Dissertation

submitted to the
Combined Faculties for the Natural Science and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
Ekaterina Lamber, MSc
born in St.-Petersburg, Russia

Oral-examination:

Structural studies on Ets1 and USF1 transcription factor complexes with DNA

Referees: Prof. Dr. Sinning
Dr. Mueller

Table of content

Table of content

Abstract **6**

Zusammenfassung **7**

Abbreviations **8**

1. Introduction

1.1 Initiation of transcription 9

1.2 Regulation of transcription 11

1.3 USF1 transcription factor 12

1.4 Ets1 transcription factor 15

1.5 Stromelysin-1 promoter and Ets1/Ets1/DNA complex 20

1.5.1 Aim of the project focused on Ets1/Ets1/DNA complex 22

1.6 Transcription factors Ets1 and USF1 on HIV1 LTR 22

1.6.1 Aim of the project focused on Ets1/USF1/DNA 23

1.7 USF1 tetramerization 25

1.7.1 Aim of the project focused on USF1 28

2. Results and discussions

USF1 and Ets1 expression and purification

2.1 USF1 expression and purification 29

2.2 Ets1 expression and purification 29

Stromelysin-1 promoter and Ets1/Ets1/DNA complex

2.3 Ets1/Ets1/DNA complex formation 30

2.4 Ets1/Ets1/DNA complex purification 31

2.5 Ets1/Ets1/DNA SAXS experiment 34
2.6 Ets1/Ets1/DNA complex crystallization 39
2.7 Ets1/Ets1/DNA structure determination 51
2.8 Comparison of SAXS model and crystallographic model 56
2.9 Comparison of crystal structure of Ets1/Ets1/DNA complex and Ets1 dimer 57
2.10 Preliminary conclusions 58
2.11 Future perspectives 58

USF1 tetramerization

2.12 USF1/DNA complex formation and purification 59
2.13 SAXS experiment on USF1/DNA complexes 60
2.14 USF1 without DNA - SAXS model 64
2.15 FRET experiment 66
2.16 Rotary shadowing electron microscopy 68
2.17 Crystallization of USF1/DNA complex 71
2.18 Conclusions 71
2.19 Future perspectives 71

Ets1/USF1/DNA complex

2.20 Ets1/USF1/DNA complex formation and purification 72
2.21 Crystallization of Ets1/USF1/DNA ternary complex 73
2.22 Conclusions 74

3. Materials and Methods

3.1 Materials

3.1.1 Chemicals 75
3.1.2 Buffers 75
3.1.3 Media 75
3.1.4 Expression vectors 76
3.1.5 Oligonucleotides 78

3.2 Methods

3.2.1 Sub-cloning **78**
3.2.1.1 Digestion of insert or vector DNA 78
3.2.1.2 Purification of digested insert DNA or digested vector 79
3.2.1.3 Ligation of DNA fragments with sticky ends 79
3.2.1.4 Transformation of plasmid DNA to chemically competent E. coli cells 80
3.2.1.5 Colonies selection 80

3.2.2 Protein expression and solubility test **81**

3.2.3 Protein purification **84**
3.2.4 USF1 expression and purification 84
3.2.5 Ets1 expression and purification 97
3.2.6 Ets1/Ets1/DNA complex formation and purification by gel filtration 89

3.2.7 SDS-PAGE 89
3.2.8 Native gels 90
3.2.9 Protein concentration 90
3.2.10 Protein or protein/DNA complex concentration determination 90

3.2.11 Fluorescence resonance energy transfer (FRET) 91
3.2.12 Rotary Shadowing Electron Microscopy 91

4. List of references **92**

5. Appendix
5.1 Fluorescence resonance energy transfer 97
5.2 Small-angle X-ray scattering 98
5.3 Protein crystallization 103
5.4 Principles of X-ray crystallography 105
5.5 list of references (for Appendix) 112
Acknowlengements **113**

Abstract

Ets1 and USF1 are transcription factors, which were shown to play a role in regulation of transcription on different viral and cellular promoters.

Ets1 has a conserved 85 amino acids DNA binding domain termed as ETS domain surrounded by two autoinhibitory regions. Autoinhibition is released when Ets1 is bound to the DNA.

Ets1 binds cooperatively to two Ets1-binding sites located on the human stromelysin-1 promoter and transactivate it (Baillat et. al., 2002). Stromelysin-1 (matrix metalloproteinase-3) is a major matrix metalloproteinase of connective tissue and is important for tissue remodeling during tissue development, growth, and wound repair (Sternlicht et. al., 1999). Since, stromelysin-1 misregulation can lead to pathological processes development and the Ets1 protein is involved in regulation of stomelysin-1 promoter, understanding of the mechanism of Ets1/Ets1/DNA complex formation is of interest.

Small angle X-ray scattering (SAXS) model for Ets1/Ets1/DNA complex was built. The complex was crystallized. The data set was collected to a resolution of 2.58 Å. The structure was solved by molecular replacement. SAXS model is in a good agreement with crystal structure.

The distal enhancer region of the human immunodeficiency virus 1 (HIV1) long terminal repeat LTR (-130 to -160) is known to be important for transcriptional activity and viral replication in T cells (Sieweke et. al., 1998). The DNA sequence of this region contains binding sites for the transcription factor USF1 (E-box) and for the transcription factor Ets1.

It has been shown that besides the E-box in the distal enhancer, USF1 can bind to two initiator-type elements near the transcription start site of the HIV1 LTR (Du et. al., 1993). Based on spectroscopic and biochemical evidence it has been proposed that USF1 can form homotetramers when bound to two recognition sequences (Ferre-D'Amare et. al., 1994). It was proposed that formation of the bivalent homotetramer may lead to the DNA looping recruiting USF1 and other factors from the distal region of the promoter to the initiator element.

USF1 and USF1/DNA complex were investigated in SAXS experiments. Low resolution *ab initio* model of USF1 monomer was reconstructed using GASBOR program. The tentative model of USF1/DNA bivalent homotetramers was built. It displayed the dimers arrangement similar to the crystallographic structure of Myc-Max heterotetramer (Nair et. al., 2003).

In order to validate tetrameriation two other methods were used. They were fluorescence resonance energy transfer (FRET) and rotary shadowing electron microscopy (EM). USF1 tetramerization was not proved by FRET experiment and by rotary shadowing EM.

Based on yeast one-hybrid screen assay the E-box binding protein USF1 was identified as an interaction partner of Ets1 (Sieweke et. al., 1998). The interaction between USF1 and Ets1 was claimed to be important for full transcriptional activity of HIV1 LTR in T cells. Structural studies on Ets1/USF1/DNA ternary complex were done. Unfortunately, no crystals were obtained.

Zusammenfassung

Ets1 und USF1 sind Transkriptionsfaktoren, die in der Regulierung der Transkription unter der Kontrolle von viralen und zellulären Promotoren eine wichtige Rolle spielen.

Ets1 besitzt eine 85 Aminosäuren lange konservierte Domäne, die eine DNA Bindedomäne ist und Ets Domäne genannt wird. Diese ist von zwei autoinhibierenden Regionen umgeben. Die Autoinhibition findet statt, wenn Ets1 an DNA gebunden ist.

Est1 bindet an zwei Ets1-Bindemotive auf dem Stromelysin-1 Promotor und und bedingt so eine Transaktivierung. Stromelysin-1 (auch genannt Matrix-metalloproteinase-3) gehört zu den Matrix Metalloproteinasen, die wichtige Rollen bei Gewebegenerierung, Wachstum und Wundheilung (Sternlicht et. al., 1999) spielen. Weil die Missregulierung von Stromelysin-1 pathologische Prozesse hervorruft und Ets1 eine Rolle in der Regulation des Stromelysin-1 Promotor spielt, ist es von hohem Interesse, den Mechanismus der Ets1/Ets1/DNA Komplex bildung zu verstehen.

Ein Modell des Ets1/Ets1/DNA-Komplexes basierend auf Daten erhalten aus einem Röntgenkleinwinkelstreuexperiments (SAXS) konnte erstellt werden. Weiterhin wurde der Komplex wurde mit Hilfe der Dampfdiffusionsmethode kristallisiert. Ein Röntgenstreudatensatz mit einer Auflösung von 2.58 Å konnte aufgenommen werden. Die dreidimensionale Struktur des Komplexes wurde dann mit der Methode des Molekularen Ersatzes gelöst. Das SAXS Modell ist vergleichbar mit der Kristall Struktur.

Von der distalen Verstärker-Region des humanen Immunschwäche Virus 1 (HIV1) des langen terminalen Repeat LTR (-130 to -160) weiss man, dass es wichtig für die Transkriptionsaktivität und für die virale Replikation in T-Zellen ist (Sieweke et. al., 1998). Die DNA Sequenz der Region besitzt Bindungsstellen für den Transkriptionsfaktor USF1 (E-box) und für den Transkriptionsfaktor Ets1.

Es gibt mehrere E-boxen auf HIV1 LTR, die USF1 an zwei Initiationselementen in der Nähe des Transkriptionstartelements auf HIV1 LTR binden kann (Du et. al., 1993). Es wurde vorgeschlagen, dass USF1 Homotetramere ausbilden kann (Ferre-D'Amare et. al., 1994). Außerdem wurde vorgeschlagen, dass die Bildung der bivalenten Homotetramere kann eine DNA Loop-Bildung provozieren kann. Dies wiederum kann zur Rekrutierung von USF1 und anderen Transkriptionsfaktoren der distalen Region des Promotor zum Initiatorelement führen.

Der USF1 und USF1/DNA Komplex wurde mittels SAXS Experimenten untersucht. Ein *ab initio* Modell bei niedriger Auflösung des USF1 Monomers wurde mit Hilfe des Programms GASBOR erstellt. Ein vorläufiges Modell eines bivalenten Homotetramers wurde ebenfalls gebaut. Die Komplex weist die gleiche Position der Moleküle wie in der Struktur des Myc-Max Heterotetramers auf (Nair et. al., 2003).

Um die USF1 Tetramerisierung experimentall zu überprüfen, wurden zwei weitere Methoden eingesetzt: Fluoreszenz Resonanz Energie Trransfer (FRET) und rotary shadowing Elektronenmikroskopie (EM). Allerdings konnte die USF1 Tetramerisierung weder mittels FRET noch mittels rotary shadowing EM bestätigt werden.

Das E-box Bindeprotein USF1 wurde wie auch der Wechselwirkungspartner für Ets1 mit dem Hefe one-hybrid Screen Assay gefunden (Sieweke et. al., 1998). Die Wechselwirkung zwischen Ets1 und USF1 ist vermutlich für die Transkriptionsaktivität des HIV1 LTR in T-Zellen wichtig. Strukturelle Untersuchungen des Ets1/USF1/DNA wurden durchgeführt, allerdings wurden keine Kristalle erhalten.

Abbreviations

HIV1	human immunodeficiency virus
LTR	long terminal repeat
dsDNA	double-stranded DNA
BHLHZip	basic-helix-loop-helix-leucine zipper
USF1	upstream stimulatory factor 1
SAXS	small angle X-ray scattering
FRET	fluorescence resonance energy transfer
EM	electron microscopy
MW	molecular weight
CTD	C-terminal domain
RT	reverse transcriptase
UAS	upstream activating sequences
HLH	helix-loop-helix motiv
Zip	leucine zipper motif
b	basic region
VEGF	vascular endothelial growth factor
kDa	kilo-Dalton
DR	dummy residues
Inr	Initiator element
WT	wild type
BSA	bovine serum albumin
bp	base pairs
AUS	upstream activating sequence
LEF 1	lymphoid enhancer factor 1
TAD	transactivation domain
USR	USF-specific region
ds DNA	double-stranded DNA
EBS	Ets1-binding site
Sp1	specificity protein 1
NF-kB	nuclear factor kB

1. Introduction

1.1 Initiation of transcription

Transcription defines the process in which RNA is synthesised by RNA-polymerase on the matrix of DNA. The important feature about transcription is the choice of fixed positions where the synthesis starts (transcription initiation) and where it finishes (transcription termination). Transcription starts upstream of the initial transcription sequence and the starting point of transcription is labelled +1. A promoter is located at the 5'-end of starting point. The promoter is defined as the piece of matrix DNA required for the initial binding of RNA-polymerase and transcription initiation complex prior to transciption.

The following is a brief overview of transcription initiation by eukaryotic RNA-polymerase. The eukaryotic promoters can have two basic elements functioning together or independently. The first of them is a **TATA-box** situated 25 base pairs (bp) along from the 5'-end of the initiation point, having the consensus sequence TATAa/tAa/t. The second is an **initiator element** (Inr), a pyrimidine–rich sequence having consensus YYANt/aYY (where Y is a pyrimidine, and N is any base). The initiator element is situated close to starting point of transcription (fig. 1).

The strongest promoters have both elements, but some contain only one of them. These elements (TATA and Inr) are called core promoter elements because they are required for the proper initiation of the transcription by RNA-polymerase in cell-free system.

RNA-polymerases by themselves are not able to recognise the promoters. For proper transcription initiation, basal transcription factors are required. Each of 3 RNA-polymerases has its own set of basal transcription factors. RNA-polymerase II, for example, contains transcription factors TFIID, TFIIB, TFIIE, TFIIF and TFIIH. These factors bind sequentially to the promoter together with RNA-polymerase and form pre-initiation complex. The amount of basal transcription factors in the cell is much higher than the amount of RNA-polymerase. Fig. 1 illustrates how the pre-initiation complex is formed on the TATA-box containing promoters (Kalinin 2001 p. 27).

