A3 of the same strand and A1 of strand 1.

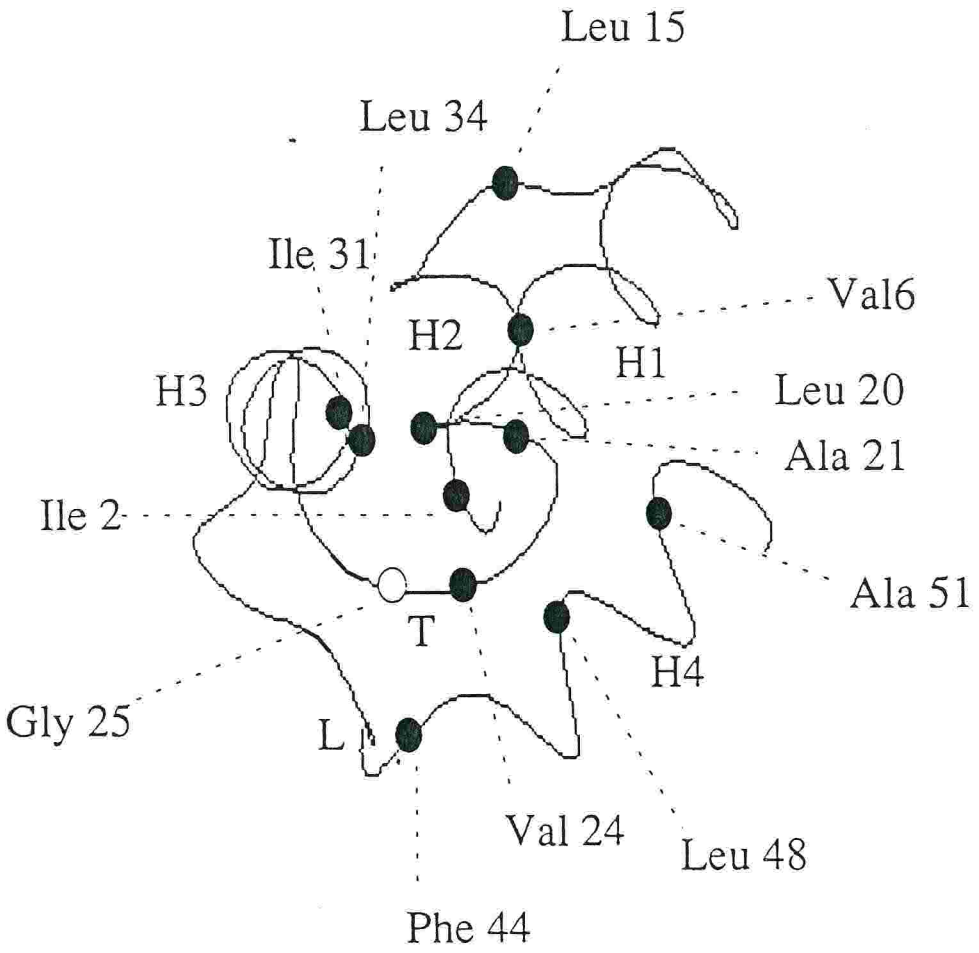


Figure 11. Schematic representation of the positions of hydrophobic residues in the hydrophobic core formed by the helices H1, H2,H3 and H4, the turn T and the loop L in the CI434 repressor N-terminal domain. Helix H5 is omitted for clarity. Hydrophobic residues are shown as filled circles, while the highly conserved residue Gly 25 in the turn as an open circle.

3) Interactions of residues outside the HTH motif with DNA. Several residues outside the HTH binding motif are involved in nonspecific binding to DNA, via either hydrophobic interactions with DNA bases, a hydrophobic or electrostatic interactions with the DNA backbone.

The loop L following the binding helix H3 contains two lysine (positions 38 and 40) and two arginines (positions 41 and 43) which are placed relatively close to the DNA phosphodiester backbone (Figure 12), and which may be involved in electrostatic interactions with the phosphate groups. Lys 38 from one subunit comes into contact with the DNA in proximity to the bases C4 (strand 2, Fig. 10) while Lys 38 from the other subunit comes into contact with the DNA in the proximity of A10 (strand 1). Regarding the arginines, one subunit points its Arg 43 directly at the DNA minor groove, where it comes into contact with the base T 10 in strand 1and T 9, T 10 and T 11 in the other strand of the cognate DNA, while that in the other subunit comes into contact with A 8 and G 9 in strand 1 and T 10 in strand 2. Arg 41 points upwards towards the DNA phosphodiester backbone in one subunit, but in the other it points away from the DNA (Figures 12 and 13).

The formation of preferential complementary surfaces between the DNA molecule itself and the protein are strongly favoured by the ability of the DNA to undergo structural changes, which in turn are favoured by the establishment of non-specific interactions between proteins and the sugar-phosphate backbone. In the case of the CI 434 repressor, the structural change essentially involves the bending of DNA and overwinding of the middle regions of the operator (Fig 8). When this does not occur, as in the binding of the 434 repressor to OR3, binding is weak. The protein-DNA complementary surface is quite extended and since each monomer of the repressor protein can see only a half operator, the DNA molecule has to undergo the abovementioned structural changes so as to coincide with the dimeric organization of the protein itself .and to accomodate the protein in the best manner. The arginine in position 43 projects into the minor groove and

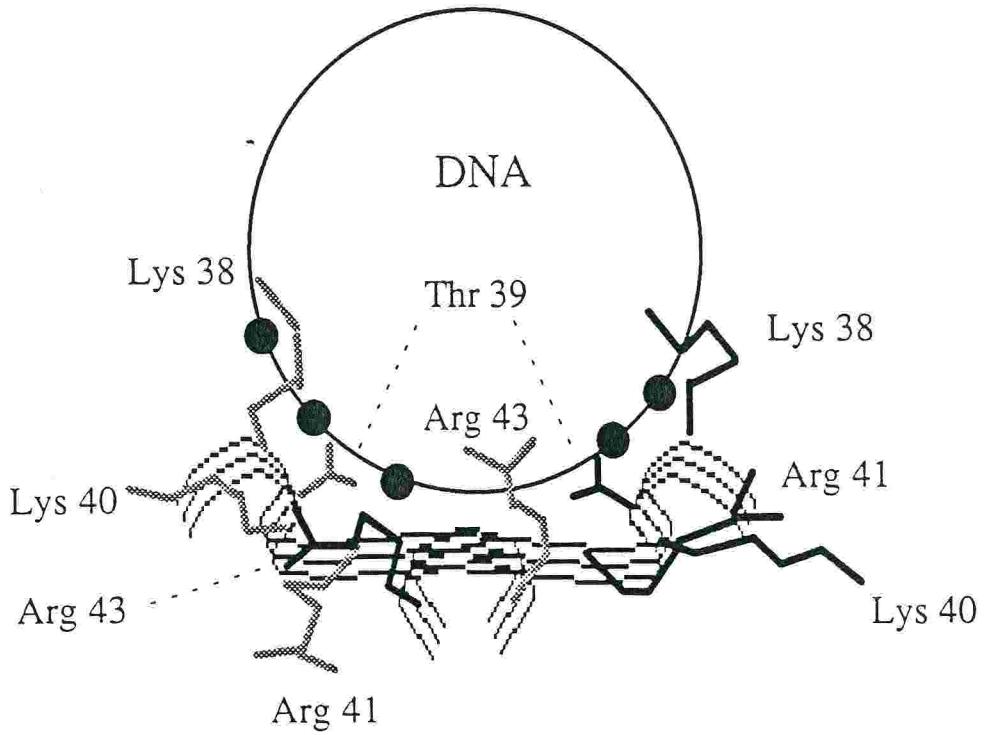


Figure 12. Schematic representation of the interactions of side-chains from residues in the loop region of the CI434 repressor N-terminal domain with DNA. The diagram is based on a computer visualization of the X-ray structure. The DNA is viewed along the major axis and phosphate groups in the general vicinity of the loop are shown as filled circles. Side chains from one subunit are in grey, those from the other in black

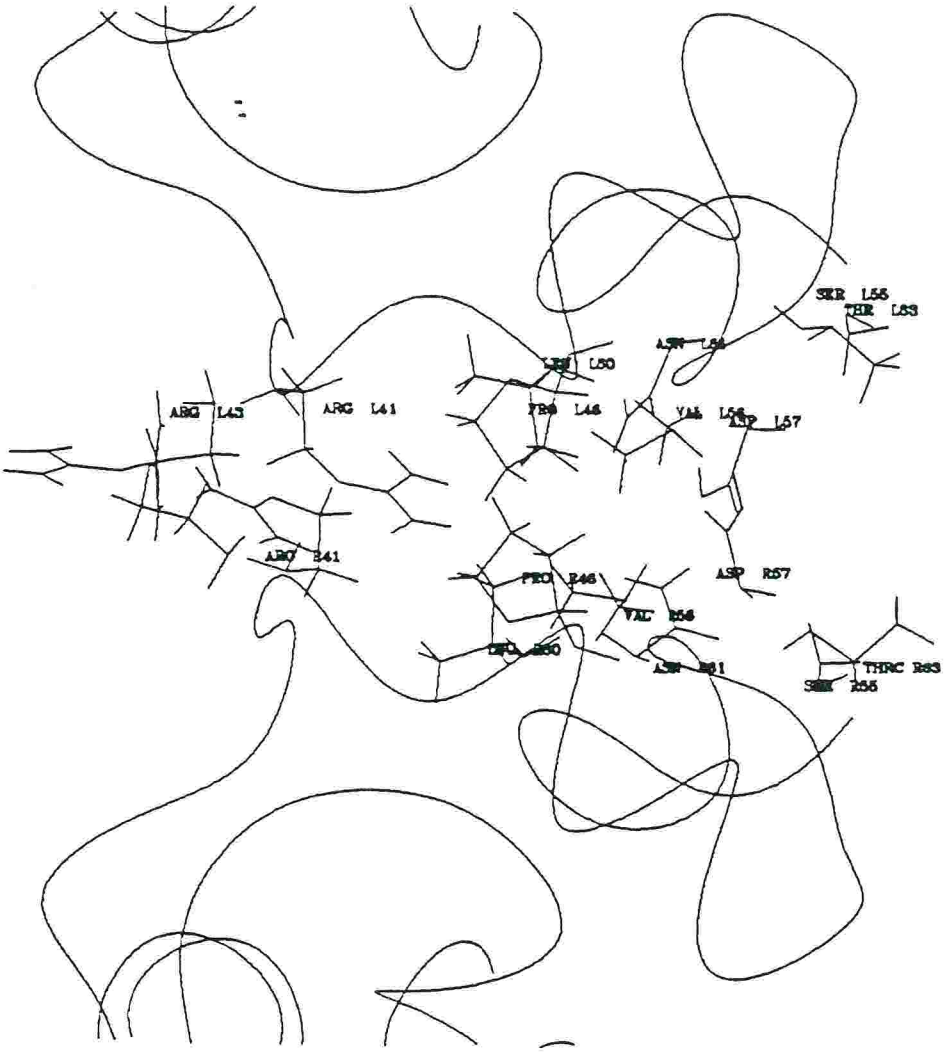


Figure 13. Side-chains of residues at the interface between the two subunits in the CI434 repressor N-terminal domain.

provides a positive charge that shields the DNA phosphate groups, thus facilitating the compression of the minor groove. This is a typical non-specific interaction with a multiple role, namely that of i) directly stabilizing interaction with DNA via electrostatic interaction, ii) stabilizing the change in DNA conformation and iii) as a result of the latter effect, increasing the complementary surface between the DNA and the protein dimer.

The arginine at position 10, and thus in the helix H1in both subunits also points at the DNA phosphodiester backbone, and may be providing an electrostatic anchoring interaction. Another residue that contacts the DNA is Asn 16.

4) Interactions between residues in the N-terminal domain

The N-terminal domain is a compact structure consisting of 5 α-helices connected by turns or loops. Hydrophobic interactions at the core of the subunit are important for the stability of the structure and its correct folding. One could detect these interactions by looking at the computer model, but a more efficient method involving a distance matrix is described in section 2.2. Modeling was useful in validating the indications given by the distance matrix, and the results of these studies is described in detail in that section.

5) Interactions between subunits. Dimerization in the CI434 repressor is mediated by the C-terminal domain. In its absence, a weak dimer is formed in the presence of DNA on crystallization but does not survive in solution. Examining the molecular model it was found that the residues in the loop L come into close contact with residues in the loop L of the other subunit and with two residues (positions 56 and 60) in the helix H5 and *vice versa* (Fig. 13). From the molecular model it was possible to establish that the point of closest approach for the two subunits is at residue Pro 46 with Cα to Cα distance of 9.9Å and Cγ to Cγ distance of only 5.85Å.

2.2. Inter-residue Distance Matrix

It is logical to suppose that sequentially distant residues within the CI 434 repressor N-terminal subunit that come spatially close to each other may be acting to stabilize the structure as a whole, by means of hydrophobic or electriostatic interactions. It is not possible to efficiently individuate all such interactions by means of a visual inspection of the molecular model. The modelling programme however allows a listing of all interactions as a side product of its energy minimization routine. By definition, the calculation of interaction energies, as used in minimization, requires a measuring of all interatomic distances within the molecule. These are stored in a data file, and it is relatively simple to select only certain interactions (say Cα to Cα) and filter out all those above a cutoff distance (say 7Å). The result of this process is best summarized in a distance matrix (Figure 14) all on the diagonal and can be disregarded, while all spatially vicinal but sequentially distant residues fall above the diagonal. Below the diagonal are represented residues from different subunits that come within 7Å.

As far as the HTH motif is concerned, results from the distance matrix confirm that there are a number of interactions between H2 and H3, but it also points to several interactions between H2 and the turn T. What is immediately apparent from the matrix is that a number of interactions occur between the loop L after H3 with helix H3 and the turn T. Returning to the molecular model, it was in fact found that hydrophobic residues (Val 24, Ile 31 and Phe 44) in T, H3 and L are part of a hydrophobic pocket, to which Leu 48 and Ala 51 from the beginning of helix H4 also contribute.

A number of interactions are also indicated beteen the binding helix H3 with helix H1. In particular, hydrophobic interactions are indicated for Ile2 and Val6 in H1 with Leu 34 in H3 (Figure 11), and electrostatic interactions for Arg 10 and Glu 35.

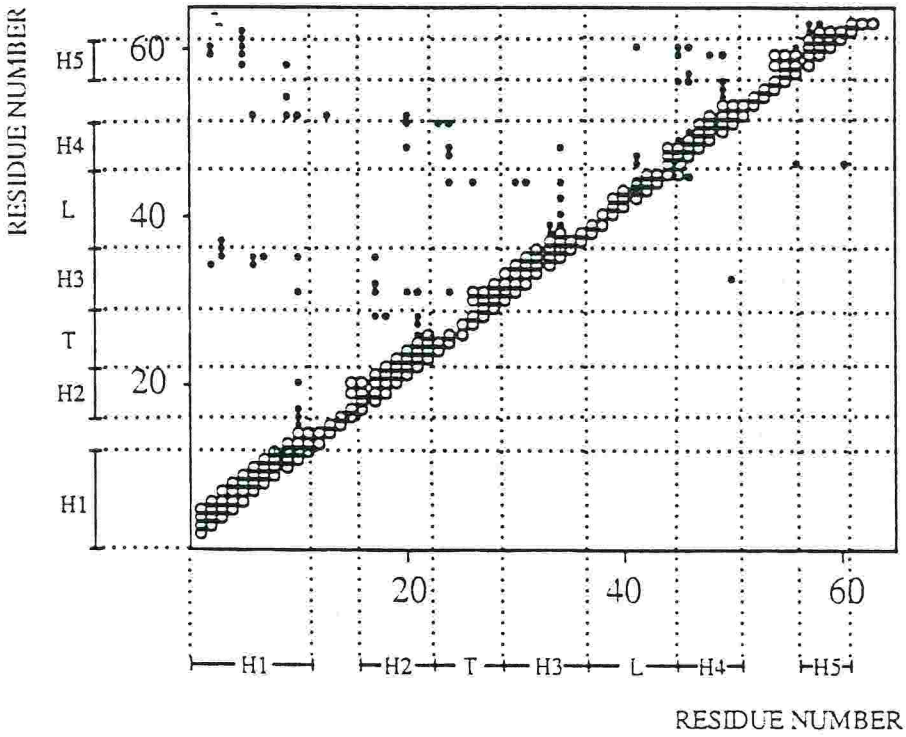


Figure 14. Distance matrix of residues in the CI 434 N-terminal domain (7Å cutoff). The grid divides into its helix (H_1 to H_5), turn (T) and loop (L) secondary structure elements. The DNA binding motif is H_2TH_3.indicates vicinal residues, indicates distal residues.

2.3 Homology Studies

A data base search was carried out on the Swissprot database for all bacterial phage repressors, which structurally resemble CI 434. The aminoacid sequence of the helix-turn-helix motif, flanked by 10-20 residues on each side, was then abstracted from the complete sequences of each repressor, and the Intelligenetics programme Genalign used to perform an alignment. It was thus possible to determine those aminoacid residues that are conserved. As this homology study was carried out in aid of structural determinations, it was not an exact homology that was of interest, but the conservation of physical characteristics, such as hydrophobicity, aromaticity or charge.

The alignment of the sequence abstracted from the CI 434 N-terminal domain with those abstracted from other phage repressors is shown in Fig. 15. Directly conserved residues, or hydrophobic residues that are conserved as such are in bold characters, while conserved charged residues are in italics. It can be seen from the alignment, that hydrophobic residues are conserved at positions 13,15,20,21,24,25,31,34 and 44/45 (with respect to numbering in the CI 434 N-terminal domain). Inspection of the molecular model and distance matrix for CI 434 shows that the leucine in the loop before helix H2 at position 13 does not appear to interact with other residues, while that at position 15 interacts with the leucine at position 20 in helix H2. This in turn forms part of a hydrophobic pocket with Val 24 in the turn T, Ile 31 in the binding helix H3 and hydrophobic residues in Helix H4. The alanine at position 21 is highly conserved among all repressors and is probably required in this position to allow packing of the helices H2 and H3 (45). A glycine at position 25 is also highly conserved and probably is required to allow formation of a tight turn at T. Residue Val 24, apart from the abovementioned interactions, also comes into contact with the Phe at position 44, as do Ile 31 and Leu 34 in the helix H3. Thus, the conservation of hydrophobic residues in these positions is explained by the need to pack the protein correctly and stabilize it via the formation of a hydrophobic core.