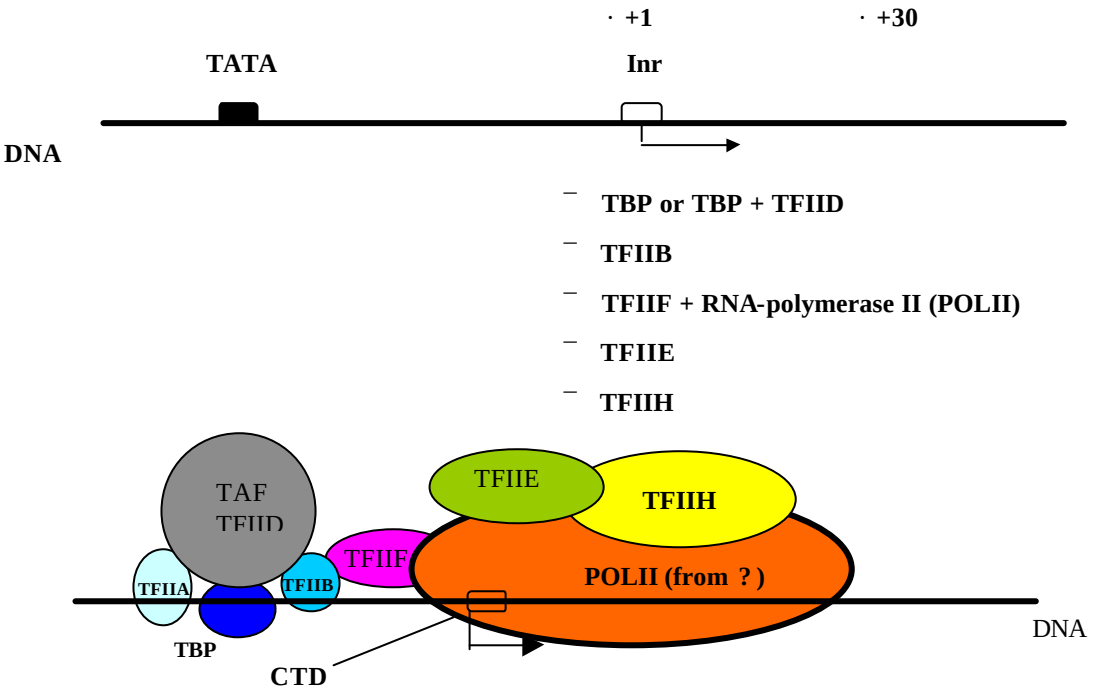


Fig. 1 **Initiation of transcription**
(CTD is C-terminal domain of RNA-polymerase)

TFIID is the first transcription factor, which binds to the promoter. This factor contains TBP (TATA-binding protein) binding specifically to the TATA-box and TAF_{II} (TBP-associated factors). TBP induces DNA bending when bound to its minor groove.

The next transcription factor in the pre-initiation complex is TIIB. Its N-terminal domain binds to TFIIF/RNA-polymerase II complex, its C-terminal domain binds to the DNA and to TBP. TFIIB does not change the structure of the TFIID/DNA complex but stabilises it. After the formation of TFIIB/TFIID/DNA complex, the promoter is ready to bind RNA-polymerase II.

RNA-polymerase II usually forms a complex with the TFIIF transcription factor. The most important function of TFIIF is to support the interaction between RNA-polymerase II and TFIIB.

The fourth transcription factor is TFIIE, which interacts with RNA-polymerase II and probably with TFIIF.

Another transcription factor playing a role in the initiation of transcription is TFIIA. It has three functions: it stabilises the TBP interaction, enhances the interaction between TAF and DNA and has a coactivation function.

The last step in the formation of the pre-initiation complex is binding of TFIIH transcription factor. TFIIH has several enzymatic activities. It can play the role of ATP-dependent helicase. In the presence of ATP, formation of pre-initiation complex leads to melting of the DNA in the region close to transcription starting point. After this TFIIH phosporylates the C-terminal domain of RNA-polymerase II, activating elongation. As a result of phosporylation, the interaction between RNA-polymerase II and specific initiation transcription factors becomes weaker and RNA-polymerase II is able to leave the promoter and to start the elongation, TFIID stays on the promoter and can initiate the transcription again (fig. 1).

1.2 Regulation of transcription

On the basis of estimates of the intracellular concentrations of RNA-polymerase II and basal transcription factors, it can be concluded that they are constant and not dependent on the cell-type. In the absence of an additional regulation of transcription all the genes would be transcribed at the same speed equal to the power of their promoters. This would lead to a loss of the cell's ability to regulate different processes and differentially express the genes. The differential expression is determined by epigenetic information present in the each cell type and dependent on the internal and external factors. In the regulation of differential expression, the system of transcription factors transmitting the signals to the specific genes plays a crucial role. These factors can either decrease or increase the transcription level of certain genes in comparison to the basal level. If activation occurs transcription factors are called ***transcription activators***, whereas for repression they are called ***transcription repressors***. In both cases they are special proteins, which bind to regulatory elements.

The sequences recognized by transcription activators are often situated at the 5'-end of the promoter and called ***5'-upstream activating sequences (UAS).*** Usually, they are placed between positions –100 and –150 for eukaryotes. Apart from the

proximal activator sequences eukaryotes also have ***enhancers.*** The enhancers have also specific sites where transcription factors can bind, and can regulate the promoter from large distances (up to several thousand base pairs) and are not dependant on the promoter orientation and can be situated either at the 5'-end of the promoter or at the 3'-end (Kalinin 2001 p. 42).

The ability of the enhancers to stimulate transcription when they are situated at a distance from the promoter, is possible by the formation of DNA loops. Many transcription factors can bend DNA. For example, when TBP is bound to TATA-box the DNA bends by $70\text{-}80^0$. There are also some proteins which do not play a role as transcription activators but can bend DNA, and as a result distant DNA pieces come in close contact. This leads to the formation of the ***enhancosome*** where different proteins-activators bound to the enhancer are placed close to each other and close to the promoter region. An example of such a protein is lymphoid enhancer factor 1 (LEF-1), which belongs to the high mobility protein group and recognises the consensus sequence CCTTTGAA. As a result of the LEF-1 binding, DNA bends by 130^0 (Kalinin 2001 p. 58). Such a bending can lead to the activation of the promoter because it allows direct contact between the transcription factors bound to the enhancer and these bound to the promoter.

This PhD thesis focused on transcription factors Ets1 and USF1. Both of the transcription factors have binding sites on different promoter and play a role in the regulation of those promoters.

1.3 USF1 transcription factor

USF1 (upstream stimulatory factor 1) belongs to the basic helix-loop-helix leuzine zipper (bHLHZip) transcription factor family. This family of eukaryotic transcription factors characterised by a highly conserved DNA binding domain composed of a basic region (b), followed by helix-loop-helix (HLH) and leucine zipper (Zip) motifs (fig. 2).

The contiguous presence of the HLH and the leucine zipper, two dimerization interfaces, distinguishes these proteins from both the bHLH and b/Zip transcription

factors. b/HLH/Zip family members occur widely, ranging from the mammalian nuclear proteins Myc (Murre et. al., 1989), Mad (Ayer et. al., 1993), Max (Blackwood and Eisenman, 1991), Mxi1 (Zervos et. al., 1993), USF (Gregor et. al., 1990), TFEB (Carr and Sharp, 1990), TFE3 (Beckman et. al., 1990) and AP-4 (Hu et. al., 1990), to the yeast protein CBF-1 (Cai and Davis, 1990), among many others. All of these proteins bind to the common CANNTG element known as the E-box.

USF1 was first characterized as a transcription factor, which binds to an upstream element of the adenovirus major late promoter in HeLa cell nuclei (Carthew et. al., 1985; Miyamoto et. al., 1985; Sawadogo and Roeder, 1985) stimulating transcription possibly by direct interaction with the basal factor TFIID (Sawadogo, 1988; Bungert et. al., 1992). Extensive purification of the HeLa nuclear extract yielded two polypeptides (Sawadogo et. al., 1988), the smaller of which has been cloned and sequenced revealing a protein of molecular mass 44 kDa with a highly conserved b/HLH/Zip DNA binding domain near its C-terminal (Gregor et. al., 1990). The recombinant protein expressed from this cDNA clone (which shall be henceforth referred to as USF) homo-oligomerized efficiently, bound DNA containing a CACGTG E-box motif with nanomolar affinity and activated transcription in a manner undistinguishable from that of material purified from HeLa nuclear extracts (Pognonec and Roeder, 1991).

The domain structure of USF1 is presented at fig. 2.

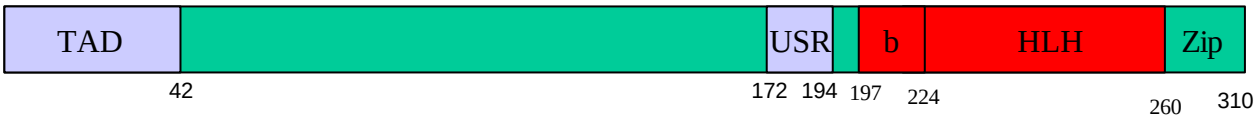


Fig. 2 **The domains of the USF1 protein**
(TAD – transactivation domains, USR – USF specific region)

USF1 has two transactivation domains: one situated at the N-terminal part of the protein and another one situated in the middle of the protein called USF-specific region (USR).

The crystal structure of USF1 bHLH domain bound to DNA from adenovirus major late promoter was solved (Ferre D'Amare et. al., 1994) to a resolution of 2.9 Å (fig. 3). The $(bHLH)_2$-DNA complex folds into a parallel, left-handed four-helix

bundle, which has a topology identical to the structure of another bHLHZip transcription factor Max bound to DNA from adenovirus major late promoter (Ferre-D'Amare et. al., 1993) (fig. 4).

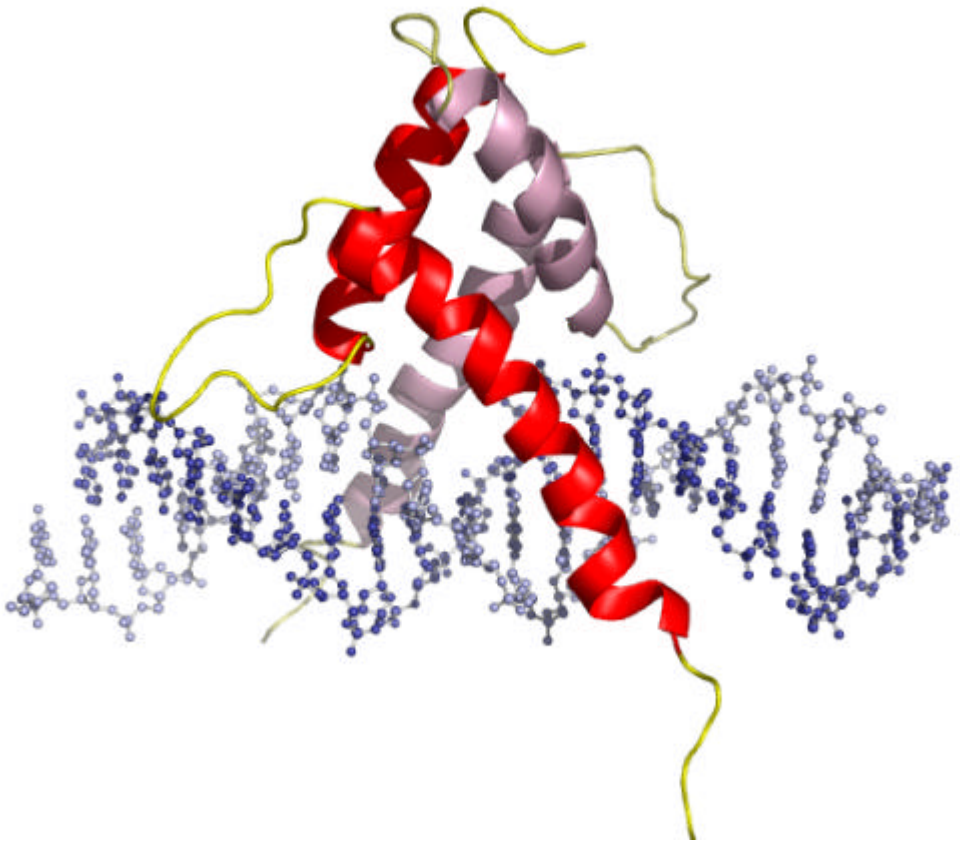


Fig. 3 **Structure of USF1 bHLH bound to DNA** (Ferre D'Amare et. al., 1994)
USF1 protein contained amino acid residues 197-260. dsDNA sequence is shown below (core binding element is colored in red).

5' CACCCGGTCACGTGGCCTACA
TGGGCCAGTGCACCGGATGTG 5'

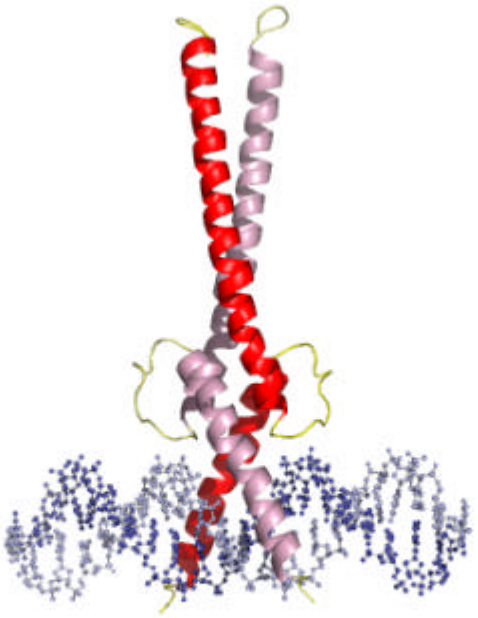


Fig. 4 **Structure of Max bound to DNA** (Ferre D'Amare et. al., 1994)
Max protein contained amino acid residues 22-113. Double-stranded DNA (dsDNA) sequence is shown below (core binding element is colored in red).

5' CACCCGGTCACGTGGCCTACAC
GTGGGCCAGTGCACCGGATGTG 5'

USF-1 is expressed in a variety of tissues and therefore it has been shown to act on a variety of cellular and viral promoters.

USF1 is highly abundant in the eukaryotic nuclei. In the HeLa nuleus the USF concentration can be estimated as 0.5 µM, assuming homogeneous distribution throughout a spherical nucleus of radius 2.5 µm with 20000 molecules/cell (Sawadogo et. al., 1988).

1.4 Ets1 transcription factor

The Ets1 proto-oncoprotein is a member of the Ets family of transcription factors that share a unique DNA binding domain, the Ets domain. The name "Ets" stems from a sequence that was detected in an avian erythroblastosis virus, E26, where it formed a transforming gene together with ? gag and c-myb. The newly discovered sequence was called E26 transforming specific sequence or Ets. Later, a

cellular homologue to the viral ets (*v-ets*), was found suggesting that *v-est* derived from *c-ets1* (Watson et. al., 1985; Ghysdael et. al., 1986).

Up to now more than 45 members of the Ets family have been characterized as transcriptional activators and inhibitors in eukaryotes.

The Ets1 protein is produced by a variety of tissues. Ets1 is expressed in lymphoid tissues (Ghysdael et. al., 1986). Ets1 is also detected in other tissues. The two main blood vessel forming types, endothelial cells and vascular smooth muscle cells, transiently produce Ets1 upon activation by angiogenic factors. Ets1 is produced by a variety of solid tumors, including epithelial tumors, sarcomas and astrocytomas. In addition to advanced solid tumors, high Ets1 expression has also been found in leukemic T-cells. Ets1 is expressed in certain cells of ovary, in hepatic stellate cells as well as in glandular epithelial cells and stromal cells of the endometrium during menstrual cycle. Ets1 is expressed in a variety of tissues throughout the embryonal development.

The human *ets1* gene contains eight exons (A, III-IX). Only two proteins are generated from RNAs, $p54^{c\text{-}ets1}$ (full length Ets1) and $p42^{c\text{-}ets1}$ (?VII-Ets1) (fig. 5).

The Ets1 protein has pointed domain, transcativation domain (TAD), DNA-binding domain (ETS) surrounded by two autoinhibitory regions (amino acid residues 301-331 and 415-441) (fig. 5).

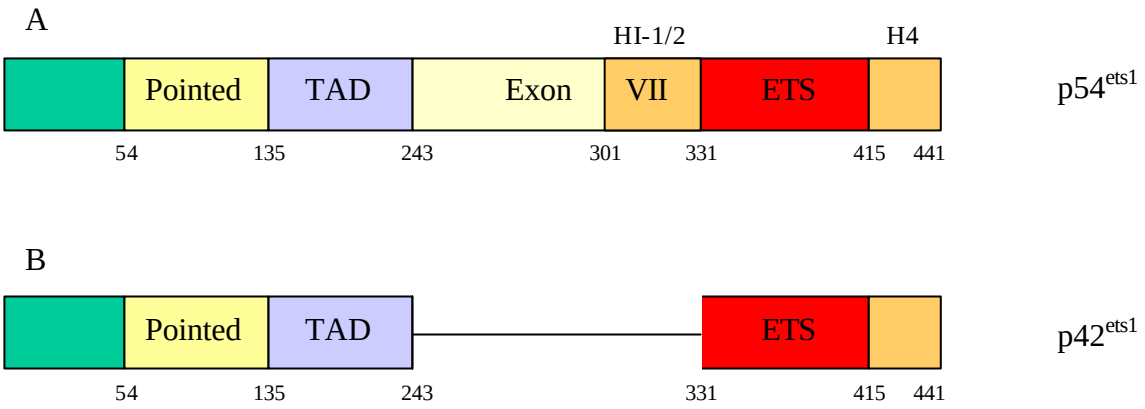


Fig. 5 **The domains of the Ets1 protein**

A. $p54^{c\text{-}ets1}$ (full length Ets1)

B. $p42^{c\text{-}ets1}$ (natural isoform of Ets1 where exon ? VII was deleted)

The Ets domain, composed of 85 amino acids, comprises three a-helixes and four ß-strands that are arranged in the order H1-S1-S2-H2-H3-S3-S4 (fig. 6 A). The Ets domain specifically recognizes DNA sequences that contain a GGAA/T core element (Nye et. al., 1992). However, Ets proteins differ significantly in their preference for the sequence flanking the GGAA/T core motif.

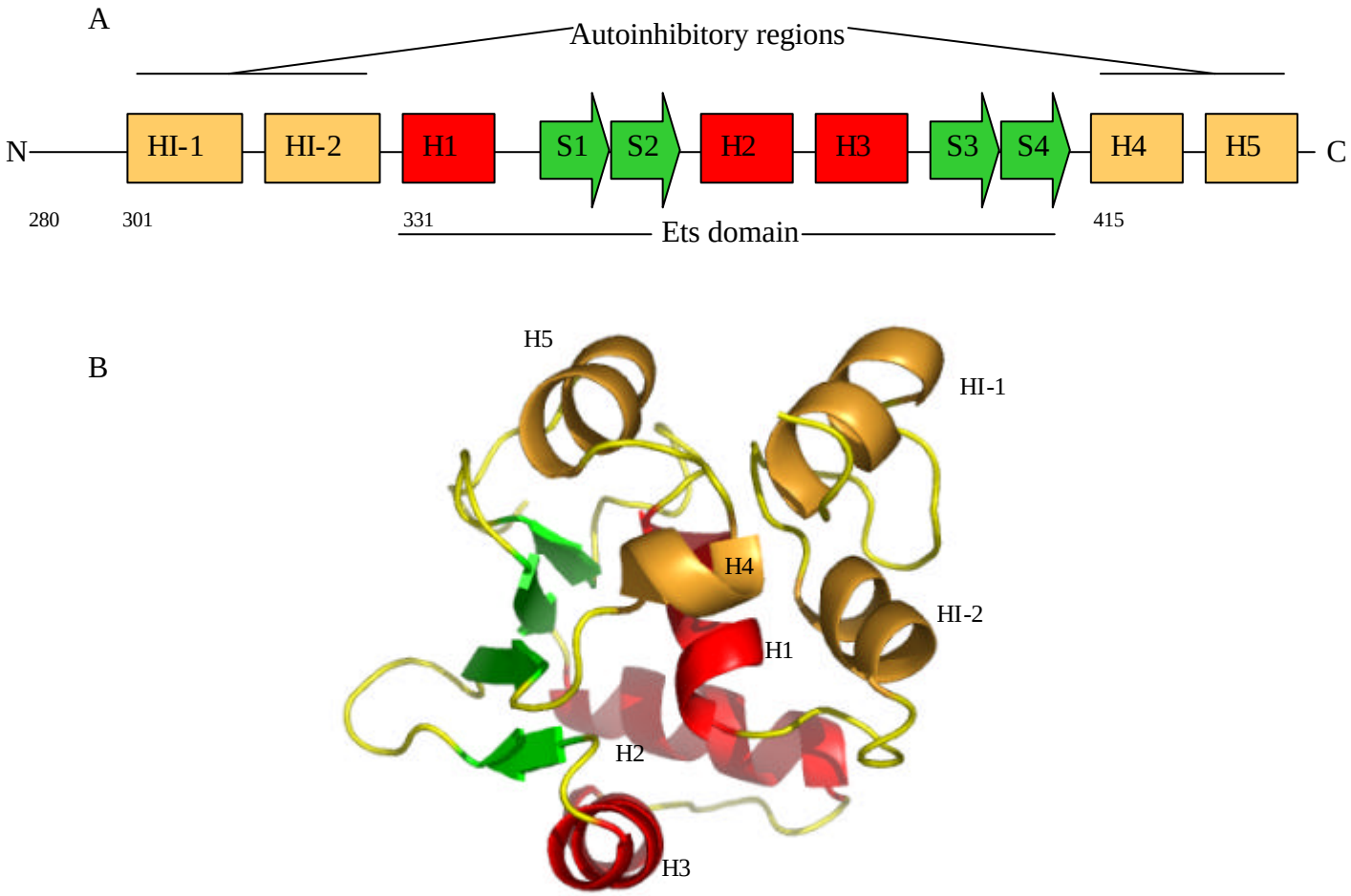


Fig. 6 **A model of the autoinhibition mechanism of the Ets1 protein**

A. Ets domain surrounded by two autoinhibitory regions (Lee et. al., 2003)

B. Structure of the Ets1 protein (residues 301-440) without DNA

In the Ets1 protein, the Ets domain stretches from residue 331 to residue 415. Ets domain is surrounded by two autoinhibitory regions and autoinhibition is released when Ets1 is bound to the DNA. There are three inhibitory helixes: HI-1 and HI-2 within exon VII domain (301-331 amino acid residues) and H4 at the C-terminus

(415-440 amino acid residues) (fig. 6 A). These helixes cooperatively block Ets DNA binding activity by interacting with H1-helix and, thereby, freeze the Ets domain in a closed conformation. Fig. 6 B shows NMR structure of Ets1 (residues 301-441) in an autoinhibited conformation. The blockage is transient and can be relieved when Ets1 binds to DNA and the HI-1 helix unfolds to form a random coil.

The structure of the Ets1 protein bound to DNA was (fig. 6C). The helix H3 of Ets1 binds to the major groove and leads to DNA bending. The "wing" is formed by the loop between strands 3 and 4 of the ß-sheet, makes contacts with the 5' minor groove (Werner et. al., 1997). The distantly related Ets factor PU.1 binds in a similar way to DNA (Kodandapani et. al., 1996) demonstrating that the Ets domain/DNA interaction is highly conserved.

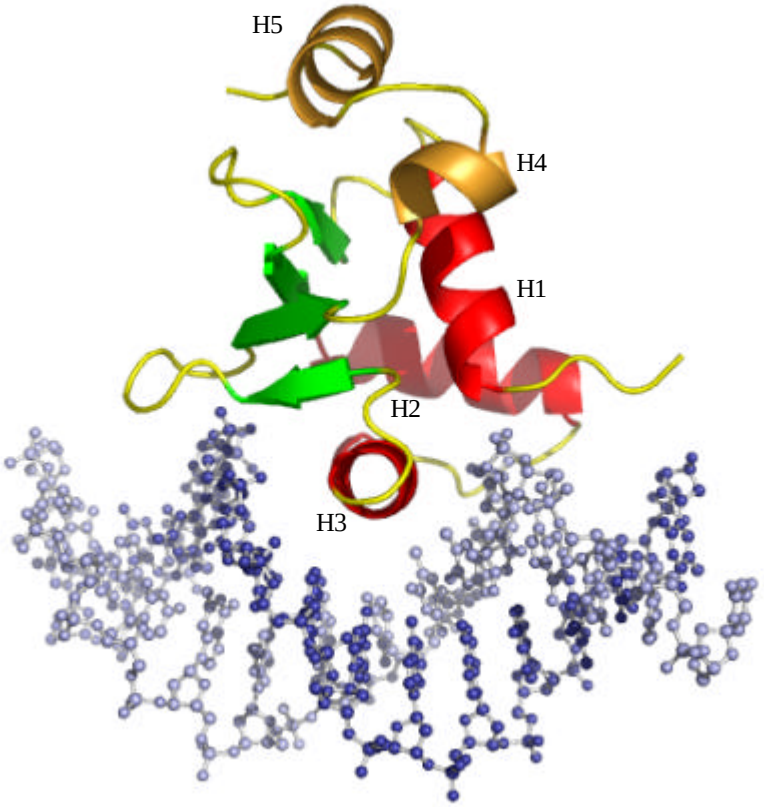


Fig. 6C **Ets domain-DNA interaction** (Garvie et. al., 2001)
Ets1 protein contained amino acid residues 331-440. dsDNA sequence is shown below (core binding element is colored in red)

5' TAGTGCCGGAAATGT
TCACGGCCTTTACAA 5'

Recently, the structure of Ets1 (residues 297-441) without DNA was solved. This structure shows domain swapping in which N-terminal parts of the protein are involved (fig. 7).

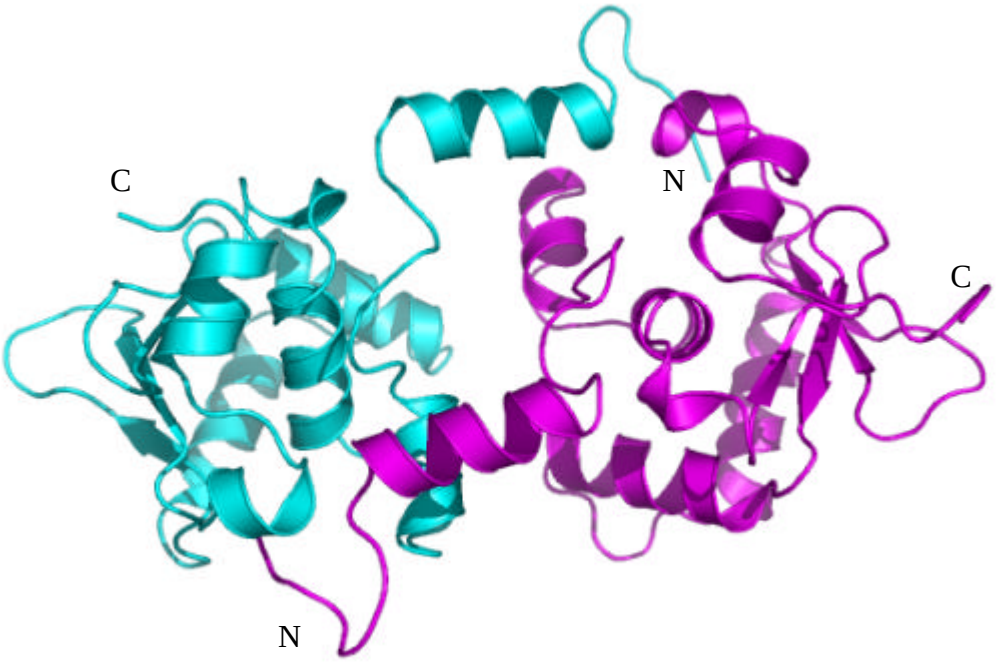


Fig. 7 **Structure of Ets1 dimer in the absence of DNA** (Tahirov et. al., 2002 to be published)
Ets1 dimer structure demonstrates domain swapping (one Ets1 molecule is coloured in cyan, another one is coloured in magenta).