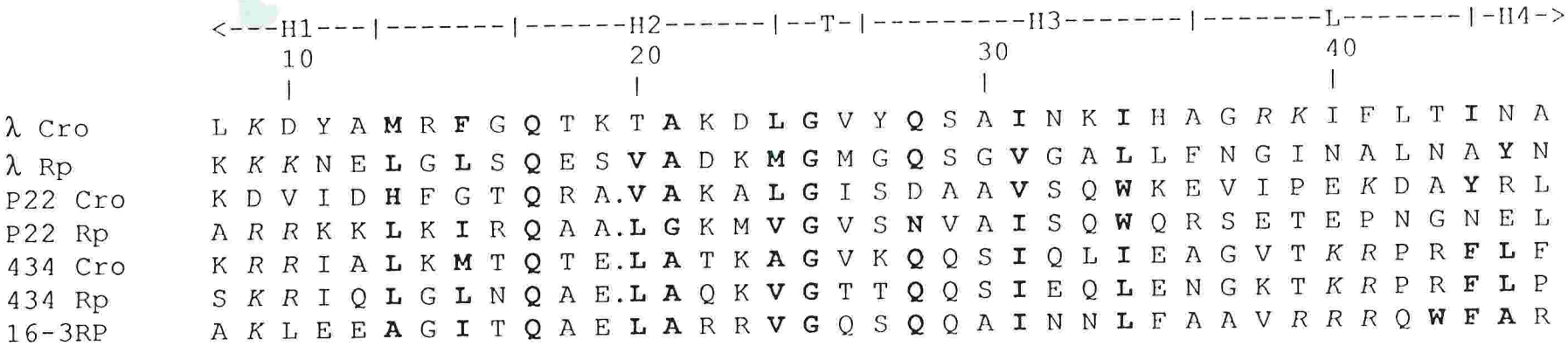


Figure 15. Sequence alignment of helix-turn-helix motifs found in DNA-binding phage repressor (Rp) and Cro proteins.

A cluster of conserved basic residues is apparent around podsition 10. From the model and distance matrix it was found that the arginine in this position contacts the DNA and it appears to be involved in electrostatic interactions and H-bonding to the bases (62). This residue is at the end of helix H1, pointing towards the DNA, and probably has similar role in different repressors. At the other end of the HTH motif, another cluster of conserved basic residues appears in the loop L at around position 40. Conservation is not so high in this case and it is possible that Van Der Waals complementarity of the loop with the DNA phosphodiester backbone (62) is more important than electrostatic interactions. An arginine at position 43 is present only in the 434 repressor and Cro, so it is possible that its role is that of mediating the conformational changes required of the cognate DNA for this particular system.

Also highly conserved are glutamines at positions 17 and 28. In the first case, the residue comes at the begining of the helix H2, pointing towards DNA and is probably acting in a similar manner as Arg 10. Gln 28 at the beginning of the binding helix is involved in recognition of the binding site for all repressors. Asn 16, which also contacts DNA in CI 434, is instead peculiar to this repressor.

Thus, a homology study of different repressors combined with structural studies can give valuable information as to those residues which play either an important role in the recognition and binding of DNA, or in maintaining the stability of the DNA-binding region in the protein. Such studies have also been carried out in an attempt to derive rules for identifying probable HTH motifs in proteins (45, 49, 62, 63).

2.4 Design of DNA-binding peptides based on the HTH motif: A minimalist approach.

Having obtained structural information as to which residues are important for binding to DNA, and which are imortant for stabilizing the structure, one can begin to design a peptide which contains the minimum ammount of structural features that will allow it to function. Reducing the length of the viable peptide to a minumum is desirable as there is a limit to the length of peptides that may be easily and inexpensively chemically synthesized in good yields.

The following features must be incorporated into the peptide: i) The HTH motif, ii) sufficient residues flanking this motif, on either side, to promote folding into the correct structure for binding iii) a number of residues apart from the binding helix which will interact with the DNA and improve the affinity of binding, iv) some means of introducing stable dimerization to give correctly oriented binding helices and v) from practical considerations, a balanced number of hydrophobic and charged residues should be present, so as to allow the peptide to dissolve in acqueous solution and to fold correctly.

Considering the information obtained from the molecular model, the distance matrix and the homology study, it was decided to begin the peptide at Pro 46. This is the point of closest aproach for the subunits in the repressor, and the distance between the two proline residues in different subunits (about 9Å) is spannable by a lysine linking the carboxylic groups of the two with its a and e amino groups. Thus a way of introducing a permanent form of dimerization is possible. Continuing the synthesis towards the N-terminal of the repressor, one introduces the loop L and then H3, the turn T and H2. It was decided initially to terminate the synthesis at this point.

Thus the synthesis would require 30 coupling reactions, including an initial branched lysine, which is well within the capacity of an automated

synthesizer, and assuming a reasonable coupling yield of 98% per reaction, would lead to a final yield of 54%. The resulting peptide, termed HTHLDIM, would contain two branches, each with 30 residues, of which 22 neutral or hydrophobic, 5 basic and 3 acidic (Fig. 16.)

Before synthesizing the peptide, a modelling study was carried out to test whether it was viable. The molecular model for the CI 434 repressor N-terminal domain (Brookhaven entry PDB2OR1, Fig. 8A, 17A), consisting of two protein subunits each of 69 resideus, bound to a 20 base-pair cognate DNA fragment (Fig. 10) and was taken and 40 counterions (Na^+) added so as to shield the DNA phosphate groups. The counterions were placed manually roughly in proximity of each phosphate, and allowed to take up a correct position by use of the molecular dynamics programme Discover (10000 cycles), with the dielectric set to water. This was carried out so as to mimic the real environment of the DNA-protein complex more closely. The residues from 47 to 69 in each residue were then excised and the C-terminus COOH reconstructed. A lysine was then bound via its a amino group to the C-terminus of the bottom subunit at Pro 46 (Fig. 17B) and a push/pull force applied between the e amino group and the C-terminus of the top subunit (also Pro 46) with equilibrium set at a 2Å distance (Fig. 17C). The model was then subjected to 10000 cycles of molecular dynamics (with all atoms except those of the last 10 residues in the bottom subunit fixed), and the two groups brought within 2Å of each other. The programme Insight was then used to create a bond between the e amino group and the C-terminus of the top subunit, but due to the characteristics of the programme, this dislodges the bottom subunit from the DNA, so as to give a physically correct bond. Editing of the molecular coordinates file, so as to replace back the correct coordinates for a DNA bound subunit then resulted in DNA-bound, covalently linked subunits, but with a physically incorrect bond between the bridging lysines and the top subunit (Fig. 17D). Energy minimization was then carried out to correct the bond (Fig. 17E).

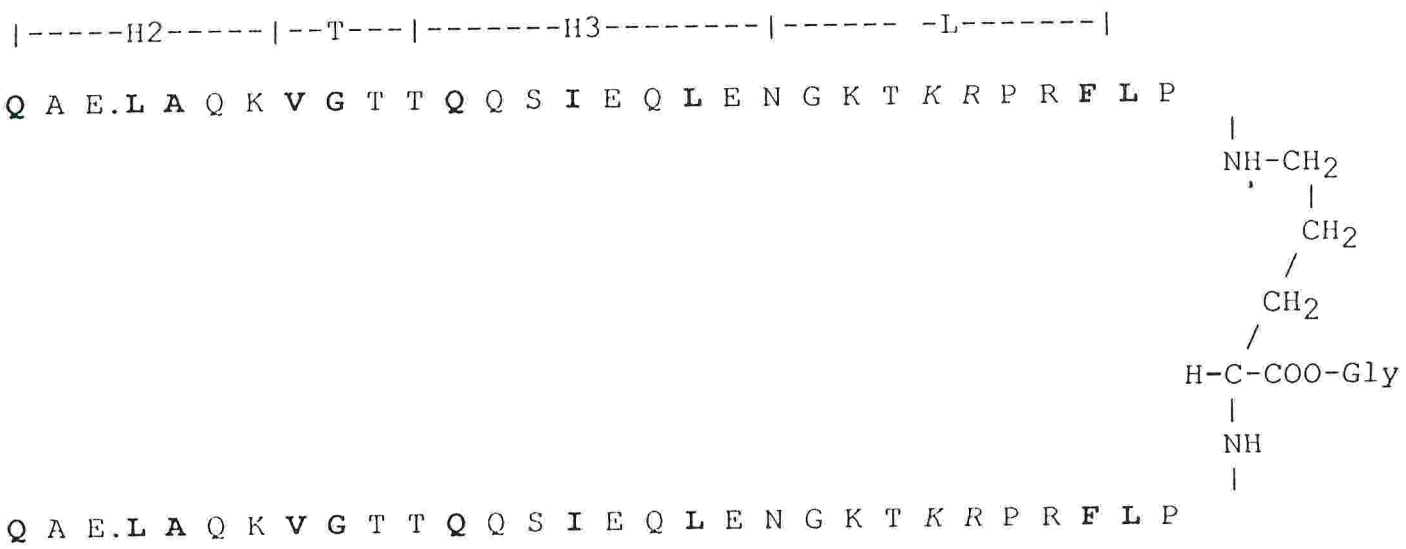


Figure 16. The model peptide HTHLDIM

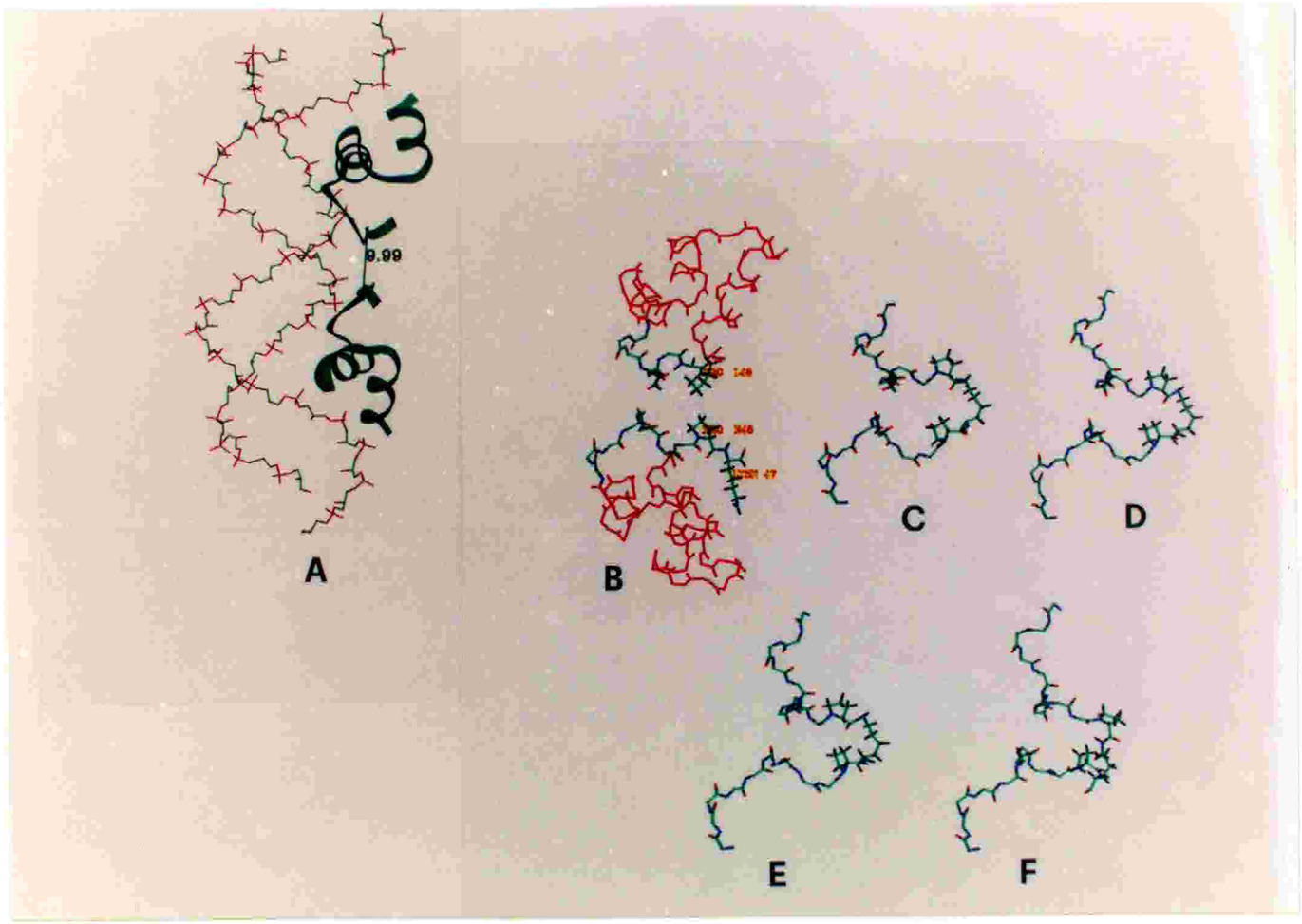


Figure 17. Modelling dimerization

17A HTHL peptide bound to DNA. The distance between Pro 46 residues in each peptide (Cα - Cα) is shown.

17B Lysine is bound via Nα to Pro 46 on one subunit (backbone for residues 1-36 in red, for 37 to 46 coloured).

17C Nε is pulled to Pro 46 on other subunit (within 2Å) using molecular dynamics (DISCOVER). All atoms except in residues 37-46 and Lys 47 are kept fixed.

17D Lys is linked to the other subunit via peptide bond formation with C-terminal carboxyl group.

17E Energy minimization of D

17F The dimerized peptide is subjected to molecular dynamics (10000 cycles, only coloured region) so as to obtain the minimum energy conformation, without showing excessive distortion

At this stage a model for the HTHDIM peptide, bound to DNA, had been obtained, but the question remained as to whether this structure was energetically acceptable. All residues except for the bridging lysine and ten residues on each side were then fixed and the structure sbjected to 10000 cycles of molecular dynamics. The resulting structure did not show a significant change in this region beyond the first two or three residues flanking the lysine (Fig. 17F), indicating that the dimer bound to DNA is an energetically acceptable structure.

This last step ended the process of designing a peptide based on the HTH motif and with a potential for binding to DNA. The next step required the synthesis of the peptide and its testing. The possibility of analysing and redesigning a structure in detail by computer greatly increases the confidence with which this next step was approached, considering the large expenditure in time, effort and resources that it would entail. Attempts have in fact already been made at constructing DNA binding peptides based on the HTH motif of 434 Cro. Grokhovsky *et al.* have recently linked this HTH motif to a multiple lysine core. Their four-stranded peptides, with substitutions rendering the HTH more basic, are quite far removed from the actual structural arrangement in the protein. They find that the peptide does not bind to OR1, OR2 or OR3 of the cognate DNA. Some binding is observed to the pseudo operator OP1, but via a b-hairpin conformation inserted into the minor groove of DNA (64). It is to avoid a similar outcome, and to attempt to design peptides that might take up structures as faithful as possible to that of the analogous fragments in the repressor, that extensive modeling studies were carried out before synthesis.

3.0 BINDING OF PEPTIDES CONTAINING THE HTH MOTIF TO DNA.

3.1 Introduction

DNA-binding peptides based on the Helix-Turn-Helix motif of the λCI 434 repressor were designed, using a minimalist approach. Molecular modelling and structural and homology considerations were used to determine the minimum sequence (chapter 2), including this motif, which might impart biological activity.

The peptides HTH, HTHL, HLDIM, and HTHLDIM (Figure 18) were thus designed, chemically synthesized and characterized. Circular Dichroism was carried out on the peptides to study their conformational features. Their biological activity was assayed via band shift and UV crosslinking studies. Control biochemical experiments such as band shift assays and DNAse footprinting were also carried out in the presence of λCI 434 repressor crude extract.

3.2 Synthesis and characterization of model peptides

As can be seen in Figure 18, the peptides HTH, HTHL, HTHLDIM all include the recognition helix (H3) flanked on the N-terminal by a stabilizing helix and turn (H2T). A loop (L) is added to peptide HTHL. The loop contains residues that form stabilizing interactions with the HTH motif and residues that contact the DNA backbone. Dimerization was built into the peptide by linking two HTHL units via a branched lysine (HTHLDIM). The peptide HLDIM is a truncated form containing only the binding helix.

The four peptides were synthesized using the Fmoc chemistry in automated, continous flow, solid phase. In all cases the syntheses presented difficulties because of the strong tendency of such peptides to give conformational collapse on the resin. This problem was partly overcome by using a polyethyleneglycol/polystyrene (PEG-PS) graft resin, in particular in the case of

(A)

HTH	NQAELAEKVGTTQQSIEQLENG-OH
HTHL	NQAELAEKVGTTQQSIEQLENGKTKRPRFL-OH
HTHLDIM	NQAELAEKVGTTQQSIEQLENGKTKRPRFLP K-G-OH NQAELAEKVGTTQQSIEQLENGKTKRPRFLP
HLDIM	QQSIEQLENGKTKRPRFLP K-G-OH QQSIEQLENGKTKRPRFLP

(B)	HTH	HTHL	HTHLDIM
Measured Molecular Weights*		3414.80	7192.00
Apparent Molecular Weights**	2500	3500	14500
Theoretical Molecular Weights	2387.577	3414.853	7192.048

* FAB-MS **SDS-PAGE

Figure 18. A) Schematic diagram showing the sequences of the synthesized peptides and B) theoretical and experimentally calculated molecular weights of the synthesized peptides.

HTHLDIM, allowing a higher degree of solvation and thus increasing the efficiency of each coupling and the total yeild of the synthesis (see chapter 4.0, sec 4.2.2). The peptides were purified by RP-HPLC, eluting the column with a gradient from A (0.05% TFA in water) to B (90% acetonitrile in A). During the purification process difficulties were faced in solubilizing the peptides. In particular HTH was soluble only in a mixture containing 50% acetonitrile in 0.05% TFA. HTHL, HLDIM and HTHLDIM showed increased solubility in aqueous solvents due to the presence of the basic residues found in the loop (L) which is not present in HTH.