Ets1 interacts with a variety of transcription factors, other proteins and with itself. In the presence of the palindromic Ets binding site on the DNA, the exon VII domain also mediates homodimerization of Ets1 proteins (Baillat et. al., 2002). The dimerization blocks the autoinhibitory mechanism allowing these proteins to mutually increase their DNA binding activities and to bind cooperatively to DNA. The Ets1 binding sequence was shown to be situated 12 bp apart from the E-box sequence which binds USF1 in the distal enchancer of HIV1 LTR and direct interaction between the two proteins was demonstrated (Sieweke et. al., 1998).

1.5 Stromelysin-1 promoter and Ets1/Ets1/DNA complex

Stromelysin-1 (matrix metalloproteinase-3) is a major matrix metalloproteinase of connective tissue and is important for tissue remodeling during tissue development, growth, and wound repair (Sternlicht et. al., 1999). Stromelysin-1 misregulation can lead to the development of the diseases such as rheumatoid and osteoarthritis (Flannery et. al., 1992; Malemud et. al., 1999), Alzheimer's disease (Yoshiyama et. al., 2000), tumor invasiveness, and metastasis (Zucker et al., 2000; Nelson et. al., 2000; Liu et. al., 2001). In addition, it was recently reported (Sternlicht et. al., 1999; Sternlicht et. al., 2000) that stromelysin-1 by itself promoted mammary carcinogenesis in a mouse model system. Stromelysin-1 expression is mainly controlled at the transcription level. A number of specific DNA elements in the human stromelysin-1 promoter have been shown to be important for the regulation of its transcription (Ye et. al., 1999; Rekdal et. al., 2000). Many cytokines and growth factors such as interlikin-1ß, tumor necrosis factor-a activate stromelysin-1 gene transcription.

Two palindromic head to head Ets1-binding sites (EBS) with a 5'-GGA(A/T)-3' core motif at the positions -216 to -209 and at -208 to -201 were found to be important for stromelysin-1 promoter regulation. Ets1, Ets2 and PEA3 (members of Ets transcription factor family) were reported to activate stromelysin-1 promoter (Wasylyk et. al., 1991). Nevertheless, in a heterologous system Ets1 and Ets2 showed no cooperative binding to the palindrome (Wasylyk et. al., 1991). In contrast, other studies proved that the functional EBS palindromes present in GATA-1 and p53 promoter (Seth et. al., 1993; Venanzoni et. al., 1996) cooperatively bound Ets1 and Ets2.

Increased level of coexpression of both Ets1 and stromelysin-1 has been shown for rheumatoid arthritis, glomerylonephritis, angiogenesis and tumor invasion. This leads to the hypothesis that misregulation of stromelysin-1 could be mediated by Ets1 (Bajllat et. al., 2002).

The mechanism of Ets1 binding to two palindromic Ets1-binding sequences located on the stromelysin-1 promoter (positions -216 to -209 and at -208 to –201) was investigated (Bajllat et. al., 2002) by several methods: by surface plasmon resonance, by electrophoretic mobility shift assay and by photo-crosslinking

experiments. It was shown that the full-length human Ets1 protein (p51) binds with a positive cooperativity to the EBS palindrome of the human stromelysin-1 promoter. This cooperativity is due to head to head topology of Ets-1 binding sites. Different topology of the binding sites abolish cooperativity. Studies with N-terminal deletion mutants of Ets1 revealed that the 245–330-residue region of the protein encoded by the exon VII of the gene is important for the cooperativity and it was proposed that autoinhibitory regions of Ets1 are important for the cooperativity as well.

The p42 (natural isoform of Ets1), lacking exon VII, is unable to bind cooperatively to the palindrome, despite a better binding to each individual EBS. Transient transfection experiments show for p51 a good correlation between DNA binding and promoter transactivation. In contrast, p42 shows a poorer transactivation reinforcing the significance of cooperative binding for a full Ets1-mediated transactivation of the promoter. This is the first time that Ets1 is shown to be able to counteract its own autoinhibition (Baillat et. al., 2002).

Based on kinetic and equilibrium analyses of Ets-1 interaction with the EBS palindrome of the stromelysin-1 promoter, the model of the mechanism by which Ets1 cooperative binds to the stromelysin-1 promoter was proposed (fig. 8) (Baillat et. al., 2002).

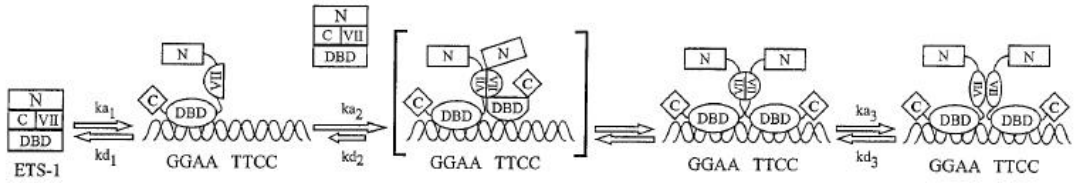


Fig. 8 **Proposed model for ETS-1 cooperative binding to the stromelysin-1 promoter.** *N* and *C* represent the N-terminal and C-terminal part of Ets-1, respectively. DBD is Ets1-binding domain, VII is Ets1 domain encoded by exon 7. It was proposed that after binding to DNA, the first Ets-1 molecule facilitates the binding of the second one by positioning and helping it to reach its uninhibited state through a contact involving the exon VII-encoded region. The intermediary state where the first Ets-1 molecule contacts the second one before the complete formation of the ternary complex is represented by *square brackets.* After formation of the complex, Ets-1 molecules are able to adopt a relaxed conformational state. (Figure is taken from Bajllat et. al., 2002).

It was proposed that it could be a conformational change between the exon VII-encoded regions in the initial stages of the Ets1/Ets1/DNA complex formation and some rearrangments could occur once again when the ternary complex is formed. It was proposed that during Ets1/Ets1/DNA complex formation, first one Ets1 molecule binds to the promoter and facilitate the binding of the second Ets1 molecule (fig. 8).

However, there is no experimental evidence of direct protein-protein interaction. It is predicted that protein-protein interaction should involve an autoinhibition counteraction process and could lead to conformational change in Ets1 protein.

Since, stromelysin-1 misregulation can lead to pathological processes development and the Ets1 protein is involved in regulation of stomelysin-1 promoter, understanding of the mechanism of Ets1/Ets1/DNA complex formation is of biological interest.

X-ray crystallography gives a possibility to inverstigate protein-protein interaction in Ets1/Ets1/DNA complex and conformational change which could occur in DNA or in one of Ets1 molecules or in both of them when Ets1/Ets1/DNA complex is formed. It can help to understand the mechanism of the complex formation and to answer the question why Ets1 (p51) bind cooperatively to stromelysin-1 promoter when p42 (natural isoform of Ets1 lacking exon VII) can not, more precisely, how domain encoded by exon VII can facilitate cooperative interaction between two Ets1 molecules. The last can have a biological significance because p42, in contrast to full-length Ets1, can poorly transactivate stromelysin-1 promoter.

Aim of the project:

Thus, the aim of this project was to obtain structural information on Ets1/Ets1/DNA complex and to understand the mechanism of Ets1/Ets1/DNA complex formation.

1.6 Transcription factors Ets1 and USF1 on HIV1 LTR

The distal enhancer region of the human immunodeficiency virus 1 (HIV1) long terminal repeat LTR (-130 to -160) is known to be important for transcriptional activity for the transcriptional regulation of HIV1 and viral replication in T-cells.

Several transcription factors were shown to interact with and regulate HIV1 promoter (fig. 9). Among them there specificity protein 1 (Sp1) (Perkins et. al., 1993) and nuclear factor kB (NF-kB) which are located in proximal promoter element (Duh et. al., 1989, Israel et. al., 1989, Osborn et. al., 1989). Sp1 has three binding sites and NF-kB has two binding sites in the proximal part of HIV1. Both Sp1 and NF-kB activate HIV1 promoter.

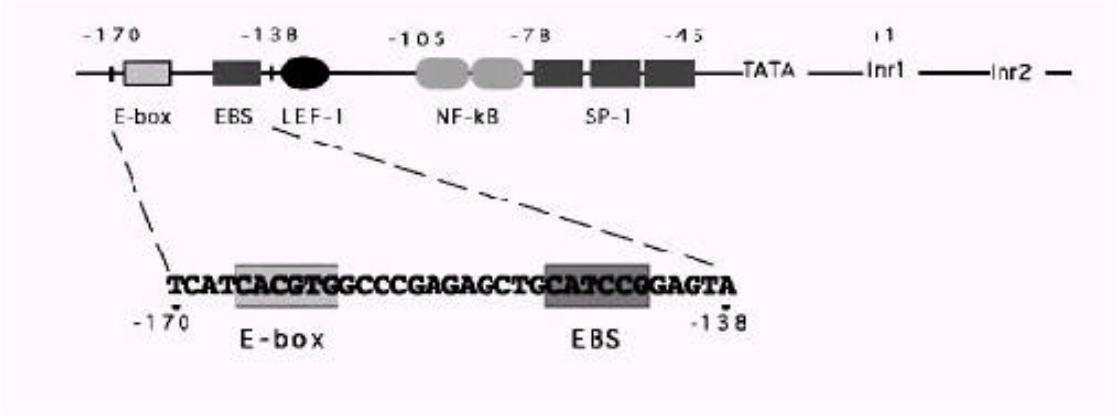


Fig. 9 **Transcription factors on HIV1 LTR**
(Figure is taken from Sieweke et. al., 1998)

It was shown that distal part of HIV1 LTR is also important for HIV1 promoter activation. Mutations in this region (–130 to –166 bp) completely prevent virus replication (Kim et. al., 1993) and reduce the activity of LTR-reporter constructs in transient transfection assays (Zeichner et. al., 1991). Binding sites for USF1, Ets1 and LEF1 transcription fators were found in this region. LEF1 is known as DNA-bending transcription factor. The DNA-bending by LEF1 factor could recruit the factor from the distal part of HIV1 LTR to the initiator region. Both Ets1 and LEF1 are highly expressed in T-cells, and *in vitro* transcription assays with reconstituted chromatin revealed that these two transcription factors in conjunction with Sp1 transcription factor could relieve nucleosomal repression of the HIV1 LTR (Sheridan et. al., 1995).

It was suggested that interaction of Ets1 with other transcription factor could be important for HIV1 promoter regulation (Sieweke et. al., 1998). The study was made to search for interaction partners of Ets1. This was done by yeast one-hybrid screen assay developed by M. Sieweke (Sieweke et. al., 1996). USF1 was identified as an interaction partner of Ets1. It was shown that these proteins interact via their DNA-binding domains in the absence of DNA (Sieweke et. al., 1998). By band shift assay binding of Ets1 and USF1 to distal part of HIV1 LTR was demonstrated (Sieweke et. al., 1998). Using reporter gene assay it was shown that both of the proteins Ets1 and USF1 transactivate HIV1 promoter. The level of activation was higher when both proteins were used in the study than for the individual proteins.

A physical and functional interaction between Ets1 and USF1 transcription factors was identified. The two proteins form a specific ternary complex on the adjacent E-box and Ets binding site of the distal enhancer region of the HIV-1 LTR and synergize in both DNA binding and transcriptional activation. The interaction appears to be important for T-cell expression of the HIV-1 LTR since mutations of each individual DNA binding site drastically reduce the activity of the enhancer element in these cells.

Since it was shown that transcription factors Ets1 and USF1 play a role in activation of HIV1 LTR, the question would be to understand the mechanism by which these two transcription factors can participate in the regulation of HIV1 transcription.

However, little is known about atomic details of interaction between Ets1 and USF1 when both proteins are bound to DNA. The distance between Ets1 and USF1 binding sites is 12 bp which is a bit more than one helical turn on DNA. Thus, the interaction between two proteins should involve either DNA bending or huge conformation change in the proteins or both. The crystal structure of the ternary complex could describe how these two proteins could interact and what happens with DNA conformation when the ternary complex is formed.

Aim of the project:

The aim of this project was to obtain structural information on ternary complex formed by USF1, Ets1 and DNA in order to understand atomic details of the

ternary complex formation, which could lead to understanding of the mechanism of HIV1 transcription regulation by transcription factors Ets1 and USF1.

1.7 USF1 tetramerization

The HIV1 E-box sequence binds USF1 protein as a homodimer, is situated in the distal enhancer region of HIV1 LTR. The mutations in the E-box showed 93% reduction in activity of the reporter in Jukat cells in the reporter gene assay (Sieweke et. al., 1998). It was shown that apart the E-box USF1 can also bind to initiator elements of HIV1 (Du et. al., 1993). Two binding sites for USF1 were discovered in this region: Inr1 (positions from –5 to +9) and Inr2 (positions from +29 to +42).

By gel retardation assay specific binding of USF1 to Inr1 and Inr2 was proven. Inr1 and Inr2 are conserved. Similar elements were found in the adenovirus major late promoter. They are situated at the positions –3 to +20 for Inr1 and at the positions +35 to +60 for Inr2. Series of Inr1 and Inr2 mutations were generated in order to assay the role of these 2 conserved DNA sequences (Du et. al., 1993). Mutated templates were compared with the wild-type samples both by *in vivo* transfection assay and by *in vitro* transcription assays to determine effects on both promoter strength and accuracy of initiation. In the transfection assay most of the mutations in the Inr1 and the Inr2 sites reduced the level of stimulation by USF1.

Another example of multiple USF1 binding sites is human telomerase reverse transcriptase promoter. There are two E-boxes: upstream (positions –165 to –160) and downstream (positions +44 to +49), USF binds to the E-boxes primary as a homodimer USF1/USF2. Mutations of each of these E-boxes significantly decreased promoter response to USF and mutations in both sites further impaired the ability of USF to activate the hTERT promoter (Goueli et. al., 2003).

Steady state fluorescence spectroscopy demonstrated that the b/HLH/Zip domain of USF binds its DNA targets with high affinity and specificity, whereas removal of the leucine zipper yielding the b/HLH minimal DNA binding region reduces both affinity and specificity (Sha et. al., 1995). Stopped flow kinetics provided evidence for a two-step binding process involving rapid formation of a protein-DNA intermediate followed by a slow isomerization step, which is consistent

with the basic region undergoing a random coil to α-helix transition on specific DNA recognition (Sha et. al., 1995). The leucine zipper is also necessary for USF to function as a bivalent homotetramer, capable of binding two distinct recognition sites simultaneously and mediating DNA looping under physiological conditions. Titration studies revealed that the first binding event has an equilibrium constant K_{eq} = (2.2 ± 0.8) x 10^9 M^{-1} for major late promoter DNA, whereas the second binding event occurs with remarkably reduced affinity, K_{eq} = (1.2 ± 0.8) x 10^8 M^{-1}. This anticooperative feature of DNA binding by homotetramer suggests that USF stimulates transcription by mediating DNA looping between nearby recognition sites located in nuclear and viral gene promoters. The model for the unusual anticooperative is shown on fig. 10. The proposed model suggests DNA looping and mismatch correction in the transcription system (Sha et. al., 1995).

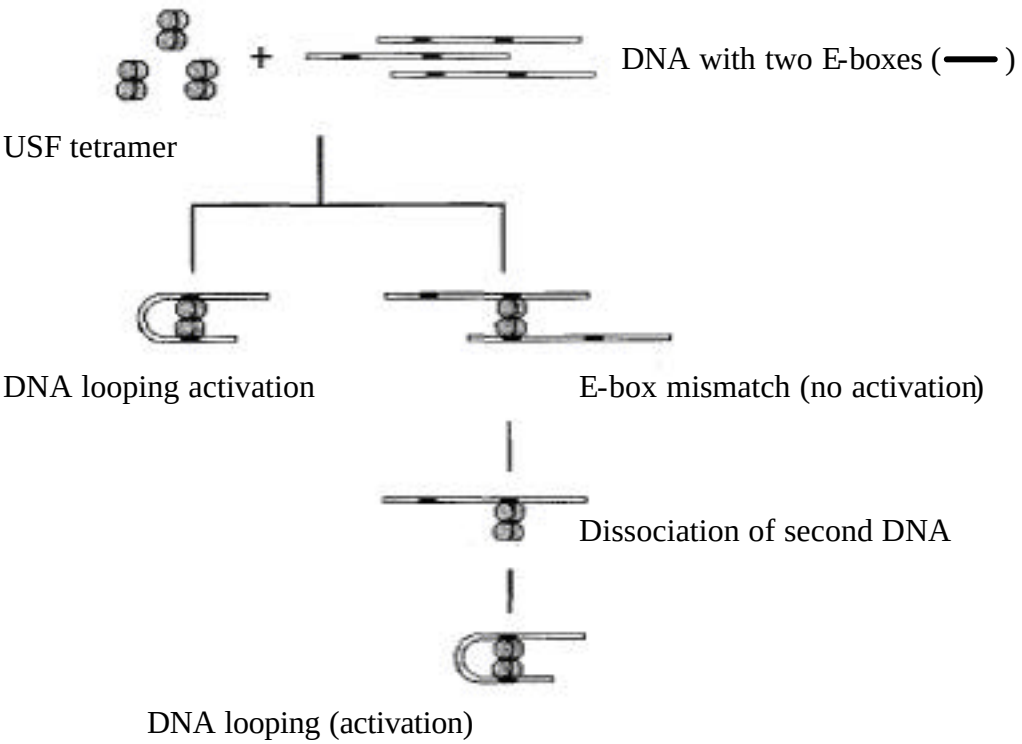


Figure 10. **Model of selectivity for looped DNA/USF1 complex formation by anticooperative binding during transcription activation** (Figure is taken from Sha et. al., 1995)

Furthermore, a LEF1 binding site located between USF1 binding sites on HIV1 LTR leads to DNA bending, and the DNA bending produced by LEF1 factor could be stabilized by USF1 bivalent homotetramer formation. This could bring together transcription factors on the HIV1 LTR and recruit transcriptional machinery.

It seems that if bivalent homotetramer is formed USF1 could play a role of scaffold protein bringing together distant transcription factors, provide a platform for multiprotein complex formation, and regulate transcription in this way. Also it could be a general mechanism of transcription regulation via USF1 homotetramer formation and DNA loop formation because multiple USF1 binding sites situated far from each other were found on several human and viral promoters transcribed by RNA-polymerase II. An example is the Myc-Max tetramer. Its crystal structure was solved (Nair et.al., 2003). Both Myc and Max proteins are homologs proteins for USF1, which bind E-box sequence and belong to b/HLH/Zip transcription factor family. The structure is shown at fig. 11. It could be that tetramer formation is an artefact of crystal packing due to high protein concentration.

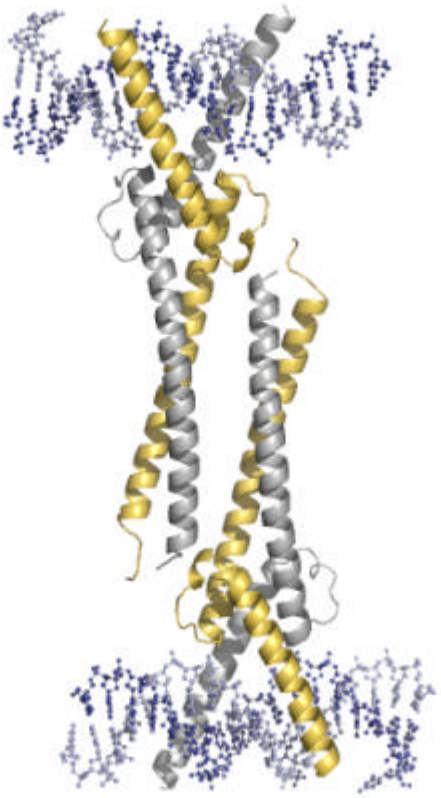


Fig. 11 **Myc-Max tetramer structure**

Myc molecule is colored in grey, Max molecule is colored in yellow

This project was focused on getting structural information of USF1 tetramer bound to HIV1 LTR. The tetramer formation could lead to DNA looping which could explain the regulation of HIV1 transcription by transcription factors binding to the distal enhancer element. There is no structural information of USF1 tetramer bound to DNA. Since, the USF1 tetrmar bound to DNA could be potentially difficult to crystallize, the method of choice was small angle X-ray scattering (SAXS). However, this method requires high sample concentration and USF1 tetramerization could be due to high concentration. In order to check USF1 tetramerization under physiological concentrations fluorescence resonance energy transfer (FRET) experiment was planned. In order to visualize DNA loop, rotary shadowing electron microscopy (EM) was planned.

Aim of the project:

The aim of this project was to study USF1 homotetramer formation upon DNA binding. In more details, this project involves structural investigations on USF1 tetramer formation and DNA loop formation which was proposed (Ferre D'Amare et. al., 1994) but was not shown. Three different methods were applied: small-angle X-ray scattering (SAXS) to get low-resolution structural information on USF1 tetramer, fluorescence resonance energy transfer (FRET) to validate SAXS model of USF1 tetramer and rotary shadowing electron microscopy to visualize DNA loop.

2. Results and discussions

2.1 USF1 expression and purification

USF1 (residues 194-310) corresponding to the bHLHZ region was provided by Francisco Fernandez (former PhD student at EMBL-Hamburg). In this construct both cysteines were mutated to serines in order to avoid oxidation.

The expression and purification protocol provided by Fransisco Fernandez was modified at several steps and can be found in "Materials and Methods" chapter.

2.2 Ets1 expression and purification

Three constructs for Ets1 (residues 280-441, 301-441 and 335-441) were provided by Francisco Fernandez (former PhD student at EMBL-Hamburg) on the plasmid pProExHTb. In these constructs both cysteines were mutated to serines in order to avoid oxidation.

The soluble protein could be expressed from the Ets1 construct containing residues 301-441 and 280-441 and purified by affinity chromatography followed by TEV-protease cleavage. Up to 80% of the protein precipitated after TEV cleavage. In order to overcome this problem different buffer conditions were tested, however, none of them gave satisfactory results.

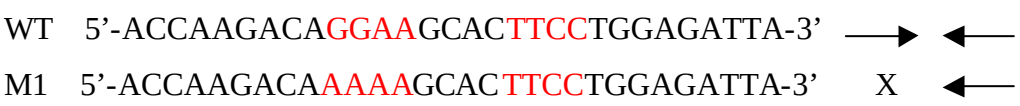
In order to improve solubility the constructs (residues 335-441, 301-441 and 280-441) were subcloned into pETM10 expression vector which has a noncleavable N-terminal His-tag (described in "Material and Methods" chapter).

Ets1 protein from the shortest construct corresponding to residues 335-441 was found in insoluble fraction, although different buffers were tested (see "expression and solubility test" in "Materials and methods" chapter). The Ets1 proteins (residues 280-441, 301-441) were found to be expressed in soluble fraction (10-15 mg protein for 1 liter of bacterial culture).

Ets1 proteins (residues 301-441 and 280-441) were expressed and purified as described in "Material and methods" chapter.

2.3 Ets1/Ets1/DNA complex formation

For the DNA binding experiment a dsDNA fragment was used:



The core binding sequences for Ets1 is colored in red, orientation of binding sites is shown by arrows. The Ets1/Ets1/DNA complex was formed with a 2:1 molar ratio (protein to DNA). The complex formation was demonstrated by native gel. As an additional control (in order to prove that 2 molecules of Ets1 binds to DNA) the complex was formed between Ets1 and DNA (M1), where one Ets1 binding site was mutated (Baillat et. al., 2002). This experiment was a repeat of the experiment done by Baiilat et. al., 2002, and the results were the same (data not shown).

In order to further examine Ets1/Ets1/DNA complex formation, dsDNA with different lengths (from 15 bp to 21 bp) were tested to establish the minimum length of DNA required for complex formation. This was done by "band shift assay". For this purpose two native gels were made and one of them was stained with coomassie blue and another with ethidium bromide in order to visualize protein and DNA respectively. The results are illustrated in fig. 48.

One can see that the complexes with higher molecular weight migrate faster on native gel than those with smaller molecular weight. It is known that migration on a native gel depends on the molecule shape and on the charge. When the DNA is getting longer the complex has a higher negative charge. This could explain that the complexes with higher molecular weight migrate faster. Another explanation could be that when longer pieces of DNA are used the complex is more elongated, and could migrate faster.

As shown on fig. 12, the complex is not formed when 15 bp dsDNA is used. Relatively weak binding can be observed when 17 bp dsDNA is used. For 19 bp, 21bp and 23 bp dsDNA one can see strong bands corresponding Ets1/Ets1/DNA complex.