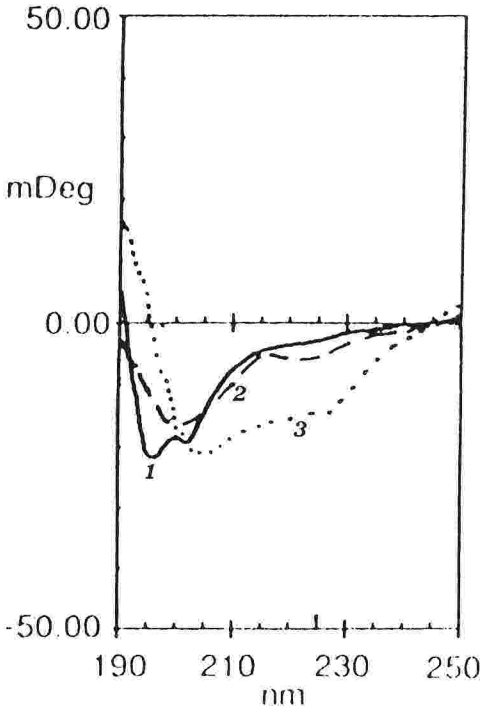
After RP-HPLC purification, HTH, HTHL and HTHLDIM were characterized by amino acid analysis on a custom made system based on the PicoTag method. HTHL and HTHLDIM were also characterized by mass spectrometry. In both cases the molecular weights and the sequence analysis were determined after enzymatic cleavage with trypsin (sec. 4.2.2). Table 4 shows the peptides generated by digestion with trypsin, their calculated molecular weights and observed retention times in the case of HTHLDIM. The molecular weights of HTHL and HTHLDIM experimentally determined by mass spectrometry coincide with the theoretically expected ones as shown in Figure 18 B.

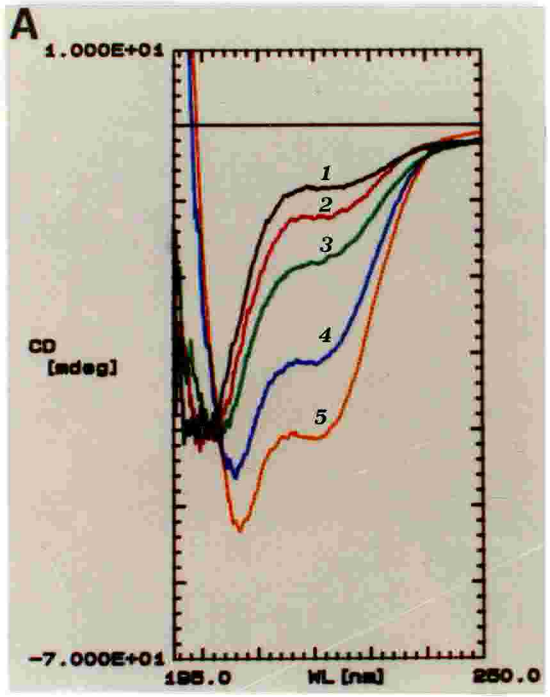
3.3 CD spectroscopy

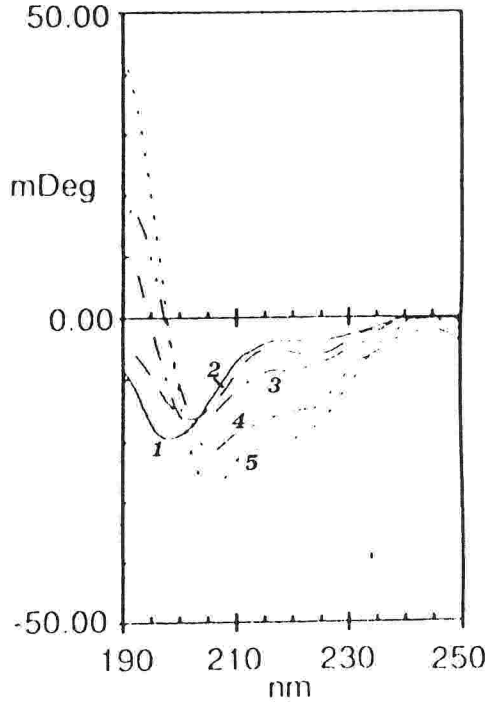
Circular dichroism was carried out on HTH, HTHL and HTHLDIM to study their tendency to fold into the appropriate α-helical conformation required for binding to DNA.

In the case of HTH the CD spectrum in aqueous solution (Figure 19) indicates that the peptide is in a random coil conformation. This is as expected for a relatively short peptide. At 10% triflouroethanol (TFE), a solvent that mimics an organic environment such as that found in the DNA major groove, a change in the spectrum may already be observed, whereas one can clearly observe the

Figure 19. CD spectra of HTH (A) HTHL (B) HTHLDIM (C) peptides (0.1mg/ml) in 10mM phosphate buffer with increasing quantities of TFE.







A

1: HTH (0.1mg/ml) in water
2: " in 10% Trifluoroethanol
3: " in 33% Trifluoroethanol

B

1: HTHL (0.1mg/ml) in water
2: in 10% Trifluoroethanol
3: " in 15% Trifluoroethanol
4: " in 25% Trifluoroethanol
5: " in 33% Trifluoroethanol

C

1: HTHLDIM (0.1mg/ml) in water
2: " in 10% Trifluoroethanol
3: " in 15% Trifluoroethanol
4: " in 30% Trifluoroethanol
5: " in 50% Trifluoroethanol

typical spectrum of an α-helical peptide when the percentage of TFE is raised to above 25%. This indicates that the peptide folds to a structure with a high α-helical content in an organic environment, which is consistent with its structure inside the CI434 N-terminal domain. Similarly the CD spectra of HTHL in the presence of TFE (Figure 19) show clearly that it also folds into an α-helix at above 25% TFE. Figure 19C shows the CD spectra obtained for HTHLDIM at increasing concentrations of TFE. The behaviour of this peptide is entirely analogous to that of the other two peptides, indicating that the dimerization process, effected by covalent binding of two HTHL units to the α and ε amino groups of a lysine, has not had a deleterious effect on folding. That some form of inter-strand interaction might occur precluding a correct folding had been reason for concern. The CD spectra of all three peptides indicate that in water or phosphate buffer all peptides are essentially present as random coils. The λCI 434 repressor *in vivo,* being a fully fledged protein, obviously exists in a folded form, but even in its case, it only achieves a complete three dimensional arrangement of its α-helical units in the presence of cognate DNA (45).

A change to the CD spectrum of the peptide, in the presence of a cognate DNA supporting an α-helical arrangement, would indicate that an interaction is occurring. Preliminary work has been carried out with HTHLDIM in the presence of the 16-3(298 bp) cognate DNA, but technical problems were faced in that the absorption of DNA was too strong at the concentration required for a reasonable binding of HTHLDIM (at least a ten-fold excess of DNA). This is clearly a difficulty as the CD spectra are measured as a difference in the absorption for alternatively right and left polarized light, and the signal to noise ratio becomes very high at increased absorbances. Use of shorter oligonucleotide fragments containing the binding site would partly resolve the problem, but as will be discussed later, binding to these is rather labile. This type of experiment will have to await an optimized peptide which will bind strongly to the cognate DNA, thus enabling binding to shorter DNA fragments and at a lower excess. It is important

to observe the CD behaviour in the presence of DNA as it is possible that binding does occur but in a manner different to that of the HTH motif *in vivo*. In fact, one group that attempted to bind a model peptide based on the HTH motif to DNA claim that it does so as a β-hairpin structure, in the minor groove of a pseudo-operator site (64).

3.4 Binding assays

Different types of binding assays were attempted, which consisted mainly in determining whether the peptides induced a band-shift in DNA or oligonucleotides containing one or more operator sites. To test for non-specific binding to DNA, the effect of competing DNA that does not contain the operator site (in this case poly[d(I-C)]), was determined. Given that the design of the peptides was appraoched from a minimalist point of view, and that a considerable part of the protein with its contribution to the overall affinity for DNA had been eliminated, it was probable that binding might well be occurring in a specific manner but with a low binding constant. It might not in this case be possible to observe binding under the conditions for band-shift experiments. Experiments were thus designed in which the peptides would be permanently covalently linked to the DNA, at the binding site, via UV cross-linking. This would trap the peptide/DNA complex and allow observation of a band-shift.

3.3.1 Band shift assays

Band shift assays were carried out in the presence of HTH, HTHL and HTHLDIM with a series of DNA probes based either on the CI434 operator site, or the equally recognized phage 16-3 repressor operator sites. These probes were

either obtained by chemical synthesis of oligonucleotides, or via insertion of oligonucleotides into plasmid DNA. The probes (see sec. 4.1 for preparation) are as follows:

1) 434OR1(259 bp).	Contains the 434 OR1 operator cloned into pBR322, and excised as the NheI-HindIII fragment.
2) 16-3(298 bp).	Contains the (OR-OR′) operator from the C 16-3 repressor, cloned into pTZ19U, and excised as the EcoRI-Sal1 fragment.
3) 16-3OR(32 mer).	Contains the 16-3 OR operator and flanking sequence, chemically synthesized.
4) 16-3OR(*32 mer)	As in (3) but with bromouridine substituting thymidines
5)16-3OR(16 mer).	Contains the 16-3 OR operator, chemically synthesized
6)[16-3OR]n(32 mer).	Enzymatically concatenated 16-3OR(32 mer) oligonucleotdes designed so as to have helically phased operator sites.
7) [16-3OR]n(16 mer).	Enzymatically concatenated 16-3OR(16 mer) designed so as to be non-helically phased with respect to the operator sites.
8) 16-3[OR-OR′](45 mer).	Contains the 16-3 (OR-OR') operator region, chemical synthesized and Klenow filling from two fragments. Thymidines are substituted with bromouridine.

Band shift experiments performed with 16-3OR(16 mer), 16-3OR(32 mer) or 434OR1(259 bp) in the presence of all four peptides have as yet failed to show any retardation of these DNAs. This either indicates that the binding is not occurring in any case, or that in cases where it does occur binding is weak and does not survive the conditions for band-shift experiments. Care was taken to

incubate the peptides with the cognate DNA at 4C, and to electrophorese them under mild conditions in a cold room. However as the peptides are slightly basic, it is possible that they were swept away from the negatively charged DNA under electrophoretic conditions if binding was weak, or reversible. Control experiments in the presence of 16-3(298 bp) are shown in Figure 20 and all other operator models (data not shown) with λCI 434 repressor crude extract showed retardation in all cases, so that all operator models were viable. Results of band-shift experiments with HLDIM were inconclusive as this tended to form an aggregate with the DNA, that remained in the well.

Band shift assays in the presence of 16-3(298 bp) show retardation only in the case of HTHLDIM (Figure 21) which tends to disappear when increasing amounts of competitor DNA, poly[d(I-C)], are added. It is not observed for HTH (data not shown) nor is it detected in the case of HTHL (Figure 21). Band-shifts cease to be observed for HTHLDIM only when a greater than 100 fold excess of aspecific competitor is added, indicating that the peptide binds in a specific manner to the cognate DNA.

Binding reactions were carried out with concatenated oligonucleotides. These were constructed by enzymatic ligation of oligonucleotides 16-3OR(32 mer) or 16-3OR(16 mer) so as to obtain respectively [16-3OR]n(32 mer) and [16-3OR]n(16 mer) where $1 \leq n \geq 15$. In the first instance the oligomers were so designed to be helically in phase with respect to the operator sites (i.e. these are separated by exactly one turn of the DNA). Figure 22A clearly shows retardation in the case of HTHLDIM and this increases with increasing n. No binding occurs with HTHL (data not shown). No shift is observed (see Figure 22B) when experiments are performed with HTHLDIM in the presence of [16-3OR]n(16 mer) which was enzymatically constructed so as to be non helically phased with respect to the operator site (i.e. separated by half a turn of the DNA duplex).

From the results described above, the following considerations can be made:

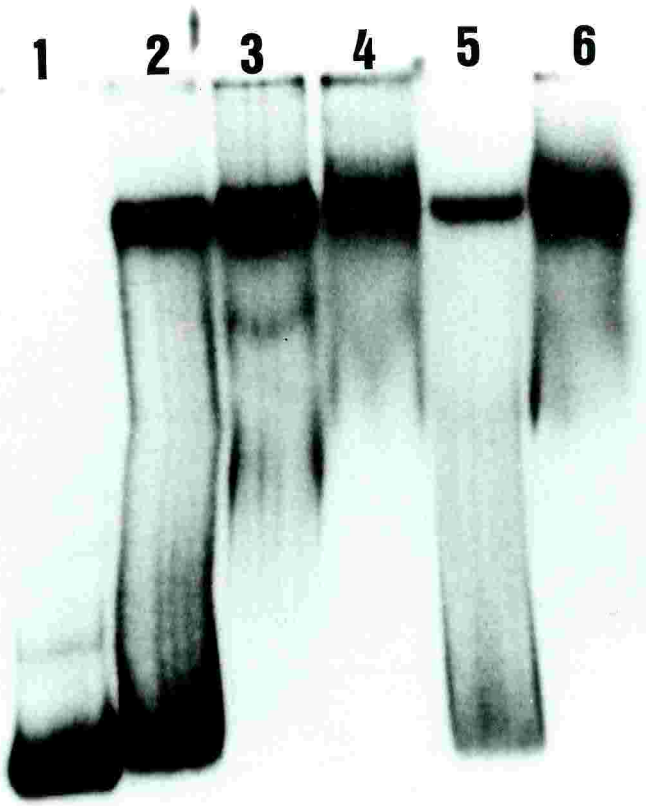


Figure 20. Band shift assay performed with 16-3 operator (298 bp) and CI 434 crude extract. Lane 1, no protein; Lanes 2-6, increasing quantities of CI 434 repressor crude extract with 1.2 μM poly[d(I-C)].

1) Binding does not occur for the monomeric HTH or HTHL to any of the operator models.

2) The dimeric HTHLDIM does not bind to those models which include only one operator site (i.e. 16-3OR(16 mer), 16-3OR(32 mer) and 434OR1(259 bp).

3) The dimeric HTHLDIM binds to a cognate DNA fragment containing two operator sites [i.e. the 16-3(298 bp) fragment].

4) This peptide binds to concatemers containing multiple operator sites only when these are in phase, and binding appears to increase as the number of operator sites increases.

5) It does not bind to concatemers containing multiple operator sites when these are out of phase, regardless of the number of operator sites present.

From these one could conclude that binding only occurs for a dimeric form of the model peptide, and that in any case does not occur with a high binding constant, although it appears to be specific. A degree of cooperativity is indicated by the fact that binding occurs to a fragment containing two operator sites [16-3(298bp)] but not a fragment of similar length with only one operator site [434OR1(259bp)]. Furthermore, stronger shifts are apparent for concatemer fragments with a higher number of operator sites. One could argue that this is expected due to a greater number of bound peptides, but the fact that the out of phase concatemer does not show any band-shifts even for equivalent n values, supports a cooperative binding of the peptide.

The cooperativity could derive from two distinct effects, i) an interaction of an incoming peptide with one already bound to the DNA, aiding the binding of the former, or ii) a conformational change in the DNA which increases the avidity of the second peptide to the DNA, or from both these effects. In the case of the protein, it is known that binding of one repressor to an operator site (say OR1) favours the binding to the adjacent operator (OR2), but this does not extend to OR3. In that case we are however dealing with quite large entities, so that a

contact between adjacently bound proteins is not unreasonable, especially as the DNA to which they are bound is bent. It is difficult to see how this could be extended to quite small peptides. However, the fact that binding occurs to concatemers that have in-phase binding sites (i.e. all bound peptides face the same way on the DNA) and not in out-of-phase concatemers (where peptides alternate on different sides of the DNA duplex) would indicate this type of cooperativity. Another possibility is that the bendability of the DNA is increased when the binding sites are in-phase, thus aiding binding.

3.4.2 UV crosslinking

To test the binding specificity of HTHLDIM for its cognate DNA while overcoming the problem of weak binding, band shift assays were carried out with 16-3OR(*32 mer) and 16-3[OR-OR'](45 mer). In this case the oligonucleotides were synthesized with 5-bromodeoxyuridine in the place of thymidine so as to use them as substrates for UV crosslinking in the presence of both HTHLDIM and HTHL. This is a general methodology in order to induce covalent adduct formation (sec. 4.3).

Strong adduct formation is observed when the binding reaction is performed with 16-3[OR-OR'](45 mer) in the presence of HTHLDIM (Figure 23A). This is direct evidence of the binding of this peptide to the cognate DNA in the case when two operator sites are present. It should be mentioned that the two binding sites were in-phase with respect to the DNA duplex, being an exact replica of the DNA in the phage genome.

In accordance with the conventional band-shift experiments described in the previous section, no adduct formation is observed when the reaction is carried out for this peptide with 16-3OR(*32 mer), as can be clearly seen in Figure 23B. This is direct evidence of the importance of the cooperative effect.

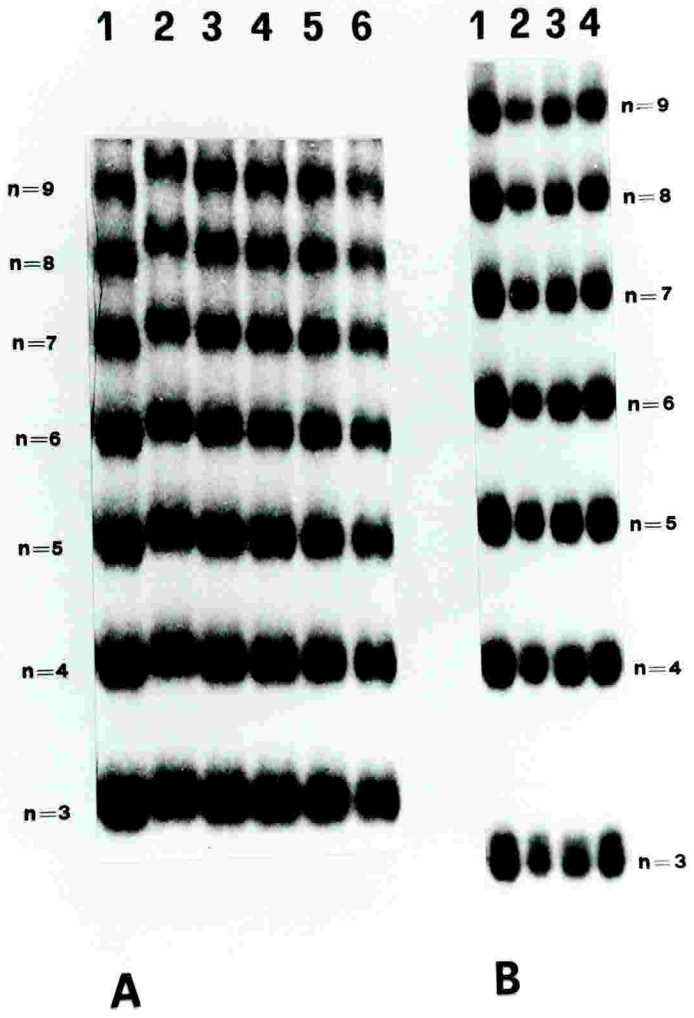


Figure 22. (A) Band shift assay performed with HTHLDIM and $[16\text{-}3OR]_n$ obtained by enzymatic ligation of 16-3OR(32-mer). The resulting concatamers are helically phased with respect to the operator site. Retardation is observed only when n>4. Lane 1, no peptide; Lanes 2-6, HTHLDIM with increasing quantities of poly[d(I-C)]. (B) Band shift assay performed with HTHLDIM and $[16\text{-}3OR]_n$ obtained by enzymatic ligation of 16-3OR(16-mer). In this case the concatenated oligonucleotides are not helically phased with respect to the operator site. No retardation is observed. Lane 1, no peptide; Lanes 2-4, HTHLDIM with increasing quantities of poly[d(I-C)].