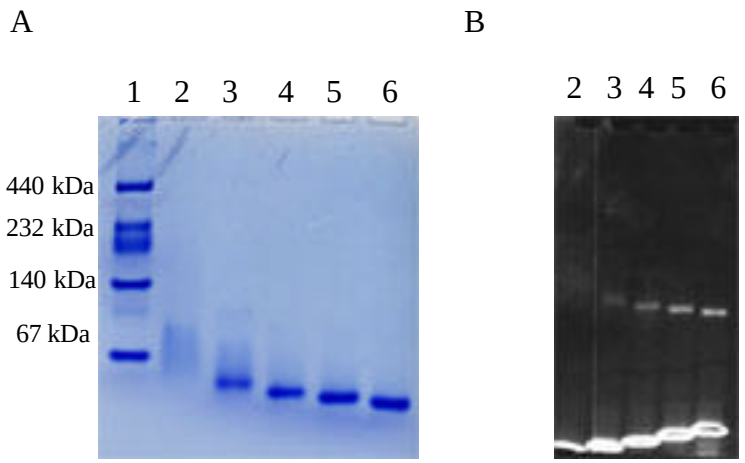


Fig. 12 **Native gel of Ets1/Ets1/DNA complexes formed with dsDNA having different length**

A. Gel stained with Coomassie blue

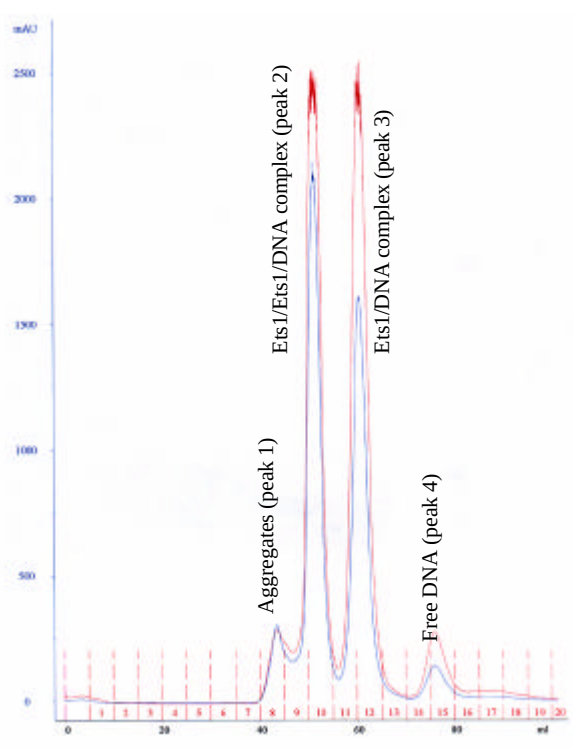
B. Gel stained with Ethidium Bromide

Lane 1 – molecular weight marker for native electrophoresis, lane 2 – Ets1 (280-441) + 15 bp DNA fragment, lane 3 - Ets1 (280-441) + 17 bp DNA fragment, lane 4 - Ets1 (280-441) + 19 bp DNA fragment, lane 5 - Ets1 (280-441) + 21 bp DNA fragment, lane 6 - Ets1 (280-441) + 23 bp DNA fragment

2.4 Ets1/Ets1/DNA complex purification

In order to purify Ets1/Ets1/DNA complex (Ets1 280-441) from Ets1/DNA complex and free DNA, gel filtration was used. Initially, after Ni-NTA column, gel filtration was used for the Ets1 protein, then the Ets1/Ets1/DNA complex was formed and purified by gel filtration. Subsequently, the gel filtration purification step for Ets1 protein was shown to be unnecessary. The results are illustrated on fig. 13.

A



B

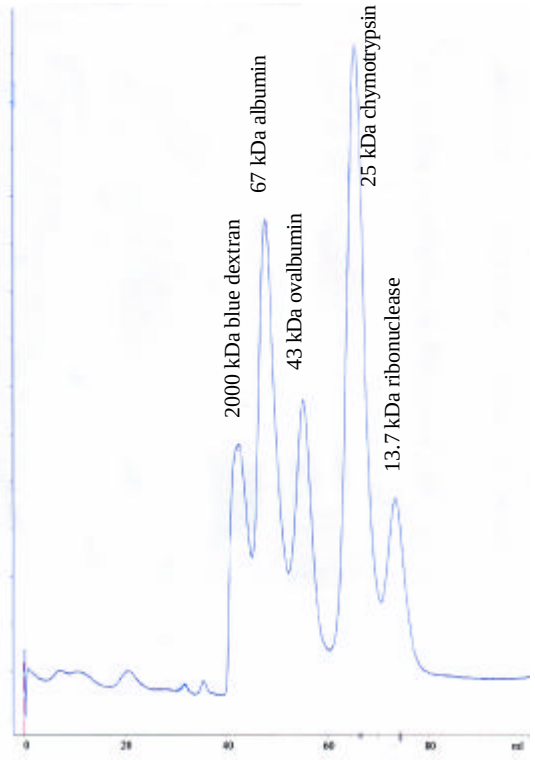


Fig. 13 **Gel filtration of Ets1/Ets1/DNA complex**

A. Gel filtration of Ets1/Ets1/DNA complex

B. Low molecular weight gel filtration standard (Amersham)

After gel filtration the samples corresponding peak 1, peak 2 and peak 3 and peak 4 were analyzed by "band shift assay". Fig. 14 demonstrates that peak 2 corresponds to Ets1/Ets1/DNA complex.

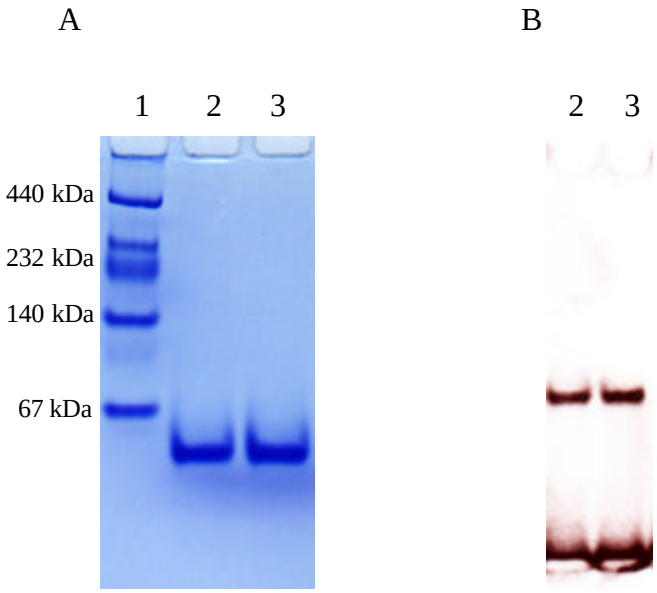


Fig. 14 **Native gel of Ets1/Ets1/DNA complex before and after gel filtration**

A. Gel stained with Coomassie blue

B. Gel stained with Ethidium Bromide

Lane 1 – molecular weight marker for native electrophoresis, lane 2 – sample before gel filtration, lane 3 – sample after gel filtration (peak 2)

2.5 Ets1/Ets1/DNA SAXS experiment

After purification by gel filtration, samples of Ets1/Ets1/ WT DNA (wild type), Ets1/M1 DNA (one of Ets1 binding sites was mutated) complexes and DNA alone (as a negative control) were concentrated and used for SAXS measurements in the concentration range of 1 mg/ml to 10 mg/ml. Ets1 corresponding to the residues 280-441 and 30 bp DNA were used. Ets1 binding site is shown in red.

WT accaagacaGGAAgcacTTCCtggagatta
tgg tt ctgt CCTT cgtgAAGGacctctaat

M1 accaagacaAAAAgcacTTCCtggagatta
tgg tt ctgt TTTT cgtg AAGGacctctaat

Higher concentrations could not be measured due to aggregation. The samples were monodisperse. The experimental curves are shown on fig. 15.

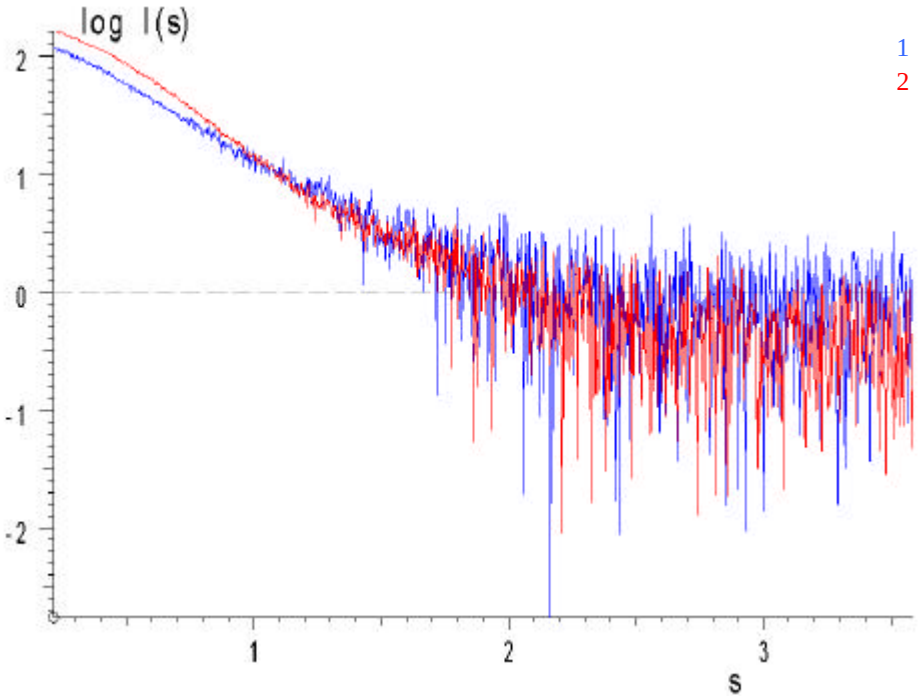


Fig. 15 **Experimental SAXS data of Ets1/Ets1/DNA complex and Ets1/M1 DNA complex at the concentration 5 mg/ml**

curve 1 – experimental SAXS data of Ets1/M1 DNA complex

curve 2 – experimental SAXS data of Ets1/Ets1/DNA complex

This graphic demonstrates dependence of scattering intensity logarithm on absolute value of scattering vector.

The shapes of the curves are different, which corresponds to the different shapes of the complexes Ets1/Ets1/DNA and Ets1/M1 DNA. Extrapolated intensities at zero scattering angle, which are proportional to molecular weights of the complexes pointed out to the difference in molecular weight of the particles.

The crystal structure of Ets1 (residues 309-441) bound to DNA as a monomer was received by extraction of Pax5 molecule from the crystal structure of ternary complex Ets1/Pax5/DNA (PDB code 1MDM) and by making DNA shorter. Using program MASSHA (Konarev et. al., 2001), tentative models of Ets1 dimer bound to DNA were built. During modeling process it was taken under consideration that the distance on the DNA between two Ets1-binding sites is 4 bp, which would correspond a turn of 137° (one DNA helical turn 360° is 10.5 bp) if the Ets1 binding sites would be located on the same DNA strand. In the example of the stromelysin-1 promoter they are located on the complementary strands (see above the stromelysin-1 promoter DNA sequence), which would correspond to a turn of 180°+137°=317°. Thus, the angle between two Ets1 molecules could be 43°. In the calculation the DNA was considered to be unbent. Modeling process is illustrated on fig. 16.

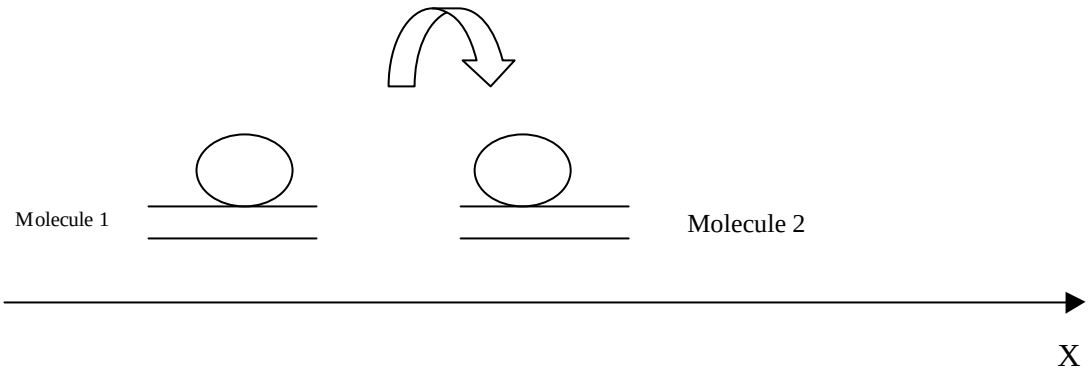


Fig. 16 **Schematic representation of modeling process**
Molecule 2 was rotated along the axis X and therefore different models of Ets1/Ets1/DNA complex were created.

10 different models (with different angles between Ets1 molecules) of the Ets1 dimer on the DNA were built. The theoretical scattering intensities of the models were calculated using program CRYSOL (Svergun et. al., 1995) and compared to the experimental one. Based on the fit the best tentative model was chosen and

investigated further. The best tentative model and the fit to the experimental data are shown in fig. 17.

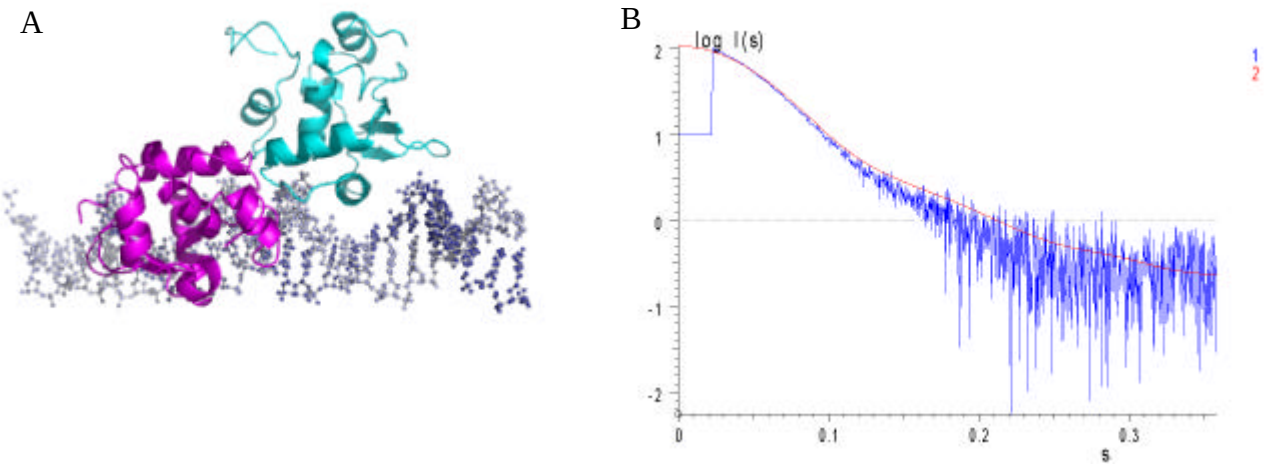


Fig. 17 **The best tentative model of Ets1/Ets1/DNA complex and the fit to the experimental data**

A. The best tentative model of Ets1/Ets1/DNA complex

B. Fit of the experimental data

Curve 1 – experimental data from Ets1/Ets1/DNA complex

Curve 2 – theoretical curve of the tentative model calculated by the program CRYSOL

This graphic demonstrates dependence of scattering intensity logarithm on absolute value of scattering vector.