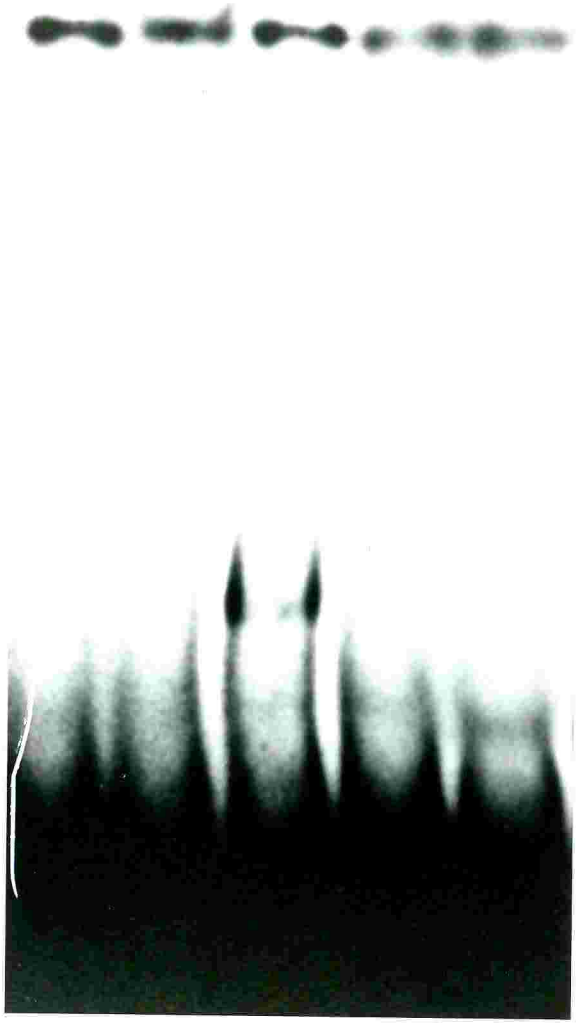


Figure 21. Band shift assay performed with 16-3 operator (298 bp) and HTHL or HTHLDIM.Retardation is obsrved only in the case of the HTHLDIM peptide. A 10-fold excess of both peptides was used with respect to the cognate DNA and the binding reactions were carried out in the presence of poly[d(I-C)] as competitor. Lane 1, HTHL with 1.2nM poly[d(I-C)]; Lane 2, HTHL with 0.12mM poly[d(I-C)]; Lane 3, HTHLDIM with 1.2nM poly[d(I-C)]; Lane 4, HTHLDIM with 0.12mM poly[d(I-C)]; Lane 5, no peptide.

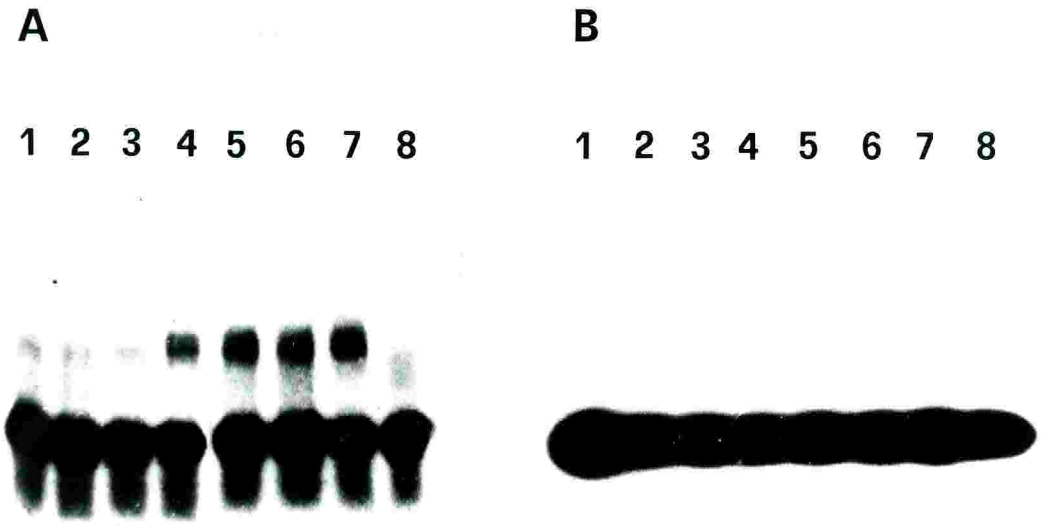


Figure 23. (A) UV crosslinking of the 16-3[OR-OR'](45-mer) to HTHLDIM. Lanes 1-7, DNA probe with increasing quantities of HTHLDIM; Lane 8, no peptide. (B) UV crosslinking of the 16-3OR(32-mer) to HTHLDIM. Lanes 1-7, DNA probe with increasing quantities of HTHLDIM; Lane 8, no peptide.

No adduct formation was observed with both DNA models in the presence of HTHL (data not shown), consistent with the results obtained in conventional band shift assays. This confirms the importance of subunit dimerization in the binding process.

Furthermore the molecular weight of the shifted bands observed in crosslinking experiments with HTHLDIM were estimated by comparison with molecular weight standards, after staining with Coomassie. The estimated MW for the complex of two HTHLDIM with 16-3[OR-OR′](45 mer) is approximately 40KDa, and the complex band falls at a position with respect to the standards consistent with this weight. This supports a 2:1 stoichiometry for the interaction. Also, the fact that there are two molecules of peptide for two operator sites suggests that the binding of HTHLDIM to its operator sites is specific.

3.3 Preliminary footprinting experiments.

Footprinting of DNA with bound peptide would provide direct evidence of binding and specificity. Preliminary footprinting experiments have been carried out using the DNAse method. These experiments were first performed in the presence of λCI 434 crude extract in the presence of the 16-3(298 bp) cognate DNA. The footprint pattern (data not shown) shows the presence of shaded areas in the regions corresponding to the operators cloned into 16-3(298 bp) thus indicating that the repressor protein present in the crude extract probably binds in a sequence specific manner. The boundaries of the protected regions are not well defined and this is probably due to a low frequency of enzyme cutting in these regions. Preliminary DNAse I footprinting experiments were also carried out with 16-3(298 bp) in the presence of HTHLDIM. Apparently regions corresponding to the operators are protected from DNAse I cleavage (data not shown). However also in this case the boundaries of these protected regions are not well defined and for this reason densitometric scanning of the autoradiographs is needed for a

clear-cut in this regard. In the meantime an alternative hydroxy radical based footprinting technique with higher resolution is being developed in order to optimize the experimental findings both in the case of the λCI repressor crude extract and for HTHLDIM and establish a clear footprint pattern.

However two main conclusions can be drawn from the preliminary footprinting results. In the case of HTHLDIM there is some further evidence indicating that this peptide binds its DNA in a sequence specific manner. Secondly, the footprint patterns of the λCI 434 repressor crude extract coupled to the earlier experimental evidence obtained with band shift assays with 16-3(298 bp) support the evidence of related repressor specificity of unrelated phages. In other words it demonstrates in a direct manner that the λCI 434 repressor capable of recognizing the phage 16-3 operator.

4.0 MATERIALS AND METHODS

4.1 DNA binding sites

Various models for the the λCI 434 repressor operator sites were considered and used, as listed below:

1) 434OR1(259 bp).	Contains only the OR1 operator cloned into it.
2) 16-3(298 bp).	Contains the full operator (OR-OR′) which is the binding site of the C 16-3 repressor.
3) 16-3OR(32 mer).	Contains the OR operator of phage 16-3 only.
4) 16-3OR(*32 mer).	As in (3) but with bromouridine substituting thymidines
5) 16-3OR(16 mer).	Contains the 16-3 OR operator.
6) [16-3OR]n(32 mer).	Enzymatically concatenated 16-3O(32 mer) oligonucleotdes designed so as to have helically phased operator sites.
7) [16-3OR]n(16 mer).	Enzymatically concatenated 16-3OR (16 mer) designed so as to be non-helically phased with respect to the operator sites.
8)16-3[OR-OR′](45 mer).	Contains the complete operator region ofthe C 16-3 repressor. Bromouridine substitutes thymidine.

4.1.1 Preparation and characterization of the 434 OR1 site

Cloning of the 434 OR1 site [434OR1(259 bp)]

Complementary oligonucleotides containing the DNA binding sites were designed and synthesized on an Applied Biosystem model 380A DNA synthesizer. After HPLC characterization, and purification through denaturing

polyacrylamide gel electrophoresis, they were annealed in water (using equimolar quantities of each complementary oligonucleotide) by heating to 95°C and then cooling overnight

The blunt ended 29 base pair oligonucleotide containing the 434 operator was then cloned into pBR322, maintaining both the ampicillin and the tetracyclin resistance genes active. For this purpose, the plasmid was first linearized with the EcoRV restriction enzyme which leaves blunt ends and then the double stranded oligo containing the 434 operator was enzymatically ligated by means of the T4 Polynucleotide Ligase, according to the protocol in Maniatis (65).

Cell transformation and preparation of plasmid pBR322 containing the 434OR1

The sample plasmid DNA was incubated with competent E.coli DH5 bacteria in ice and then heat-shocked at 42°C in a water bath. After incubation at 37°C in Luria-Bertani (LB) medium enriched with ampicillin (100 µg/ml), bacteria were plated in LB/ampicillin agar and incubated overnight at 37°C. Replica plating was then carried out by plating the transformed E.coli DH5 cells to LB/agar plates enriched with tetracyclin and incubated overnight at 37°C. Non-matching DH5 E.coli colonies were selected as the ones in which transformation had occurred successfully. Small preparations were carried out according to the following procedure (66). A 3 ml culture was grown overnight in LB medium enriched with ampicillin (100 µg/ml) at 37°C. 1.5 ml aliquots were microcentrifuged at 4°C and the pellets were resuspended and incubated in 110 µl of STETL (8% Sucrose, 5% Triton X-100, 50 mM EDTA and 0.5 mg/ml lysozyme) at room temperature. After boiling for two minutes, the sample was microcentrifuged at 4°C for 15 minutes and then the pellet was removed by means of a toothpick and discarded. RNAse treatment was carried out on the supernatant at 37°C for 30 minutes in order to remove cellular RNA. 110 µl of isopropanol were then added and the sample immediately microfuged for 15 minutes. The

supernatant was removed, the pellet dissolved in water and the plasmid DNA extracted with Phenol/Chloroform in order to remove all cellular contaminants and precipitated with 100% cold ethanol. Plasmid DNA extracted from different non-matching DH5 E.coli colonies was then digested with the EcoRV restriction enzyme to assay which of these such colonies underwent transformation by pBR322 containig the 434 OR1 operator in the EcoRV site. After overnight digestion at 37°C with EcoRV the samples were checked on a 1% agarose gel.

Large-scale palsmid DNA preparation was carried out only on transformed samples (67). Transformed DH5 E.coli cells were grown overnight at 37°C in 250ml of LB agar medium enriched with ampicillin (100 μg/ml). Cells were centrifuged at 5000 RPM, 4°C, for 15 minutes and the pellet was resuspended in 6 ml of freshly prepared buffer containing 25 mM TRIS.HCl (pH=7.5), 10 mM EDTA, 15% Sucrose and 2 mg/ml lysozyme. Cells were resuspended by rapid repipetting and the suspension was incubated in ice water for 20 minutes. 12 ml of 0.2 M sodium hydroxide with 1%SDS were added to the suspension, mixed carefully and thoroughly by inversion and then incubated in ice water for 10 minutes. 7.5 ml of 3 M sodium acetate, pH=4.6, were added and the suspension was further incubated in ice water for 20 minutes. Samples were centrifuged at 15000 RPM for 15 minutes and supernatant was removed to another tube while carefully avoiding the white precipitate. 50 μl of RNAse A (1 mg/ml) were added to the supernatant which was then incubated for 30 minutes at 37°C. The samples were then extracted with an equal volume of phenol/chloroform (v/v 1/1) and two volumes of cold 100% ethanol were added to precipitate the DNA. The resulting pellet was dissolved in 1.6 ml water, 0.4 ml of 4 M sodium chloride and 2 ml of 13% PEG (average molecular weight ca. 8000). The samples were incubated in ice water for 60 minutes and then centrifuged at 10000 RPM for 10 minutes. Supernatant was removed and the pellet washed with 70% ethanol.

Maxam-Gilbert Sequencing of 434OR1(259 bp)

Characterization of the cloned 434 operator was carried out by the chemical sequencing of a small fragment obtained by enzymatic digestion of pBR322 with restriction enzymes EcoRI and NheI. 17 µl of plasmid pBR322 (2.6 mg/ml) containing the 434 operator were overnight incubated with 10 µl of NheI (5000 U/ml) and 3 µl of EcoRI (20000 U/ml) at 37°C.

The digestion was checked on a 1% agarose gel. The bands corresponding to the EcoRI-NheI fragment (259 bp) containing the 434 OR1 operator were excised from the agarose gel and extracted according to the protocol provided in the USBioclean kit. The sample DNA was incubated with 2 µl of Calf Intestine Alkaline Phosphatase (1000 U/ml) for 30 minutes at 37°C twice. The dephosphorylated DNA was then extracted with phenol/chloroform (v/v 1/1), precipitated with cold 100% ethanol and the pellet dissolved in 19 µl of water. The labelling reaction was then carried out by the T4 Polynucleotide Kinase in the presence of 5 µl of γ-P-ATP (10 µci/µl) and the sample was incubated at 37°C for 45 minutes according to Maniatis. The labelled DNA was then extracted with phenol/chloroform and precipitated with cold 100% ethanol. The pellet was dissolved in 34 µl of water and incubated overnight at 37°C with 2 µl of the HindIII (20 U/µl) restriction enzyme in order to provide a 5′ end-labelled DNA fragment. The product of the enzymatic digestion was then checked on an agarose gel and the bands corresponding to a 259 bp DNA fragment were excised and the DNA extracted according to the USBioclean protocol. The yeild of labelling was assayed by liquid scintillation which counted 772660 CPM of radioactive material.

The sample DNA was then chemically sequenced according to the Maxam-Gilbert strategy (68). In this method the bases are first specifically modified and then the backbone is cleaved at the sites of modification by treatment with pyperidine at high temperature.

1) Cleavage at G residues

200 µl of DMS (dimethyl sulphate) buffer (i.e. 50 mM Sodium cacodylate pH= 8.0, 10 mM magnesium chloride, 1 mM EDTA) were added to 10 µl of P-labelled NheI-HindIII fragment containing the 434 OR1 binding site. The mix was chilled in ice for a few minutes and then 1 µl of DMS was added and the sample incubated at room temperature for 8 minutes. 50 µl of DMS stop solution (1.5 M sodium acetate pH=5.0, 1 M β-mercaptoethanol, 100 µg/ml t-RNA) were added and the methylated DNA was then precipitated with 750 µl of cold 100% ethanol in dry ice for 15 minutes.

2) Cleavage at G+A residues

10 µl of the labelled DNA fragment (ca. 20000 CPM) containing the 434 operator site were brought to 20 µl with water and chilled in ice. 50 µl of 88% formic acid were then added and the reaction allowed to proceed at room temperature for 8 minutes. The reaction was stopped by adding 200 µl of stop solution (0.3 M sodium acetate pH=5.0, 0.1 M EDTA, 25 µg/ml t-RNA) and the sample precipitated with 750 µl of cold 100% ethanol in dry ice for 15 minutes.

3) Cleavage at T+C residues

10 µl of the fragment (ca. 20000 CPM) containing the 434 operator site were brought to 20 µl with water and chiled in ice.

30 µl of hydrazine were added and the sample was incubated at room temperature for 8 minutes. The reaction was stopped by adding 200 µl of stop solution (0.3 M sodium acetate pH=5.0, 0.1 M EDTA, 25 µg/ml t-RNA) and the sample DNA precipitated with 750 µl of cold 100% ethanol in dry ice for 15 minutes.

4) Cleavage at C residues

15 µl of 5 M sodium chloride were added to 10 µl of the fragment (ca. 20000 CPM) containing the 434 operator site and the sample was chilled in ice. 30 µl of hydrazine were added and the sample was incubated at room temperature for 8 minutes. The reaction was stopped by adding 200 µl of stop solution (0.3 M sodium acetate pH=5.0, 0.1 M EDTA, 25 µg/ml t-RNA) and the sample DNA precipitated with 750 µl of cold 100% ethanol in dry ice for 15 minutes.

The four samples were then washed with 70% ethanol, precipitated in dry ice and lyophilized. 100 µl of 1 M piperidine were added to each sample, cleaving reactions carried out at 90°C for 30 minutes. After quenching the reactions by chilling in ice and brief spinning, the piperidine-cleaved DNA samples were precipitated with 330 µl of cold 100% ethanol with 10 µl of 3 M sodium acetate in dry ice for 15 minutes. After 15 minutes microfuging at 4°C and removal of supernatant, the samples were precipitated with 250 µl of 0.3 M sodium acetate and 750 µl of cold 100% ethanol in dry ice for 15 minutes. After further microfuging for 15 minutes at 14000 RPM, 4°C, the supernatant was removed and the pellets lyophilized. The pellets were dissolved in a denaturing buffer containing 80% deionized formamide, 10 mM sodium hydroxyde, 1 mM EDTA and 0.1% bromo phenol blue/xylene cyanol, denatured in a water-bath at 90°C and immediately cooled down in ice.

The four samples were then loaded on a 6% polyacrylamide gel and electrophoresed at 1800Volts. Figure 24 A shows the G reaction and the G+A reaction carried out on the 434OR1(259 bp) fragment (see Figure 25 A for the complete sequence of the same DNA fragment).

PCR amplification of the NheI-HindIII pBR322 fragment containing the 434 OR1 site

PCR amplification of the cloned fragment containing the434 operator OR1 was carried out using the standard TAQ polymerase method since it is ideal for automated polymerase chain reaction. For this purpose two 18-mer primer oligonucleotides were designed, 5′-TTCTCATGTTTGACAGCT (primer 1) and 5′-AGTAGCGAAGCGAGCAGG (primer 2), respectively complementary to short sequences of pBR322 contiguous and 5′ to the inserted 434 operator, on each strand. They were designed in order to frame a sequence of about 400 bp bearing the 434 operator approximately in the middle.

The polymerase chain reactions were set up in a total volume of 50 µl in which 0.5 µl of TAQ polymerase (5 U/µl) were added to a mixture containing 1 µl of sample plasmid DNA (0.4 ng/µl), 1 µl of 20 mM dNTPs, 2 µl DMSO, 5 µl 10 times PCR buffer (100 mM TRIS.HCl pH=8.3, 500 mM potassium chloride, 80 mM magnesium chloride, 0.01% gelatin), 2µl of primer 1 (20 pmoles/µl) and 2 µl of primer 2 (20 pmoles/µl). 15 µl of sterile mineral oil were added to overlay the reaction volume in order to avoid evaporation during the high temperature steps of the chain reaction. For this purpose the sample DNA was first denatured for 5 minutes at 95°C and then the temperature profile per cycle was set up as follows:

(1) T=94°C 1 minute Denaturation

(2) T=50°C 1 minute Annealing

(3) T=72°C 1 minute Chain elongation

The cycle was iterated 30 times, the reactions being carried out in an automated PCR machine (Perkin Elmer 480 Thermocycler) for about three hours. The annealing temperature was derived from the approximate melting temperatures of the two primer oligonucleotides, evaluated according to the equation, Tm=4(G+C)+2(A+T), where G+C and A+T represent the base composition of each oligonucleotide (69, 70, 71).

The amplified samples were then electrophoresed on a 1% agarose gel and showed a high level of purity. Bands were excised, the sample DNA extracted from the agarose gel according to the USBioclean protocol and UV spectrophotometrically quantitated.

4.1.2 Preparation and characterization of the 16-3 operator sites

Cloning of the 16-3 operator region containing OR and OR' [16-3(298 bp)]

The 16-3 operator region was kindly donated to us by G. Dallman (54). It had been cloned as a Sau 3A fragment (S280) obtained from the EcoRI-N fragment of 16-3 DNA into a truncated form of pBR328. In this vehicle both the tet and CAT resistance genes are disrupted by Bal 31 digestion starting from the EcoRV site whereas the amp resistence gene and the EcoRI site remain intact. Sau 3A fragment was excised and cloned in M13 by ligating it between the EcoRI and the SalI sites. The fragment was then excised and put into the BamHI site of the high copy number (70-200) plasmid pTZ19U.

Cell tansformation and preparation of plasmid pTZ19U containing the 16-3 operator region (OR, OR')

Competent E.coli DH5 cells were transformed with pTZ19U containing the 16-3 operator region according to the procedures already described above in section 4.1.1 for the preparation of the 434 OR1 site.

Maxam-Gilbert sequencing of the cloned 16-3 operator region into pTZ19U

60 μg of sample plasmid DNA were enzymatically digested overnight with EcoRI (20 U/μl) at 37°C and the digestion product, after extraction and

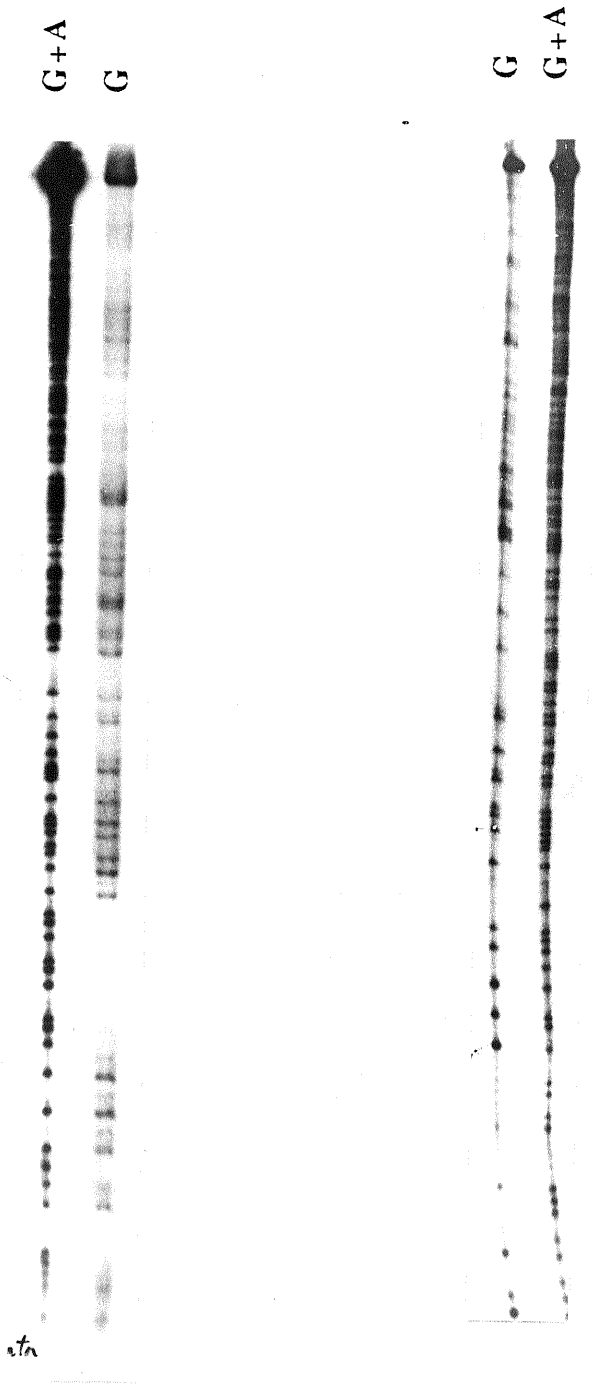


Figure 24. Chemical sequencing of DNA fragments 16-3(298 bp) (A) and 434OR1(259 bp) (B). In both cases only the G reaction and the G+A reaction are shown.

A)

5'-

TTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCT

CGTCATCCTCGGCACCGTCACCCTGGATGCTGTGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATAAGCTTCA**ACAA**TTGTAG**TTGT**ACGAGT

CATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGT-3'

------**OR1**-----

B)

5'-

GATCGCCTGTTGGCTTTGACCAACGCGCCGTGCGAGTTCGGCTTGCGTCTCCTGCCTCTTCCAGCTTCGCGGCCAGAGTGTCTGTTAATCGGCTGTGAAATGTCCCT

TTATGCATCAACAACTTACCATGGCCTACAAAAAGGCA**AACAA**TGGTAC**TTGA**CGACTCATCAC**AACAA**TTGTAG**TTGT**AGATTGTAACGAACGGATGTACGAAG

TTCCGACAGACAAAAAAGAAAAAAGCCAGCAGGGCGCTAGACCCAAAGAACGGCAAAGGAGTG

------OR'-------

------OR-------

Figure 25 A) DNA sequence of the 434OR1(259 bp) and B) DNA sequence of 16-3(298 bp)determined by Maxam and Gilbert sequencing.

purification from a 1% agarose gel, was enzymatically dephosphorylated by Calf Intestine Alkaline Phosphatase (1000 U/ml) at 37°C for 30 minutes, the procedure being carried out twice. After phenol/chloroform extraction the sample DNA was enzymatically 5′-end labelled with 8 µl of γ-32P-dATP (10 µci/µl) by using 3 µl of T4 polynucleotide Kinase (5 U/µl), 37°C, for 45 minutes. After phenol/chloroform extraction and 100% ethanol precipitation to remove unincorporated ATP, the sample DNA was enzymatically digested overnight with 3 µl of SalI (10 U/µl) at 37°C and the resulting small DNA fragment, containing the 16-3 operator region and labelled only at the 5′ EcoRI end, was excised from a 1% agarose gel, purified and quantified by liquid scintillation, the yeild being about CPM.

The sample DNA was then sequenced by the Maxam and Gilbert method (68) as described above in section 4.1.1 for the 434 OR1 cloned into pBR322 (Figure 24 B and Figure 25 B).

PCR amplification of the EcoRI-SalI fragment of pTZ19U containing the 16-3 operator region

The primers for the polymerase chain reaction (69, 70, 71) were selected according to the polycloning region of this pUC9 derivative which allows the use of universal prmers, namely a reverse primer located in the lac operator of pUC19 (5′-TGAGCGGATAACAATTTCACAC) and a pKO primer in pUC19 (5′-GTAACGCCAGGGTTTTCCCAGT).

The reaction was set up in a total volume of 50 µl in which 5 µl of pTZ19U (0.44 ng/µl) were added in the presence of 1 µl of dNTPs (20 mM), 2 µl of DMSO, 5 µl of 10 times concentrated PCR buffer (100 mM TRIS.HCl pH=8.3, 500 mM potassium chloride, 80 mM magnesium chloride, 0.01% gelatin), 2 µl of reverse primer (20pmoles/µl), 2 µl of pKO primer I (20 pmoles/µl) and everything brought to volume with sterile water. Then 0.5 µl of TAQ polymerase

(5 U/μl) were added to start the reaction. 15 μl of sterile mineral oil were added to overlay the reaction volume and then PCR was initiated by first denaturing the sample DNA at 95°C for 5 minutes. The temperature profile per PCR cycle was set up to be as follows:

(1) T=94°C 1 minute Denaturation

(2) T=58°C 1 minute Annealing

(3) T=72°C 1 minute Elongation

and was iterated for 35 cycles.

The annealing temperature was calculated on the basis of the nucleotide composition of each primer oligonucleotide by using the equation Tm=4(G+C)+2(A+T). The DNA obtained from PCR was electrophoresed on a 1% agarose gel, extracted and then spectrophotometrically quantitated.

Preparation of concatenated oligonucleotides containing the 16-3 OR site

The 32-mer 5′-TAAGCTTCAACAATTGTAGTTGTACGAGTCAC (A strand) was synthesized together with its complementary (B strand) 5′-CTTAGTGACTCGTACAACTACAATTGTTGAAG oligomer and the 16-mer 5′-TTGACAACTACAATTG (C strand) also synthesized with its complementary counterpart 5′-CAACAATTGTAGTTGT (D strand). In all cases automated synthesis was performed on an Applied Biosystem model 380A DNA synthesizer (chemistry of phosphoramidites) and the oligonucleotides were loaded on a urea denaturing polyacrylamide gel electrophoresis. The pure oligonucleotides (0.6 μg) were labelled separately in a total volume of 20 μl (70 mM Tris-HCl pH=7.6, 10 mM magnesium chloride and 5mM dithiothreitol) with 10 μCi of γ-32P-dATP (SA>3000 Ci/mmol) and 6 units of T4 polynucleotide kinase, at 37°C for 30 minutes. The labelled and unlabelled complementary strands were mixed

according to a concentration ratio of 1:2, respectively. The annealing of the complementary oligonucleotides was carried out by heating the mixtures up to 90 °C for few minutes and then allowing a very slow cooling down so as to form the hybrids.

The double-stranded oligonucleotides were enzymatically ligated so as to produce helically phased or non-helically phased concatemers with respect to the operator site. These concatenated oligonucleotides (72) were respectively named 16-3[OR-OR'](32 mer) and 16-3OR(16 mer). 2 µl of the mixtures were used in the ligation reactions with 1 unit of T4 DNA ligase in a total volume of 10 µl (70 mM Tris-HCl pH=7.6, 10 mM magnesium chloride and 5 mM dithiothreitol) and 2.3 nM cold ATP. The ligation reaction was carried out for 4-5 hours at room temperature and the concatemers were then run on 8% polyacrylamide gels in the presence of molecular weight markers.

4.2 Synthetic peptides

4.2.1 Synthesis and characterization of the HTH peptide

Synthesis protocol

The synthesis of HTH (Figure 18 A.) was performed by the automated continous solid phase method on a Milligen 9050 Pepsynthesizer. *In situ* activated chemistry (HOBt/TBTU) was used in the presence (73) of a 4 molar excess of Fmoc-AA-OH and double coupling at every step. The peptide chain was elongated on 1.2 g of the acid labile Fmoc-Gly-Pepsyn-KA resin (Keiselguhr, Milligen).

Cleavage and deprotection

After completion of the synthesis, the resin was washed respectively with isoamyl alcohol, acetic acid, isoamyl alcohol and diethyl ether and then freeze-dried. After lyophilization the peptide was cleaved from the resin by treatment with a solution containing TFA/Phenol/EDT/TA/Tris-isopropylsilane/water (90:2:2:2:2:2 %) and the reaction allowed to proceed for about two hours at room temperature. The resin was then filtered on a syntered funnel and washed with the cleavage mixture and then few times with TFA. The filtrate was collected in a round-bottomed flask and evaporated at 40°C. The dry crude material was dissolved in water/acetic acid and extracted several times with diethyl ether. The aqueous phase was then collected and freeze-dried.

Quantification of the yeild of the synthesis was spectrophotometrically carried out. For this purpose the absorbance of a diluted solution of the peptide was measured at different wavelengths (i.e. λ=280 nm and λ=205 nm) in order to evaluate the extinction coefficient ε(1 mg/ml) at 205 nm according to the equation ε=27.0/1-3.85(A280/A205), where A280 and A205 are the absorbances respectively measured at 280 nm and 205 nm (74). The calculated extinction coefficient was then used to calculate the total actual yeild which resulted 83 mg.

Purification and characterization

The crude peptide was analyzed and purified by RP-HPLC on a Waters HPLC system equipped with a Waters 484 UV detector. For analytical runs a Delta Pak C-18 300A (3.9mm×15cm) column was used and the product eluted from buffer A (0.05% TFA in water) to buffer B (90% acetonitrile in buffer A) in 30 minutes with a flow rate of 1 ml/min. The HPLC spectrophotometer was set at 214 nm. The same gradient was used for the purification of the crude material on

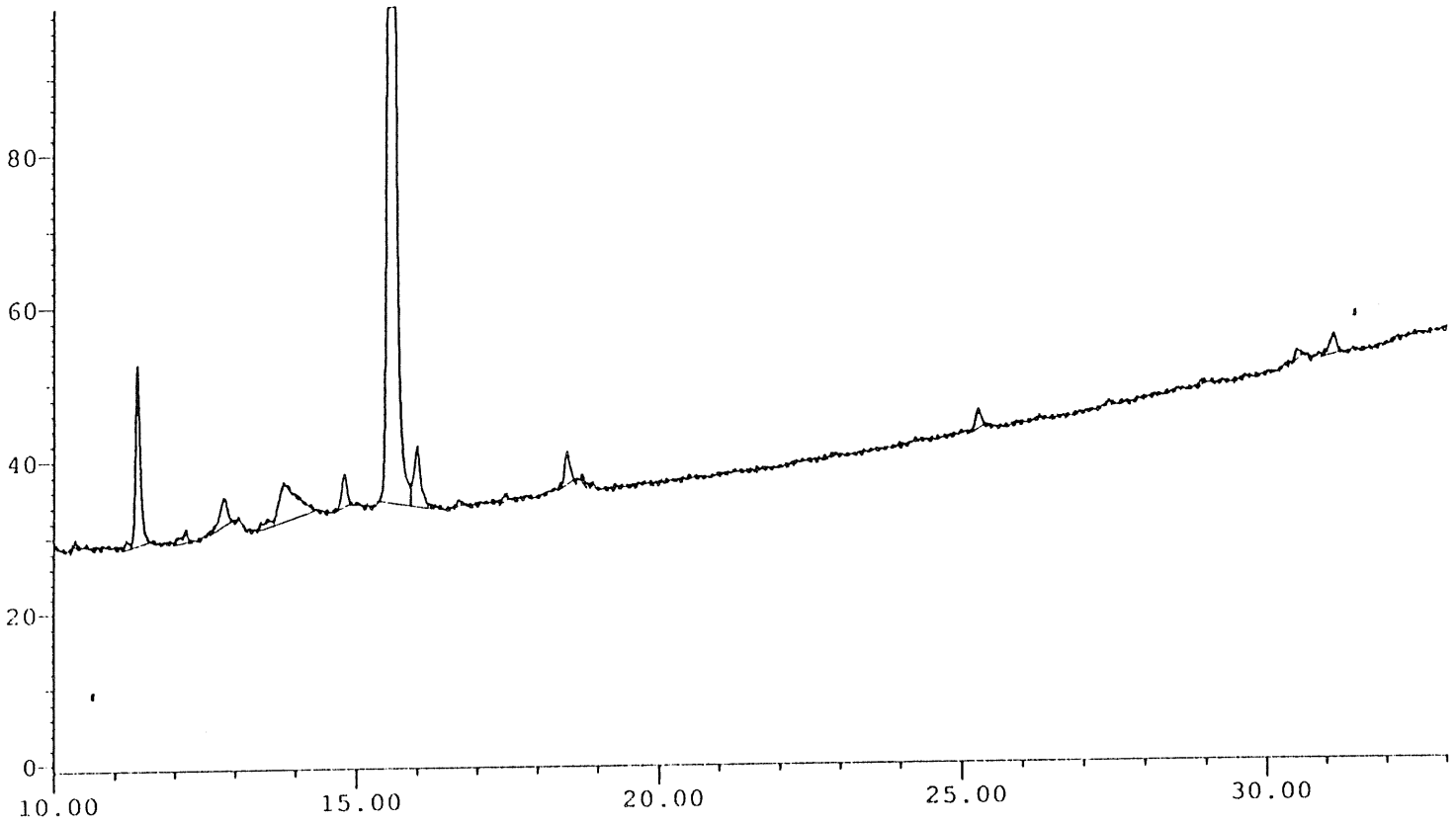


Figure 26. Analytical HPLC traces of the HTH peptide.
Column: Delta Pak RP-C 18 300 Å (3.9mm x 15cm). Buffers: 0.05% TFA in water=A; 90% acetonitrile in buffer A=B. Gradient: elution from A to 50% B in 30 minutes. Flow rate: 1ml/min. Detector: 214 nm.

a semipreparative column, Delta Pak C-18 300A (19mm×30cm), with a flow rate of 10 ml/min. HPLC traces of the purified HTH peptide are shown in Figure 26.

Amino acid analysis on the HTH peptide was performed using a custom made reactor, using the Waters PicoTag system according to the protocol reported in the Waters instruction manual.