The construct for Ets1 used in SAXS experiment corresponded to amino acid residues 280-441 and had a N-terminal His-tag. The crystal structure of Ets1/DNA (PDB code 1MDM) used to build tentative model of Ets1/Ets1/DNA complex showed no electron density for residues 280-308. This could explain the difference in the beginning of the theoretical curve calculated from the model and the experimental curve. In order to a build more accurate model the program BUNCH was used (D. Svergun and M. Petoukhov, to be published). This program automatically build models of macromolecular complexes from high resolution structures or homology models of their subunits or domains against SAXS data. In the case of the Ets1/Ets1/DNA complex the missing N-terminal residues were added. 20 different

models were built by the program BUNCH (Petoukhov et. al., 2005). The overlay of two representative models and the fit are illustrated in fig. 18.

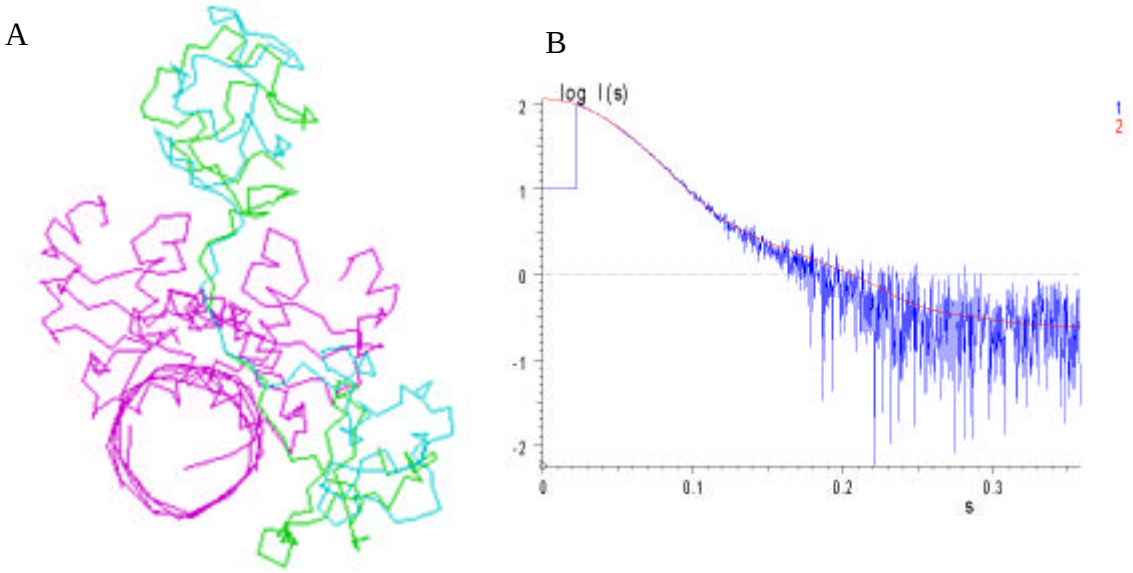


Fig. 18 **The overlay of two representative models built by BUNCH and the fit to the experimental data**

A. The overlay of two representative models built by BUNCH. The tentative model of Ets1/Ets1/DNA complex is colored in magenta. The N-terminal parts of Ets1 built by BUNCH are colored in green and cyan.

B. Theoretical curve calculated from one of the models built by BUNCH and the experimental curve

Curve 1 – experimental data from Ets1/Ets1/DNA complex

Curve 2 – theoretical curve of the tentative model calculated by the program CRYSOL

This graphic demonstrates dependence of scattering intensity logarithm on absolute value of scattering vector.

From fig. 18 it can be seen that the N-terminal parts from different models look different, which could be due to flexibility of these regions.

Based on the model built by BUNCH for Ets1/Ets1/DNA complex, the model of Ets1/M1 DNA was built by subtraction of one Ets1 molecule. The model of Ets1/M1 DNA and the fit are illustrated in fig. 19.

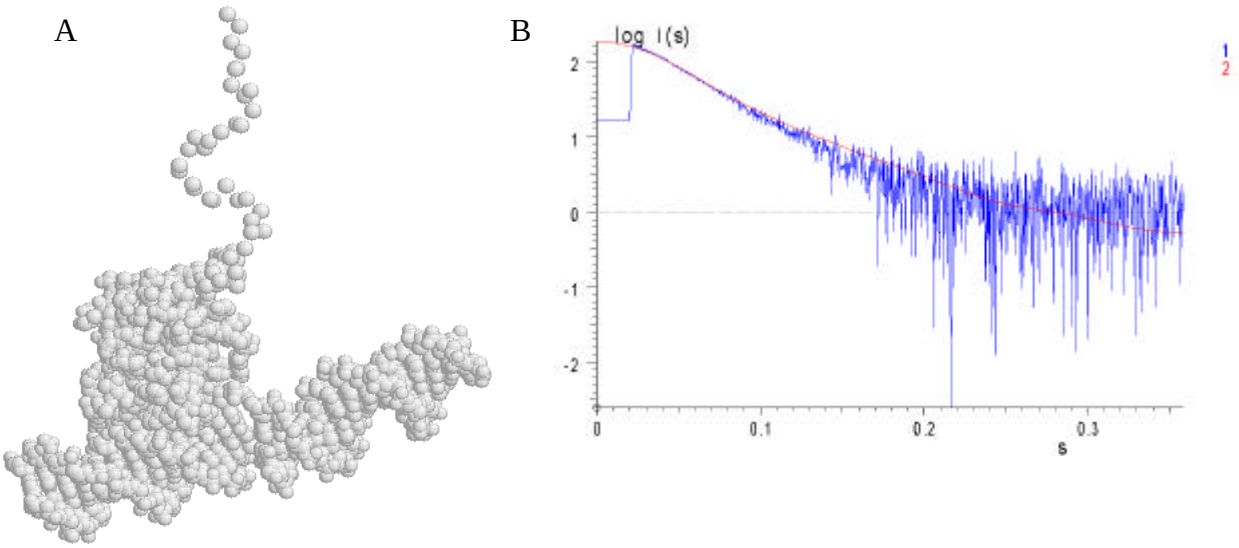


Fig. 19 **Model of Ets1/M1 DNA and the fit to the experimental data**

A. Model of Ets1/M1 DNA

B. The fit of Ets1/M1 DNA model to the experimental curve

Curve 1 – experimental data from Ets1/M1 DNA complex

Curve 2 – theoretical curve of the tentative model calculated by the program CRYSOL

This graphic demonstrates dependence of scattering intensity logarithm on absolute value of scattering vector.

2.6 Ets1/Ets1/DNA complex crystallization

Ets1/Ets1/DNA complexes (Ets1 280-441) were formed using different DNA length with and without a TA-overhang, purified by gel-filtration, analyzed on native and SDS gels, concentrated up to 10 mg/ml and used for crystallization.

DNA contacts are very important for crystal packing, so a number of DNA length were used. TA-overhang can help to establish DNA contacts.

DNA used for crystallization is shown below. Ets1 binding sites are colored in red and TA-overhangs are colored in green.

21 TA TgacaGGAAgcacTTCCtgga
ctgt CCTT cgtg AAGGacctA

22 TA TagacaGGAAgcacTTCCtgga
tctgt CCTT cgtgAAGGacctA

23 TA TagacaGGAAgcacTTCCtggag
tctgt CCTT cgtgAAGG acctcA

25 TA TaagacaGGAAgcacTTCCtggaga
ttctgt CCTT cgtgAAGGacc tctA

26 TA TcaagacaGGAAgcacTTCCtggaga
gttctgt CCTT cgtg AAGGacc tctA

27 TA TcaagacaGGAAgcacTTCCtggagat
gttctgt CCTT cgtg AAGGacc tctaA

28 TA TccaagacaGGAAgcacTTCCtggagat
ggttctgt CCTT cgtg AAGGacctc taA

30 accaagacaGGAAgcacTTCCtggagatta
tgg tt ctgt CCTT cgtgAAGGacctctaat

Initially, 30bp DNA (Baillat et al., 2002) was used for crystallization and no crystals were obtained. Then 26 bp DNA with TA-overhang was used. This DNA was chosen based on the theory that protein-DNA complexes are better crystallized when DNA length is equal to 1, 1.5, 2, 2.5 (Batchelor et. al. 2002) and etc helical turns on the DNA. A length of 26 bp corresponds to 2.5 helical turns. Crystals were grown by sitting drops vapor diffusion method. Temperatures 20º C and 4º C were tested in

crystallization trials. The commercial crystallization screens were used. Crystals appeared under 5 different conditions at 20º C after 1-2 months. Crystals pictures are shown on fig. 20.

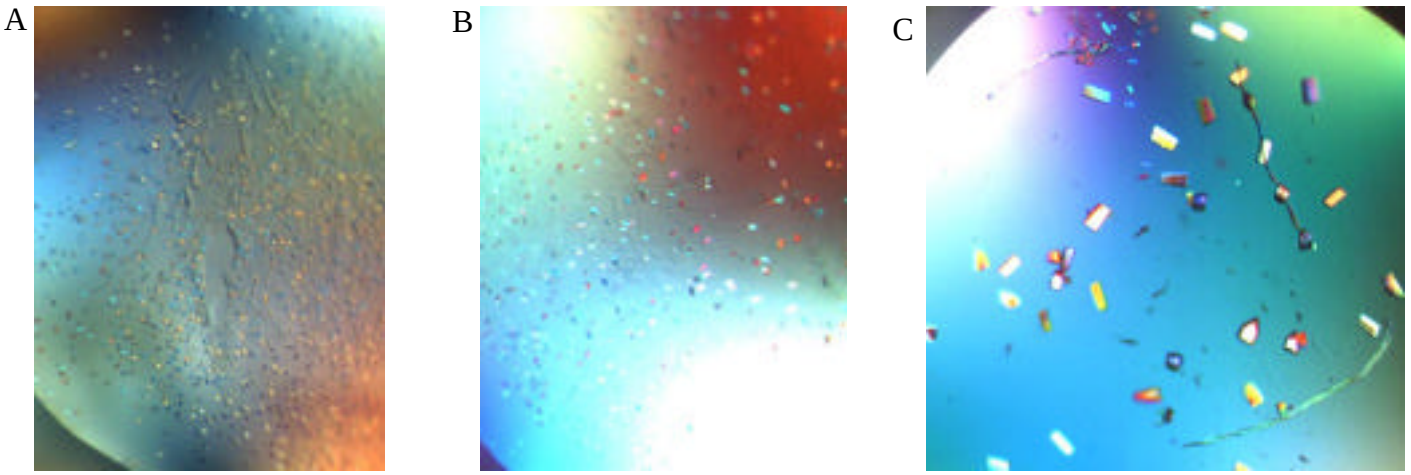


Fig. 20 **Picture of the crystals obtained for the complex of Ets1 dimer and 26 bp DNA with TA-overhang**

A, B, C – three different conditions where the crystals were formed.

After optimization these crystals diffracted to a resolution of 5-6 Å. For one of the crystals (conditions from fig. 56 C) a full data set was collected (fig. 21) in order to try to get some structural information and to validate SAXS model. However, the data quality was poor.

When PEG 4000 (for the conditions shown on fig. 20 A) was changed to PEG 1500 small needles were grown. They are shown on fig. 22. However, the attempts to improve these crystals failed.

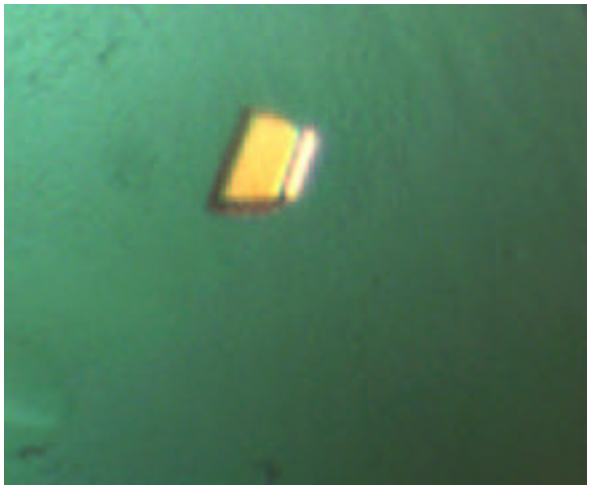
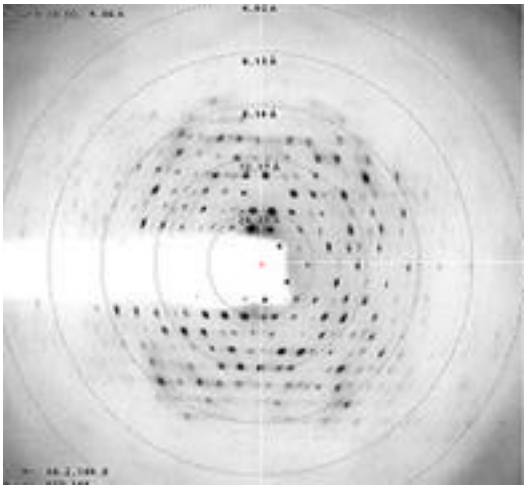
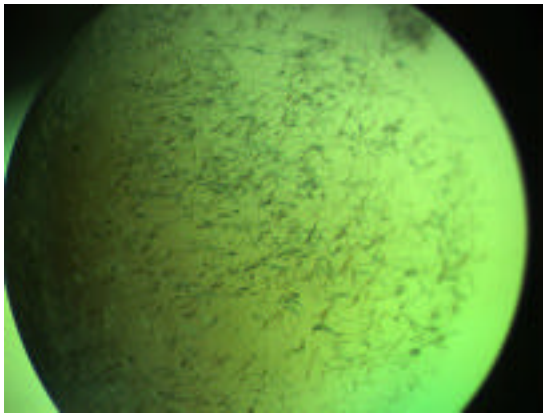
A  B 

Fig. 21 **Picture of the crystal of the complex of Ets1 dimer and 26 bp DNA with TA-overhang formed under the same conditions as shown on fig. 20 C and its diffraction pattern**

A. Crystal picture

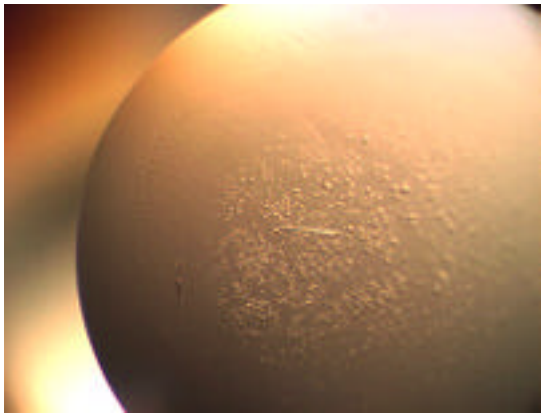
B. Diffraction pattern of the crystal.