4.2.2 Synthesis and characterization of the HTHL peptide

Synthesis protocol

The HTHL peptide (Figure 18 A) was synthesized on a solid phase support using the Fmoc strategy (73) on a Milligen 9050 Pepsynthesizer by automated continuos flow. The synthesis was carried out on the acid labile Fmoc-Leu-Pepsyn-KA resin (strong acid labile Keiselguhr resin from Milligen) using 0.94 g of resin (0.08 mmole/g) and a 6 fold excess of each Fmoc-AA-OH protected on the side chains so as to avoid secondary reactions. A mixture of active ester (OPfp or ODhbt) and in situ activated (HOBt/TBTU) chemistry was used as shown in Table 3.

Cleavage and deprotection

After the synthesis the resin was washed with isoamyl alcohol, acetic acid, isoamyl alcohol and diethyl ether and then lyophilized. The peptide-resin was treated with a mixture containing TFA/EDT/Phenol (92.5:5:2.5%) and the cleavage and deprotection reaction allowed to proceed for 3 hours at room temperature. The cleavage mixture was separated from the resin by filtration on a syntered funnel and the resin further washed with fresh cleavage solution and consecutively with small aliquots of TFA. The TFA was then removed by evaporation, the precipitate was dissolved in a mixture containing water/acetic acid (1:1) and finally extracted several times with equal volumes of diethyl ether.

TABLE 3. Amino acid sequence and side-chain protecting groups used in the synthesis of the HTHL peptide, starting from the C-terminus. Double coupling cycles are indicated by asteriscs.

Position	Aminoacid	Derivative	Quantity (grams)
1	Leu -	Fmoc-Leu	-
2	Phe	Fmoc-L-Phe-OPfp	0.250
3	Arg	Fmoc-L-Arg(PMC)-OH	0.250
4	Pro	Fmoc-L-Pro-OPfp	0.228
5	Arg	Fmoc-L-Arg(PMC)-OH	0.250
6	Lys	Fmoc-L-Lys(Boc)-OPfp	0.286
7	Thr	Fmoc-L-Thr(But)-ODhbt	0.245
8	Lys	Fmoc-L-Lys(Boc)-OPfp	0.286
9	Gly	Fmoc-Gly-OPfp	0.200
10	Asn	Fmoc-L-Asn(Tmob)-OH	0.280
11	Glu	Fmoc-L-Glu(OBut)-OPfp	0.178
12	Leu	Fmoc-L-Leu-OPfp	0.156
13	Gln	Fmoc-L-Gln(Trt)-OH	0.184
14	Glu	Fmoc-L-Glu(OBut)-OPfp	0.178
15	Ile	Fmoc-L-Ile-OPfp	0.156
16	Ser	Fmoc-L-Ser(But)-ODhbt	0.159
17	Gln	Fmoc-L-Gln(Trt)-OH	0.184
18	Gln	Fmoc-L-Gln(Trt)-OH	0.184
19	Thr	Fmoc-L-Thr(But)-ODhbt	0.163
20	Thr	Fmoc-L-Thr(But)-ODhbt	0.163
21	Gly	Fmoc-Gly-OPfp	0.139
22	Val	Fmoc-L-Val-OPfp	0.152
23	Lys	Fmoc-L-Lys(Boc)-OPfp	0.191
24	Gln	Fmoc-L-Gln(Trt)-OH	0.184
25	Ala	Fmoc-L-Ala-OPfp	0.144
26	Leu	Fmoc-L-Leu-OPfp	0.156
27	Glu	Fmoc-L-Glu(OBut)-OPfp	0.178
28	Ala	Fmoc-L-Ala-OPfp	0.144
29	Gln	Fmoc-L-Gln(Trt)-OH	0.184
30	Asn	Fmoc-L-Asn(Trt)-OH	0.180

After extraction the aqueous phase was lyophilized, the dry material dissolved in a minimum volume of water and spectrophotometrically quantified. For this purpose the same equation as reported in section 4.2.1 was used (74). The total actual yeild was thus evaluated to be 68mg.

Purification of the HTHL peptide

Analysis of the crude HTHL peptide was carried out on a reverse phase analytical column, Delta Pak C-18 300A (3.9mm×15cm), and eluted by running a gradient from 100% buffer A (0.05% TFA) to 100% buffer B (90% acetonitrile in buffer A) in 45 minutes with a flow rate of 1 ml/min and detected on the HPLC spectrophotometer at 214 nm. The same gradient as above was utilized for the purification on a semipreparative Delta Pak C-18 300A (19mm×30cm) RP-HPLC column with a flow rate of 10ml/min. The fraction corresponding to the main peack shown in Figure 27 (RT=20.5) was collected and used to carry out amino acid analysis on a Waters PicoTag workstation as in the case of the HTH peptide. HTHL was also characterized by mass spectrometry. Its exact molecular weight was determined (Figure 18 B) according to the standard methodology described in details in the case of the HTHLDIM peptide (see section 4.2.3) consisting of the tryptic digestion of HTHL and spectrometric investigation on the various fragments. It should be noted that the apparent weight of HTHLDIM (estimated using SDS PAGE) is twice the actual molecular weight. An overestimation of molecular weight in SDS PAGE appears to be typical of peptides based on a lysine core (Tossi A., personal comunication).

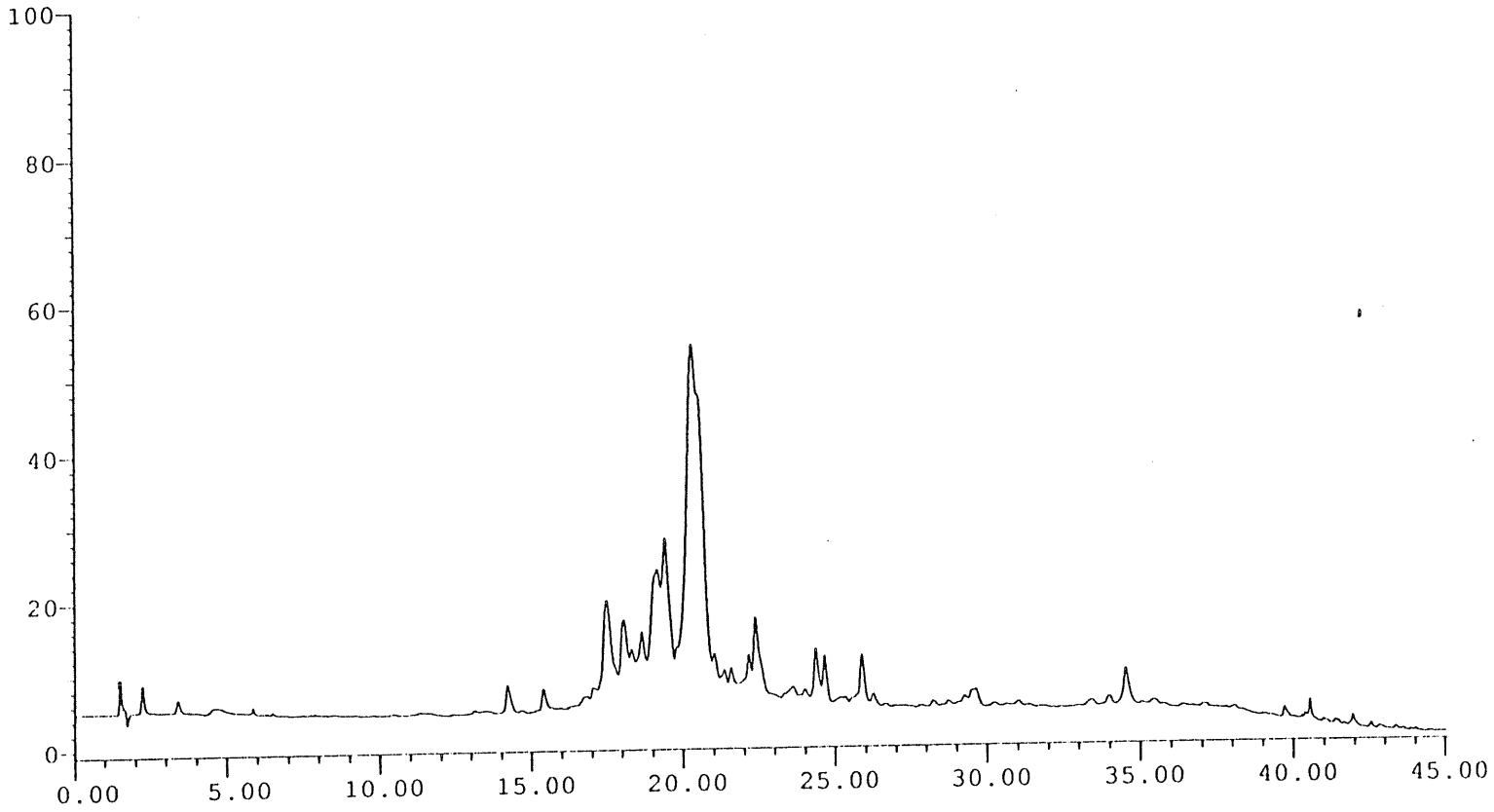


Figure 27. Analytical HPLC traces of the crude HTHL peptide.
Column: Delta Pak RP-C 18 300 Å (3.9mm x 15cm). Buffers: 0.05% TFA in water=A; 90% acetonitrile in buffer A=B. Gradient: elution from A to 50% B in 45 minutes. Flow rate: 1ml/min. Detector: 214 nm.

4.2.2. Synthesis and characterization of the HTHLDIM peptide

Synthesis protocol

The HTHLDIM peptide (Figure 18 A) was synthesized on a Milligen 9050 Pepsynthesizer by using the Fmoc chemistry (73). As the HTHLDIM peptide is relatively long (30 residues per starnd), the synthesis was carried out on a PEG-PS resin which permits greater accessinility to incoming aminoacids and reagents, and has better solvation properties than kieselguhr-based resins. 0.400 g of Fmoc-Gly-PEG-PS resin (0.190 meq/g) were used for the synthesis, which was carried out on a 0.152 mMoles scale. A six-fold aminoacid excess with equimolar quantities of HOBT and TBTU was utilized according to three differentiated sets of automated cycles. During the first, Fmoc-Lys(Fmoc)-OH was anchored to the solid support so as to provide the core for the branching of the HTHLDIM peptide. The second set consisted in the coupling of the first two Pro residues. For this purpose the deprotection step, concerning the removal of the two Fmoc groups protecting the α and ε amino groups of the Lys core, was very fast so as to avoid the formation of dichetopiperazine as non-desired product (given the presence of Gly as the first residue). The third set consisted in the elongation of the branched peptide. The synthesis was interrupted at the level of the helix H3 and a small amount of resin removed, so as to obtain the truncated peptide HLDIM.

Cleavage and deprotection

After the synthesis the PEG-PS resin was washed several times with methanol and then diethyl ether and vacuum dried. The resin was then treated with the cleavage mixture TFA/Phenol/ETD/TA/Tris-isopropylsilane/water (90:2:2:2:2:2 %), and the reaction allowed to proceed for about 4 hours at room temperature. The resin was then filtered on a syntered funnel and further washed with cleavage mixture and few times with TFA. The filtrate was collected in a

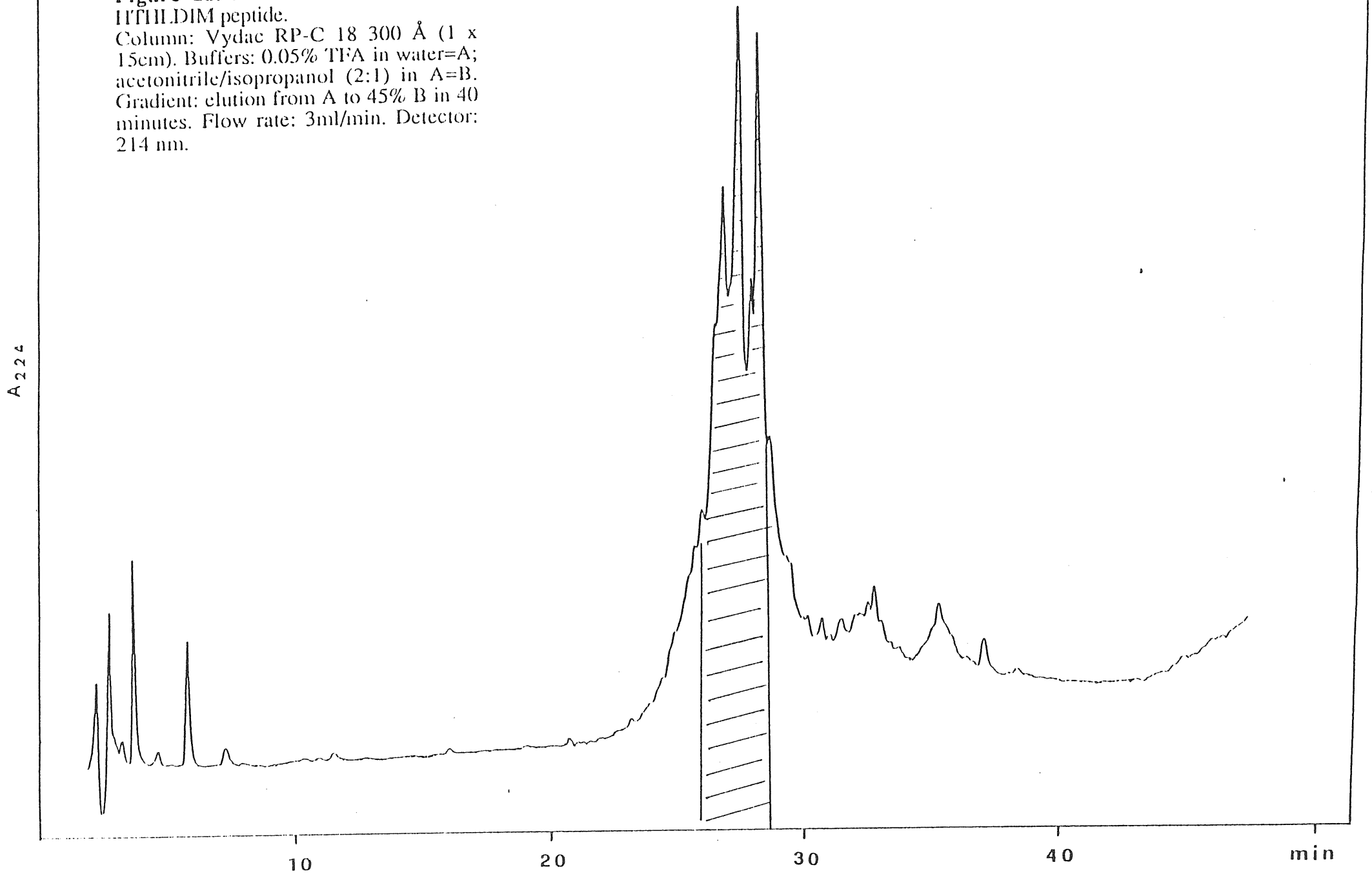
round-bottomed-flask and evaporated at 40°C. The dry crude material was then dissolved in a mixture of water/acetic acid (1:1) and extracted several times with diethyl ether. The aqueous phase was collected and lyophilized. The yeild of the synthesis was then spectrophotometrically assayed as in section 4.2.1.

Purification and characterization of the HTHLDIM peptide

The HPLC purification of HTHLDIM was carried out on a Varian 9010 chromatographer equipped with a Varian 9010 detector. The sample (8 mg) was dissolved in 0.1% TFA, filtered on Millex-HV filters (Millipore) and injected in a Vydac 5RP18 (1.0×25cm), 300 A pore size reverse-phase column. The column was initially eluted at room temperature with a mixture of 10% B (acetonitrile/isopropanol+0.05% TFA) in 90% A (0.05% TFA) for 5 minutes. The sample was then eluted with a linear gradient until 45% B in 40 minutes with a flow rate of 3 ml/min. The detector was set at 224 nm. The three main peaks were collected together, as shown in Figure 28, and lyophilized so as to use them for the tryptic digestion. Similar conditions were used to purify HLDIM.

Dry sample (4.7 mg) was then dissolved in 0.02 M ammonium acetate containing 0.001 M calcium chloride, pH=8.3. Trypsin was added with an enzyme/substrate ratio of 1:50 (mol/mol) and the sample incubated for 4 hours at 37°C. The enzymatic reaction was stopped by freezing the sample which was then lyophilized and redissolved in 0.1% TFA, filtered and injected. The HPLC separation of the tryptic fragments was carried out on the same Vydac 5RP18 (1.0 ×25cm) column used for the sample purification. The column was firstly eluted at room temperature with a mixture of 5% B (acetonitrile/5% water/0.08% TFA) in 95% A (0.1% TFA) for 5 minutes and then with a linear gradient until 47% B in 55 minutes. The flow rate was 3 ml/min and the wave length was set at 224 nm. The chromatographic traces are shown in Figure 29. All five peaks collected were freeze-dried so as to be analyzed by FAB-MS.

Figure 28. HPLC traces of the crude HTHLDIM peptide.
Column: Vydac RP-C 18 300 Å (1 x 15cm). Buffers: 0.05% TFA in water=A; acetonitrile/isopropanol (2:1) in A=B. Gradient: elution from A to 45% B in 40 minutes. Flow rate: 3ml/min. Detector: 214 nm.



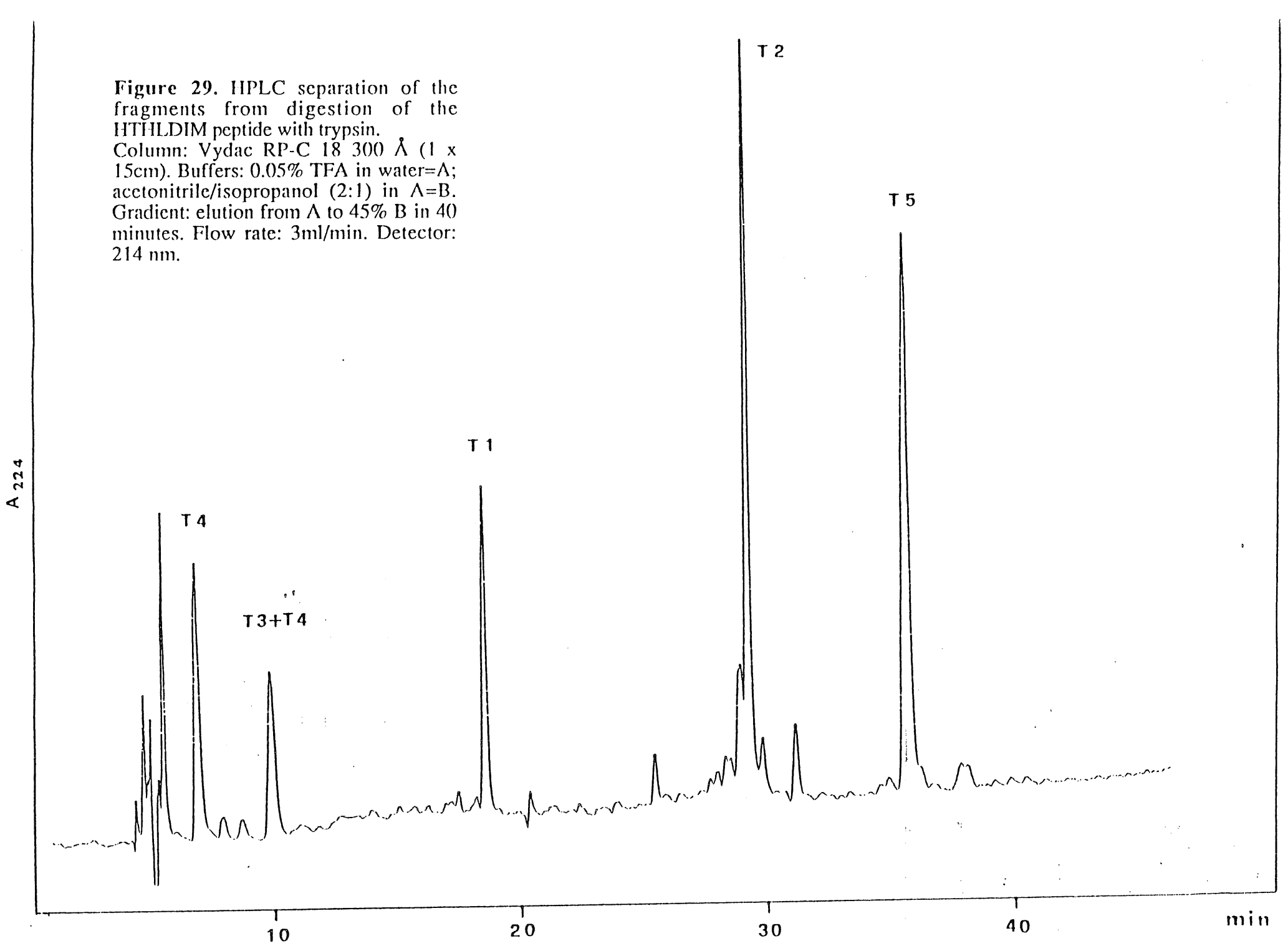


Figure 29. HPLC separation of the fragments from digestion of the HTHLDIM peptide with trypsin.
Column: Vydac RP-C 18 300 Å (1 x 15cm). Buffers: 0.05% TFA in water=A; acetonitrile/isopropanol (2:1) in A=B. Gradient: elution from A to 45% B in 40 minutes. Flow rate: 3ml/min. Detector: 214 nm.

Table 4 . Peptides originated by digestion with trypsine

		MW*	RT**
T1	Asn-Gln-Ala-Glu-Leu-Ala-Glu-Lys	901	19.05
T2	Val-Gly-Thr-Thr-Gln-Gln-Ser-Ile-Glu-Gln-Leu-Glu-Asn-Gly-Lys	1630	29.51
T3	Thr-Lys	247	
T4	Arg-Pro-Arg	427	6.92
T3+T4	Thr-Lys-Arg-Pro-Arg	656	9.53
T5	Phe-Leu-Pro-Lys-Pro-Leu-Phe | Gly	917	35.88

* Molecular Weight

**RetentionTime

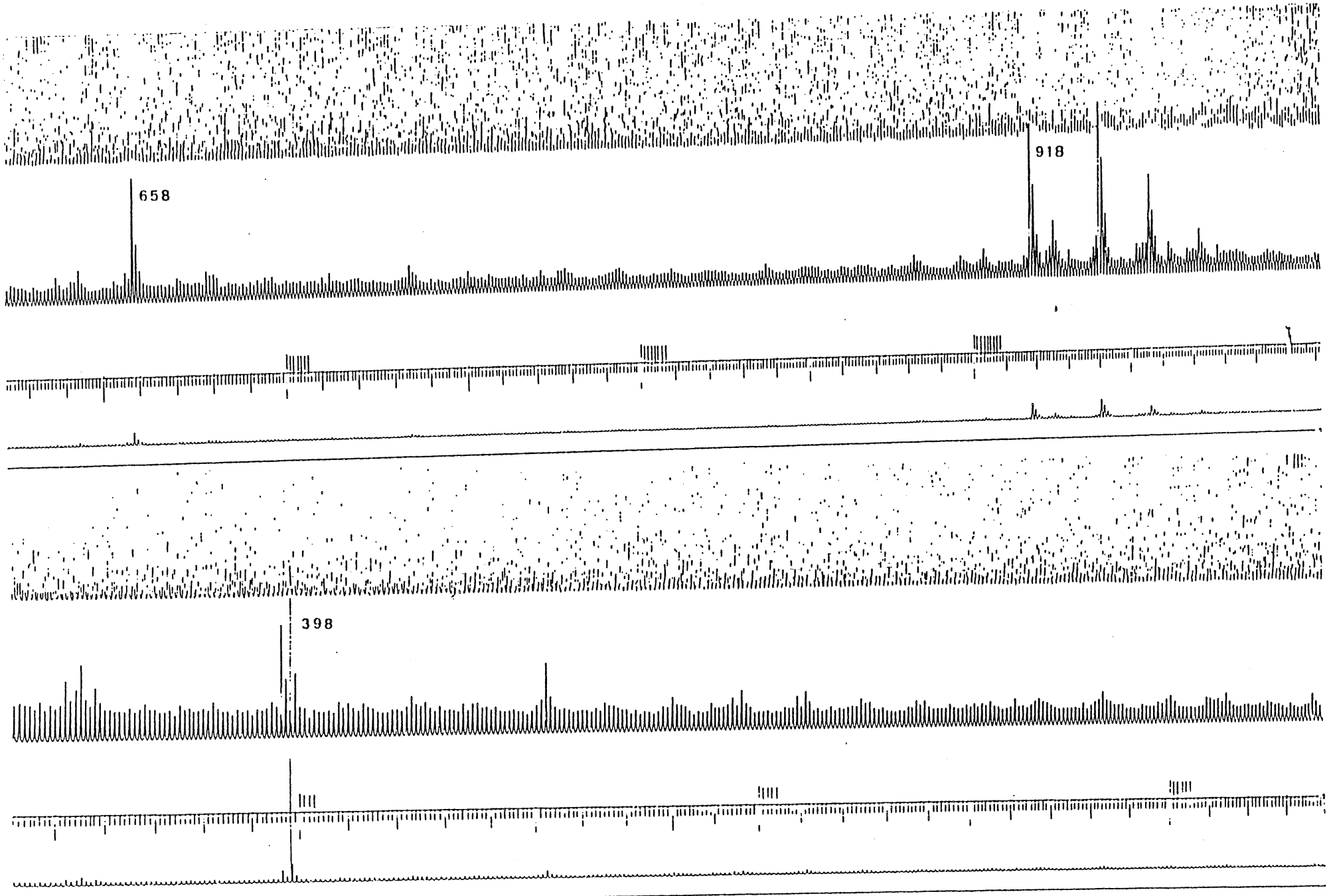


Figure 30. FAB-MS spectrum of the T5 fragment coming from the digestion of the HTHLDIM peptide with trypsin (see Table 4).

The mass spectra were obtained by a VG ZAB-2SE mass spectrometer equipped with a FAB source emitting cesium ions. Lyophilized samples were placed on top of the probe for evaporation and then about 2 µl of glycerol were added so as to solubilize them. Samples were then bombed with a cesium ion beam of 35 KeV. The spectra show that all of the five peaks collected during the HPLC runs correspond to the theoretically predicted tryptic fragments (Table 4). As an example, Figure 30 shows the FAB-MS spectrum of tryptic fragment T5 (MH+=918), one of the peptides derived from the digestion with trypsin and reported in Table 4. The fragment at 658 corresponds to peptide T5-(Leu-Phe) whereas the fragment at 398 corresponds to a further loss of (Leu-Phe) from the previous peptide.

4.3 Isolation and characterization of the 434 repressor protein

Isolation of the crude extract containing the434 repressor protein

E.coli DH5 bacteria infected with lysogen imm434 (kindly donated by Dr R. Weinberg) were grown overnight in LB plates at 37°C. A small colony was then picked up and further grown overnight in 250 ml of LB medium at 37°C and then centrifuged at 5000 RPM for 30 minutes at 4°C. The pellet was resuspended in a buffer containing 30 mM TRIS.HCl pH=7.4, 0.15 M sodium chloride, 5 mM magnesium chloride, 1 mM PMSF (protease inhibitor) and then sonicated for 5-10 minutes. The suspension was then centrifuged for 30 minutes at 15000 RPM at 4°C and the pellet discarded. The supernatant was then incubated at 4°C with 25% ammonium sulphate for 40 minutes and then centrifuged at 10000 RPM for 30 minutes at 4°C. The pellet was dissolved in TM buffer (i.e. 2 M TRIS.HCl pH=7.4, 1 M zinc chloride, 2 M magnesium chloride, 0.5 M EDTA, glycerol) and tested for the presence of the 434 repressor protein by band shift and footprinting experiments. The supernatant was then treated with 45% ammonium sulphate for

40 minutes at 4°C and then centrifuged at 10000RPM for 30 minutes again at 4° C. The supernatant was discarded whereas the pellet was also dissolved in TM buffer and its biological activity assayed again by means of band shift and footprinting reactions.

Characterization of the crude extract

Characterization of both crude extracts was carried out by SDS polyacrylamide gel electrophoresis. The presence of λCI 434 repressor is assayed by SDS polyacrylamide gel electrophoresis run together with a molecular weight standard and stained with Coomassie (data not shown). Final experimental proof of the presence of λCI 434 repressor is provided by band shift assays and footprinting reactions discussed in details in the following paragraphs.

4.4 Binding assays

4.4.1 Band shift reactions

Band shift assays were carried on all the available DNA fragments, double-stranded oligonucleotides bearing the 16-3 OR site and enzymatically concatenated oligonucleotides. The experiments were performed using the synthetic peptides HTH, HTHL, HTHLDIM as well as the λCI 434 repressor crude extract (Figures 20, 21 and 22) The latter was utilized to run control experiments in the presence of all operator models so as to have clear-cut evidence that that binding to the operator sites in each system is possible, and by comparison determine which peptides have the highest affinity for the DNA (see chapter 3.0 for results and chapter 5.0 for discussions).

The band shift reactions were performed with a 5′ end-labelled DNA probe (20000-30000 CPM in each slot) incubated with the peptides or the protein

extract in a buffer containing 20 mM HEPES (N-[2-Hydroxyethyl]piperazine-N′-[2-ethanesulfonic acid]) pH=7.3, 50 mM sodium chloride, 5 mM magnesium chloride, 2 mM DTT, 0.2 mM EDTA, 5% glycerol. Incubation was carried out in ice for 30 minutes in the case of the λCI 434 repressor protein extract whereas it was prolonged to 1 hour in the case of all the assayed peptides. Since highly specific binding and affinity was tested, poly[d(I-C)] (1 μg/μl) was used in order to remove any aspecific binding to the DNA probes. At the end of the incubation period the samples were then loaded on an 8% non-denaturing polyacrylamide gel and electrophoresed at very low voltage (5-10 V/cm) at 4°C.

4.4.2 UV crosslinking

Preparation of the DNA binding sites for UV crosslinking

1)32bp oligonucleotide containing the 16-3 operator site (OR).

The 32-mer oligonucleotide 5′-CTTAGTGACTCGTACAACTACAATTGTTGAAG-3′ was chemically synthesized on an Applied Biosystem 380A DNA Synthesizer and designed so as to comprise the 16-3 palindromic operator (OR). It was purified by precipitating it with n-butanol and then it was electrophoresed on a denaturing polyacrylamide gel and the band corresponding to its size excised. After extracting it from polyacrylamide it was phenol/chloroform extracted and further precipitated with 100% ethanol. The oligonucleotide was then used as template for E.coli polymerase I Klenow fragment in order to enzymatically prepare the double-stranded oligonucleotide to be used in the binding reaction. In a total reaction volume of 25 μl containing 2 μl of sample oligo (3e-3M), 2.5 μl of 10 times concentrated reaction buffer (130 mM potassium phosphate, pH=7.4, 6.5 mM magnesium chloride, 1 mM DTE, 32 μg/ml BSA), 1.25 μl of dATP (100 mM), 1.25 μl of dGTP (100 mM), 2 μl of α-32P-dCTP (25 μCi/μl), 4 μl of dUBr (i.e.

25 mM deoxybromourydine), 2.5 µl of water, 1 µl of Klenow Fragment (2 U/µl) of DNA Polymerase I was added. The polymerization reaction was carried out for 1 hour at 37°C. The double-stranded oligonucleotide, named 16-3OR(32 mer) was then purified on a G25 Nick column by eluting the sample with TE buffer (i.e. 10 mM TRIS.HCl pH=7.5, 1 mM EDTA) (65).

2) 45bp oligonucleotide with the 16-3 operator region (OR and OR′)

Partially complementary oligonucleotides K1, 5′-CAAACAATGTACTTGACGACTCATCAC, and K2, 5′-TCTACAACTACAATTGTTGTGATGAGTCG, were synthesized on an Applied Biosystem 380A DNA Synthesizer. They were precipitated with n-butanol and then electrophoresed on a denaturing polyacrylamide gel and the bands excised. After extraction from polyacrylamide the samples were phenol/chloroform extracted and then further prcipitated with 100% ethanol and UV quantitated ([K1]=15.7 µg/µl, [K2]=16.2 µg/µl). Equimolar quantities of K1 and K2 were then annealed in water by heating them at 95°C and then slowly cooling down at room temperature. The annealed oligonucleotides were then used as templates for Klenow fragment. For this purpose in a total volume of 50 µl containing 20 µl of sample DNA, 5 µl of 10 times concentrated Klenow buffer (i.e. 130 mM potassium phosphate, pH=7.4, 6.5 mM magnesium chloride, 1 mM DTE, 32 µ g/ml BSA), 1.25 µl of dATP (100 mM), 1.25 µl of dGTP (100 mM), 2 µl of α-32P-dCTP (25 µCi/µl), 4 µl of dUBr (i.e. 25 mM deoxybromourydine), 11 µl of water, 1 µl of E.coli DNA Polymerase I Klenow fragment (2 U/µl) was added. The polymerization reaction was carried out for 1 hour at 37°C. The double-stranded oligonucleotide, named 16-3[OR-OR′](45 mer), was purified through a G25 Nick column by eluting it with TE buffer (i.e. 10 mM TRIS.HCl, pH=7.5, 1 mM EDTA) (65).

Binding reactions

Binding reactions were carried out as already discussed in paragraph 4.4.1. In this case the main difference relies on the fact that when deoxybromouridyne is incorporated into the oligonucleotides, it becomes light-sensitive. It is therefore necessary to work in a dark environment.

Uv crosslinking

The crosslinking reaction (75, 76) was carried on all peptides i.e. HTH, HTHL (data not shown) and HTHLDIM. Samples were loaded on a strip of parafilm placed on ice and then irradiated with a UV lamp (λ=254nm) placed at 5 cm from the samples. UV irradiation was carried out for about 1 hour. Samples were then loaded on a 12% reducing SDS polyacrylamide gel and then electrophoresed at 20mA and 160V. The gel was then stained with Coomassie and then exposed (Figure 23).

4.4.3 DNAse Footprinting

DNAse footprinting reactions were carried out on the EcoRI-SalI pTZ19U fragment with the 16-3 operator region in the presence of both the 434 repressor-containing crude extract and the HTHLDIM peptide. The digestion reactions were also carried out on the naked DNA fragments in order to have positive controls. Both the HTHLDIM peptide (1 mg/ml) and the 434 crude extract were incubated with the EcoRI-SalI pTZ19U fragment (ca.20000 CPM), 16-3(298 bp) for 60 minutes in ice water. The reaction was assembled as follows:

	Crude extract	HTHLDIM
Water	12 µl	15 µl
3× Solution D	8 µl	8 µl
0.1 M ATP, pH=7.0	0.25 µl	0.25 µl
10 mg/ml yeast RNA	0.4 µl	4 µl
Protein(s)	5 µl	2 µl
Labelled DNA	2 µl	2 µl

3× Solution D contains 60 mM HEPES, pH=7.9, 150 mM potassium chloride, 0.3 mM EDTA, 1.5 mM DTT and 30% glycerol. 2µl of Mg/Ca mix (125 mM magnesium chloride and 25 mM calcium chloride) were added and the samples incubated for 1 minute at 20°C in a water bath. Then 2µl of DNAse I (Boehringer Mannheim) differently diluted from a stock solution (3 mg/ml diluted 1:100, 1:500, 1:1000, 1:10000, 1:20000 with 1× Solution D containing 100 µg/ml BSA) were added to the reaction volumes and then incubated again for 1 minute at 20°C. The digestion reaction was stopped by adding 200µl of a stop solution (20 mM Tris, pH=7.5, 0.1 M sodium chloride, 1% SDS, 5mM EDTA, 50 µg/ml proteinase K, 25 µg/ml yeast RNA) and incubating for 30 minutes at 45°C. The samples were then precipitated with 750 µl of 100% ethanol, washed with 70% ethanol and the pellets lyophilized. They were then dissolved in formamide containing dye solution, heaten up for 3 minutes at 95°C, immediately cooled in ice and loaded on a 6% denaturing polyacrylamide gel and electrophoresed at high voltage (1800-1900 V).

4.5 Circular Dichroism

CD spectra were measured on a JASCO 600 dichrograph of peptides HTH, HTHL, HTHLDIM (Figure 19) in a thermostated environment at room temperature, using a 1 cm path-length cell. The samples were dissolved in 10 mM phosphate buffer pH=7.00 and spectra measured in presence of increasing quantities of the α-helical inducing agent trifluoroethanol (TFE, 0%, 10%, 15%, 20%, 30%). Concentration of peptides were [HTH]=1.4e-4M, [HTHL]=3.8e-5M, [HTHLDIM]=1.5e-5M. All spectra were averaged three times and smoothed with the software provided by JASCO Inc. (Easton, Maryland).

Preliminary CD studies were also performed on peptides in the presence of their cognate 16-3(298 bp) DNA and difference spectra were obtained by subtracting the peptide spectrum from the corresponding peptide-DNA spectrum and data were collected and smoothed as described above.

4.6 Computer methods

The x-ray structure coordinates for CI 434 N-terminal domain (53) and for the λ repressor (51) bound to DNA were abstracted from the Brookhaven Database. They were loaded into the Biosym Technologies Insight II program and standard procedures from the Insight module used for the visualization process.

All energy minimization and molecular dynamics experiments were carried out using the Discover module of the same program, using the standard protocol. Detailed instructions are available in the Biosym Insight II/ Discover instruction manual. The dielectric was set to 1.0, temperature was 300 K and 10000 iterations were carried out after an initial 100 cycles of energy minimization.

Database searching was carried out using the Findseq module in the Intelligenetics program, on the Swissprot Database. Alignments of repressor

sequences were performed using the Genealign module of the same program, on edited sequences containing only the HTH motif with short flanking regions (Fig. 15).

The distance matrix was calculated by running one cycle of energy minimization on the CI 434 N-terminal domain structure and abstracting all the Cα-Cα distances from the resulting molecular definition files.

5.0 DISCUSSION AND CONCLUSIONS

Sequence specific DNA binding proteins use several different types of structural strategies to bind to their recognition sites, the main classes being illustrated in Table 1. Attempts have been made to design peptides that contain the DNA-binding motifs of at least three of these, namely the Leucine zipper (77, 78, 79), the Zinc finger (80) and the HTH motif (64, 81). In the first case, peptides were designed that contained only the basic motif and leucine zipper of GCN4 region, and these successfully bind to DNA. Peptides containing only basic motifs linked by a cystine bridge were less successful (77, 78). In the second case, the synthesis of peptides based on the zinc finger region of the transcription factor Sp1, which specifically recognizes GC-rich regions in the long terminal repeat (LTR) of the HIV genome, has been reported, showing remarkable biological activity (80).

Attempts at producing DNA-binding peptides based on the HTH motif have not been fully successful (64).

Harrison and Aggarwal in the review on DNA recognition by the HTH motif (62) make the rather daunting statement in their conclusion <<Thus, the unit that "recognizes" DNA is really an entire binding domain, appropriately dimerized, and nearly all the protein/DNA contacts in a complex with a correct target site contribute to specificity>>. This leaves very little room for designing peptides based on a minimalist approach, and to be perfectly honest, the project described in this work was initiated with the enthusiasm of inexperience. However, it was felt that a rigorous modelling approach, using all the facilities placed at our disposal by molecular modelling programs, might produce peptides with some hope of functioning as sequence specific DNA binders, albeit probably not as well as the parent protein. The possibility of chemically synthesizing such peptides, rapidly, in good yields and in easily purifiable form is attractive. They could, for example, be subsequently used as DNA probes by addition of moieties

that can chemically or photochemically cleave DNA at the binding site. To date, this has been carried out with CAP, trp and met repressors and requires labourious isolation and purification procedures.

Computer modelling is a useful tool in two respects. Firstly it allows a clear visualization of a protein bound to its cognate DNA, and thus a full understanding of DNA/protein interactions both specific and general, as well as those inter-residue interactions that stabilize the protein. An added facility is that of providing exact interatomic distances, allowing the building of a distance matrices. Much information exists in the literature describing the exact interactions of the CI434 repressor with its cognate DNA, but a full understanding can be achieved only by direct visualization. However, even the powerful display provided by the computer has its limitations, and the building of a wire model based on the computer model was necessary at a certain stage, so as to give a full three dimensional appreciation of the arrangement of residues in the repressor.

The second invaluable service provided by computer modelling is the possibility of designing and testing peptides before synthesis. This allows one to evaluate the viability of the design. The first prototype peptide that was considered, before modelling facilities were available, was based on a flexible polypeptide linker attached to the binding helix at one end, and providing a dimerization point at the other by means of a cystine bridge. The design of this peptide was based merely on the distance of 34 Å between DNA half/sites and the putative size of the binding helix. This peptide was rejected when modelling was carried out, as it became obvious that the linker was far too flexible. It is by no means certain that a designed peptide, which gives correct structures on modelling, will actually function in reality. The proof can only come from synthesis and testing. It is however quite unlikely that a structure that does not function in a modelling experiment, will function in reality.

Four peptides were synthesized, two (HTH and HTHL) containing the HTH motif but no dimerization devices, the other two (HLDIM and HTHLDIM)

containing a permanent form of dimerization, by means of a covalent linkage. One contained only the binding helix, while the other contained flanking regions that modelling had indicated might provide a stable structural context for the binding helix. CD spectra of the HTH containing peptides indicated a predominantly a-helical conformation in a partially organic environment (20-30% TFE) so that the peptides appear to have the potential of folding to the correct structure for binding , given the proper conditions.

Band-shift assays indicated that only HTHLDIM, which contains both a more extended structural apparatus and an inbuilt dimerization, was successful in specifically binding to the cognate DNA. Binding does not appear to occur to any extent for the monomeric petides and HLDIM appears to aggregate with the cognate DNA. HTHLDIM binds only to DNA models containing at least two operator sites, which denotes a fundamental role of cooperativity in binding. This is supported by the fact that HTHLDIM binds to concatemers containing multiple operator sites only when these are in phase, and that binding appears to increase as the number of operator sites increases.

UV crosslinking experiments were carried out so as to overcome the problems posed by the intrinsic lability of binding. Results entirely confirmed those of conventional band-shift experiments, underlining the importance of cooperativity. Thus crosslinking occurs only for HTHLDIM, and only to a DNA fragment containing two operator sites.

The experiments described in this work have demonstrated that it is possible to design peptides based on the HTH motif, using a minimalist approach. Thus, some form of specific binding is possible with only part of the apparatus normally used by the repressor. The peptide HTHLDIM is however only a first attempt, and is far from having the affinity for its binding site required for a functioning DNA probe. The experience and knowledge gained in constructing it will be fed back into the designing of a new generation of peptides. In this respect, the following experiments are planned for the future:

1) DNA suitable for cross-linking will be designed so as to contain operators in an in-phase and out-of-phase arrangement, so as to further test the role of cooperativity.

2) DNA suitable for crosslinking will be designed so as to contain operators that are mutated, for example with the addition of one or two more base-pairs in the region between half-sites. This will test the specificity of binding, and wheater it occurs in a manner similar to that of the full repressor.

3) High resolution footprinting techniques will be applied so as to clearly define the binding site.

4) A more stable peptide will be constructed that contains also the helix H1, thus increasing the number of inter-residue contacts in the peptide (this can be clearly seen from the distance matrix in Figure 14).

5) Synthetic "mutant" peptides will be designed in which residues in key positions are changed, so as to determine their role in affinity and specificty of binding

The problems that were faced with the peptides described in this work have also underlined the weaknesses of the minimalist approach, so that another approach based on engineering the full N-terminal domain has also been initiated. The design of a fully synthetic, covalently linked, N-terminal domain dimer is under way. Modelling studies indicate that this is possible and three different strategies have been considered, namely i) synthesis of peptide fragments followed by chemical condensation; ii) synthesis of peptide fragments followed by enzymatic ligation by novel ligases (82) and iii) direct one-shot chemical synthesis of a dimerized N-terminal domain.

The work described in this thesis has been presented in the form of posters to the NATO/EMBO/FEBS course on Protein Structure, Function and Design, Spetsai, August 30-September 12, 1993 and the 38th Congress of the Italian Biochemistry Society, Trieste, September 7-10, 1993.

BIBLIOGRAPHY

(1) Freemont, P.S., Lane, A.N. and Sanderson,M.R. (1991) Biochem. J. **278**, 1-23

(2) Suck, D., Lahn, A. and Oefner, C. (1988) Nature **322**, 464-468

(3) Travers, A.A. and Klug, A. (1990) in DNA Topology and its Biological Effects (Cozzarelli, N.R. and Wang, J., eds), pp. 57-106, Cold Spring Harbor Laboratory Press

(4) Pjura, P.E., Grzeskowiak, K. and Dickerson, R.E. (1987) J. Mol. Biol. **197**, 257-271

(5) Teng, M., Usman, N., Frederick, C.A. and Wang, A. H-J. (1988) Nucleic Acid Research **16**, 2671-2690

(6) Mair E.A. Churchill, Travers, A.A. (1991) Trends Bbiochem. **16**, 92-97

(7) Warwicker J., Engelman B.P., Steitz T.A. (1987) Proteins **2**, 283- 289

(8) Schultz S.C., Shields G.C., Steitz T.A. (1990) J. Mol. Biol. **213**, 159-166

(9) Travers, A.A. (1988) in Nucleic Acids and Molecular Biology (Eckstein F., Lilley D.M.J., eds), pp. 136-148, Springer-Verlag Berlin Heidelberg

(10) Berg J.M. (1993) Curr. Op. Struct. Biol. **3**, 11-16

(11) Harrison S.C. (1991) Nature **353**, 715-719

(12) Pabo C.O., Sauer R.T (1992) Annu. Rev. Biochem. **61**, 1053-1095

(13) Marmorstein R., Carey M., Ptashne M., Harrison S.C. (1992) Nature **356**, 408-414

(14) Kraulis P.J., Raine A.R.C., Gadhavi P.L., Laue E.D. (1992) Nature **356**, 448-450

(15) Baleia J.D., Marmorstein R., Harrison S.C., Wagner G. (1992) Nature **356**, 450-453

(16) Desjarlais J.R., Berg J.M. (1992) Proteins **12**, 101-104

(17) Desjarlais J.R., Berg J.M. (1992) Proc. Natl. Acad. Sci. USA **89**, 7345-7349

(18) Pavletich N.P., Pabo C.O. (1991) Science **252**, 809-817

(19) Thukral S.K., Morrison M.L., Young E.T. (1992) Mol. Cell. Biol. **12**, 2784-2792

(20) Bess J.W. Jr., Powell P., Issaq H.J., Shumak L.J., Grimes M.K., Henderson (L.E.,Arthur L.O. (1992) J. Virol. **66**, 840-847

(21) Niklov D.B., Hu S.H., Lin J.P., Gasch A., Hoffmann A., Horikoshi M., Chua N.H., Roeder R.G., Burley S.K. (1992) Nature **360**, 40-46

(22) Summers M.F., Henderson L.E., Chance M.R., Bess J.W. Jr., South T.L., Blake P.R., Sagi I., Peret Alvarado G., Sowder R.C. III, Hare D.R., Arthur L.O. (1992) Protein Sci. **1**, 563-574

(23) Starr D.B., Hawley D.K. (1991) Cell **67**, 1231-1240

(24) Ellenberger T.E., Brandl C.J., Struhl K., Harrison S.C. (1992) Cell **71**, 1223-1237

(25) Wolberger C. (1993) Curr. Op. Struct. Biol. **3**, 3-10

(26) Wolberger C., Dong Y., Ptashne M., Harrison S.C. (1988) Nature **335**, 789-795

(27) Saudek V., Pastore A., Morelli M.A., Frank R., Gausepohl H., Gibson T. (1991) Protein Eng. **4**, 519-529

(28) O'Shea E.K., Rutkowski R., Kim P.S. (1989) Science **243**, 538

(29) Landschulz W.H., Shortle D., Mc Knight S.L., unpublished observations; Landshulz W.H., Johnson P.F., McKnight S.L. (1989) Science **243**, 1681-1688

(30) Agre P., Johnson P.F., McKnight S.L.(1989) Science **246**, 922

(31) O' Neal K.T., Hoess R.H., DeGrado W.E. (1990) Science **249**, 774-778

(32) Vinson C.R., Sigler P.B., McKnight S.L. (1989) Science **246**, 911-916

(33) Busch S.J., Sassone-Corsi P. (1990) Trends Gen. **6**, 36-40

(34) Sassone-Corsi P., Ransone L. J., Lamph W.W., Verma I.M (1988) Nature **336**, 692-695

(35) Gentz R., Rauscher F.J. III, Abate C., Curran T. (1989) Science **243**, 1695-1699

(36) Hedge R.S., Grossman S.R., Laimins L.A., Sigler P.B. (1992) Nature **359**, 505-512

(37) Otting G., Qian Y.Q., Billeter M., Mueller M., Affolter M., Gerhing W.J., Wutrich K. (1990) EMBO J. **9**, 3085-3092

(38) Wolberger C., Vershon A.K., Lju B., Johnson A.D., Pabo C.O. (1991) Cell **67**, 517-528

(39) Kissinger C.R., Lju B., Martin-Blance E., Kornberg T.B., Pabo C.O. (1990) Cell **63**, 579-590

(40) Qian Y.Q., Billeter M., Otting G., Muller, Gehring W.J., Wuthrich (1989) Cell **59**, 573-580

(41) Lee M.S., Gippert G.P., Soman K.V., Case D.A., Wright P.E. (1989) Science **245**, 635-637

(42) Klevit R.E., Herriott J.R., Horvath S.J. (1990) Proteins **7**, 215-226

(43) Hard T., Kellenbach E., Boelens R., Maler B.A., Dahlman K., Freedman L.P., Carlstedt Duke J., Yamamoto K.R., Gustafson L. A., Kaptein R. (1990) Science **249**, 157-160

(44) Swaba J.W.R., Neuhaus D., Rhodes D. (1990) Nature **328**, 458-461

(45) Branden C., Tooze J. (1991) Introduction to Protein Structure, Garland Publishing, Inc., New York and London

(46) Ptashne M. (1992) A Genetic Switch, Cell Press & Blackwell Scientific Publications

(47) Johnson A.D., Meyer B.J., Ptashne M. (1979) Proc. Natl. Acad. Sci. **76**, 5061

(48) Johnson A.D., Pabo C.O., Sauer R.T. (1980) Methods Enzymol. **65**, 839

(49) Brennan R.G., Matthews B. (1989) Trends Biochem. **14**, 286-290

(50) Brennan R.G., Roderick S.L., Takeda Y., Matthews B.W. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 8165-8169

(51) Jordan S.R, Pabo C.O. (1988) Science 242, 893-899

(52) Hochschild A., Ptashne M., (1986) Cell **44**, 925-933

(53) Aggarwal A.K., Rodger D.W., Drottar M., Ptashne M., Harrison S.C. (1988) Science **242**, 899-907

(54) Dallmann G., Papp P., Orosz L. (1987) Nature **330**, 398-401

(55) Otwinowski Z., Schevitz R.W., Zhang R.G., Lawson C.L., Joachimiak A., Marmorstein R.Q., Luisi B.F., Sigler P.B. (1988) Nature **335**, 321-329

(56) Neri D., Billeter M., Wuthrich K. (1992) J. Mol. Biol. **233**, 743-767

(57) Mondragon A., Subbiah T., Almo S.C., Drottar M., Harrison S.C. (1989) J. Mol. Biol. **205**, 189-200

(58) Neri D., Billeter M., Wider G., Wuthrich K. (1992) Science **257**, 1559-1563

(59) Otwinowski Z., Shevitz R.W., Zang R.G., Lawson C.L., Joachimiak A., Marmorstein R.Q., Luisi B.F., Sigler B.P. (1988) Nature **335**, 321-329

(60) Arrowsmith C., Patcher R., Altman R., Iyer S.B.,Jardetzky O. (1990) Biochemistry. **29**, 6332-6341

(61) Arrowsmith C., Carey J.,Treat-Clemons L., Jardetzky O.(1989) Biochemistry **28**, 3875-3879

(62) Harrison S.C., Aggarwal A. (1990) Annu. Rev. Biochem. **59**, 933-969

(63) Dodd I.B., Egan B.J. (1990) Nucleic Acid Research **18**, 5019-5026

(64) Grokhovsky S. L., Surovaya A.N., Brussov R.V., Chernov B.K., Sidorova N.Yu., Gursky G.V. (1991) J. Biomol. Strcture Dynamics **8**, 989-1025

(65) Sambrook J., Fritsch E.F., Maniatis T (1989) Molecular Cloning; A Laboratory Manual Cold Spring Harbor Laboratory Press

(66) Del Sal G., Manfioletti G., Schneider C., (1989) Biotechniques 7(5) 514-519

(67) Birnboin H.C., Doly J. (1979) Nucl. Acids Res. **7**, 1513-1523

(68) Maxam A.M., Gilbert W. (1980) Methods Enzymol. **65**, 499

(69) Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A., Arnheim N. (1985) Science **230**, 1350-1354

(70) Mullis K.B., Faloona F. (1987) Methods in Enzimology **155**, 335-350

(71) Saiki R.K., Gelfand D.H., Sttoffel S., Sharf S., Higuchi R., Horn G.T., Mullis K.B., Erlich H.A. (1988) Science **239**, 487-491

(72) Brukner I., Dlakic M., Savic A, Susic S., Pongor S., Suck D. (1993) Nucl. Acid Res. **21**(4), 1025-1029

(73) Atherton E., Sheppard R.C. (1989) Solid Phase Peptide Synthesis, IRL PRESS

(74) Scopes T. (1974) Anal. Biochem. **59**, 277

(75) Markovitz A. (1972) Biochim. Biophys. Acta **281**, 522-534

(76) Shettar M.D., (1980) Photochem. Photobiol. Rev. 5 (Smith K.C., ed.) pp. 105-197 Plenum

(77) Talanian R.V., McKnight J.C., Kim P.S (190) Science

(78) O' Neal K.T., Hoess R.H., DeGrado W.F. (1990) Science

(79) Cregut D., Liautard J.P., Heitz F., Chiche L. (1993) Protein Eng. **6**, 51-58

(80) Kriwacki R.W., Schulz S.C., Steitz T.A., Caradonna J.P. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 9759-9763

(81) Benevides J.M., Weiss M.A., Thomas G.J. Jr. (1991) Biochemistry **30**, 4381-4388

(82) Wells J., lecture on Protein Engineering, Spetsai course on Protein Structure, Function and Design (August 30-September 12, 1993))

ACKNOWLEDGEMENTS

This thesis work was carried out at the International Centre for Genetic Engineering and Biotechnology, Protein Structure and Function Group, under the supervision of Prof. Sandor Pongor. Dr. Alessandro Tossi, formerly in the same group and now at the University of Trieste, Dept. of Biochemistry, was of invaluable help for discussion and brilliant advices and for the revision of the manuscript.

I wish to thank Dr. Sotir Zakhariev and Dr. Corrado Guarnaccia for advices and suggestions concerning the syntheses of the various peptides. Moreover I would like to acknowledge Prof. Salvatore Foti, University of Catania, for providing mass spectra of the synthetic peptides and Dr. Annalisa Pastore, EMBL, Heidelberg, for providing mass spectra and useful discussions.

I also wish to acknowledge Dr. Geza Dallman and Prof. Laslo Orosz, ABC, Godollo, for providing the plasmid construct containing the operator region of phage 16-3 and for useful advices.

I would also like to thank Prof. Arturo Falaschi for encouragments throughout my studies and for introducing me to the field of DNA-binding proteins.