0.2 M $CaCl_2$
0.1 M HEPES-NaOH pH 7
20% PEG 1500

Fig. 22 **Picture of the crystals of the complex of Ets1 dimer and 26 bp DNA with TA-overhang grown with PEG 1500**

A



0.2 M $(NH_4)H_2PO_4$
25% PEG 8000
0.1 M MES 6.4

B

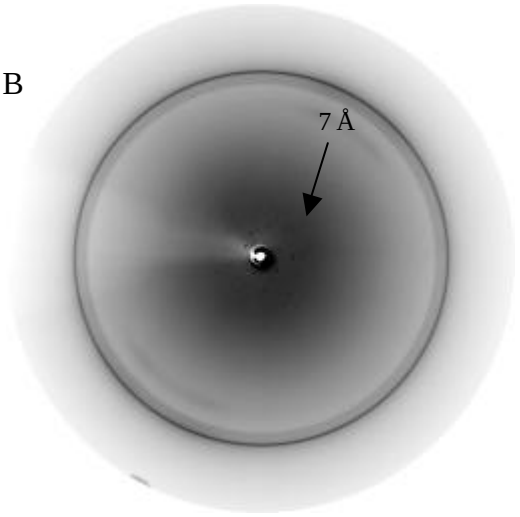


Fig. 23 **Picture of the crystals of the complex of Ets1 dimer and 26 bp DNA with TA-overhang after optimization and their diffraction pattern**

A. Crystal picture

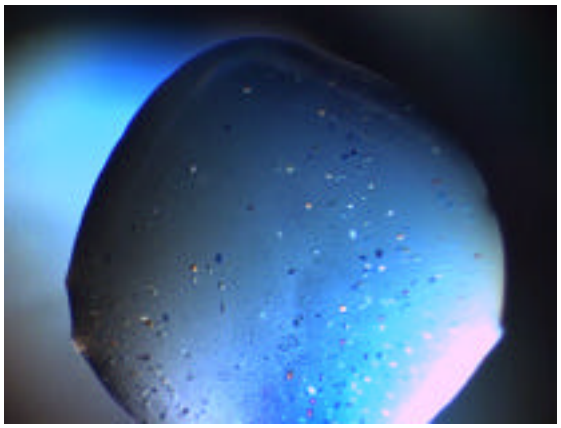
B. Diffraction pattern of these crystals.

When the salt was changed from calcium acetate to ammonium phosphate long thin needles were obtained. They and their diffraction pattern are shown on fig. 23.

In spite of the fact that different crystal forms were grown under different conditions, optimization did not improve the resolution of the crystals. It was decided to work with a different DNA construct. The work was continued in two different directions. The first direction was to vary the DNA length around 26 bp. Therefore, the following DNA lengths were tried: 25 bp with TA-overhang, 27 bp with TA-overhang and 28 bp with TA-overhang. The second approach was to use as short DNA piece for crystallization as possible in order to avoid the flexibility given by DNA ends which are not involved in binding. Based on the results of binding assay (see above, fig. 12), the following DNA lengths were chosen: 21 bp with TA-overhang, 22 bp with TA-overhang and 23 bp with TA-overhang and used for crystallization.

The DNA lengths 25 bp with TA-overhang and 28 bp with TA-overhang did not give any crystals. But the crystals were grown with the DNA length 27 bp with TA-overhang under 3 different conditions. All the conditions were similar to the previous one when the DNA length 26 bp with TA-overhang was used.

A



0.2 M Ca Acetate
0.1 M MES pH 6.5
16% PEG 8000

B

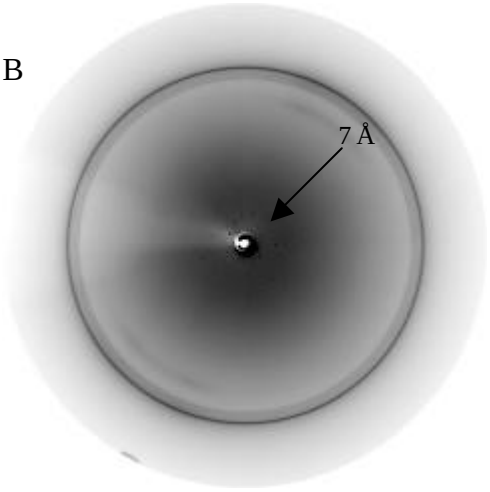


Fig. 24 **Picture of the crystals of the complex of Ets1 dimer and 27 bp DNA with TA-overhang after optimization and their diffraction pattern**

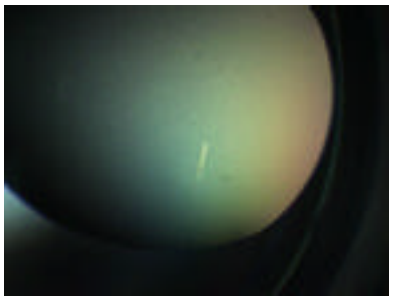
A. Crystal picture

B. Diffraction pattern of these crystals.

Fig. 24 shows how one type of the crystals grown for the complex of Ets1 dimer and 27 bp DNA with TA-overhang after 2-4 weeks at 20º C. They diffracted to a resolution of 7 Å. In comparison to the crystals grown with 26 bp DNA with TA-overhang the spots on the diffraction pattern do not looked diffused. But the resolution was similar to previously observed for the DNA length 26 bp with TA-overhang .

On the fig. 23 and 24 ice rings can be seen. In order to check that low diffraction was not due to the freezing, crystals were mounted in the capillary or used with a cryo solution, but it did not improve the resolution.

Another type of crystals was obtained after 2-4 weeks at 20º C for the DNA length 27 bp with TA-overhang. The crystal picture is illustrated on fig. 25.

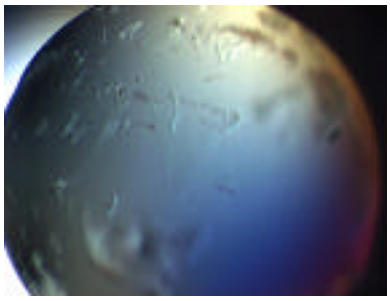


0.15 M $(NH_4)_2SO_4$
0.1 M Na Citrate pH 6.1
30% PEG 8000

Fig. 25 **Picture of another crystal type of the complex of Ets1 dimer and 27 bp DNA with TA-overhang after optimization**

The crystal illustrated on fig. 25 diffracted up to 3.7 Å. The full dataset was collected for this crystal. Initial trials to work with this data set were made but then better crystals were obtained.

For shorter pieces of DNA the crystals could be obtained for the complex of Ets1 dimer and 22 bp, 23 bp DNA with TA-overhang, no crystals appeared for the complex of Ets1 dimer and 21 bp DNA with TA-overhang. For the complex of Ets1 dimer and 22 bp DNA with TA-overhang the crystals (needles) were grown at 20º C after 5-10 days. The crystals are shown on fig. 26.

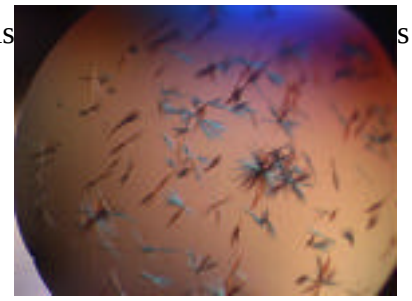


0.2 M KCl
0.05 M MES 5.6
5% PEG 8000
0.01 M $MgCl_2$

Fig. 26 **Picture of the crystals of the complex of Ets1 dimer and 22 bp DNA with TA-overhang after optimization**

Change of PEG molecular weight, percentage of PEG, buffer did not make the crystals bigger. But change of salt from KCl to $CaCl_2$ enlarged the size of the crystals. $CaCl_2$ was chosen based on previous experience with 26 bp DNA with TA-overhang

when the crystals were obtained under conditions containing 200 mM $CaCl_2$. The crystals are of thin needles. The crystals are shown on fig. 27.



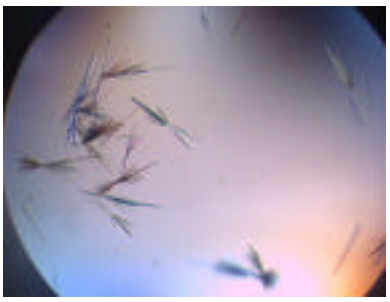
0.2 M $CaCl_2$
MES 5.6
5% PEG 8000
0.01 M $MgCl_2$

Fig. 27 **Picture of the crystals of the complex of Ets1 dimer and 22 bp DNA with TA-overhang when KCl was changed to $CaCl_2$**

These crystals diffracted up to 5-6 Å. But they were very difficult to manipulate because they were very thin and easy to break.

After increasing the concentration of additive ($MgCl_2$) to 90 mM the crystals become bigger and the best of them diffracted up to 3.5 Å. But the intensities of the spots were very weak and data collection was not possible because of this reason. The crystals and their diffraction pattern are shown on fig. 28.

A



0.2 M $CaCl_2$
0.05 M MES 5.6
5% PEG 8000
0.09 M $MgCl_2$

B

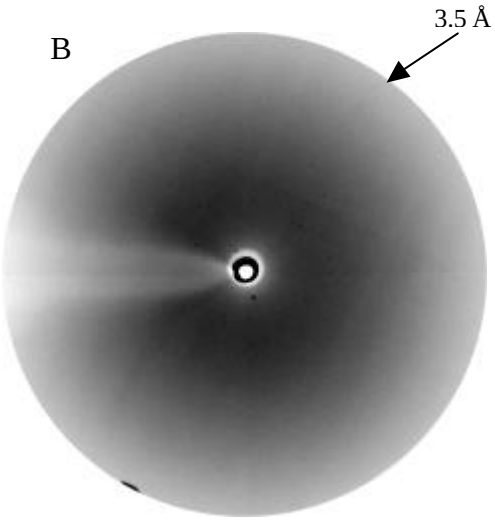


Fig. 28 **Picture of the crystals of the complex of Ets1 dimer and 22 bp DNA with TA-overhang grown in presence of 90 mM $MgCl_2$ and their diffraction pattern**

A. Crystal picture

B. Diffraction pattern of these crystals.

In order to make crystal manipulation easier the crystals were cross-linked using 0.25% of glutaraldehyde for 5 hours. After cross-linking the crystals became very stable but they lost their diffraction properties. The best diffraction observed for cross-linked crystals was 6 Å. Annealing was tried on these crystals as well but it just dropped the resolution from 4 to 6-7 Å. All further attempts to improve the crystals quality did not lead anywhere.

In order to make sure that the crystals corresponds to the Ets1/Ets1/DNA complex, some crystals were dissolved and analyzed on native gel. It was done for the crystals grown with 22 bp with TA-overhang and 27 bp with TA-overhang DNA. The results are shown on fig. 29.

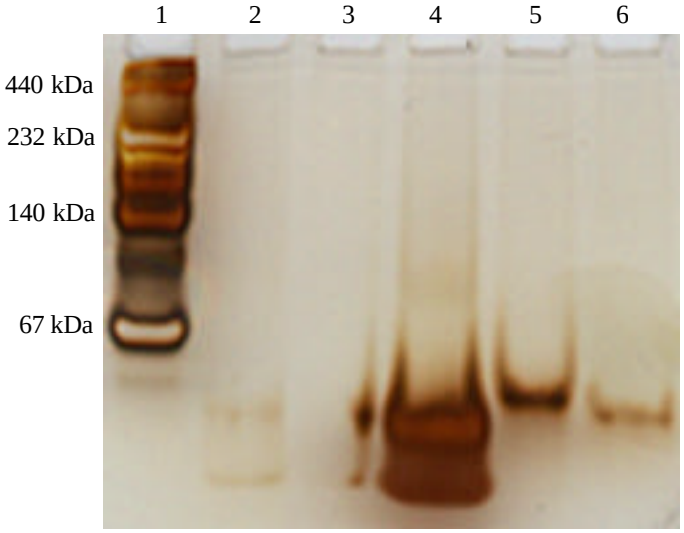


Fig. 29 **Native gel of dissolved crystals when 22 bp and 27 bp DNA with TA-overhang were used (silver staining)**

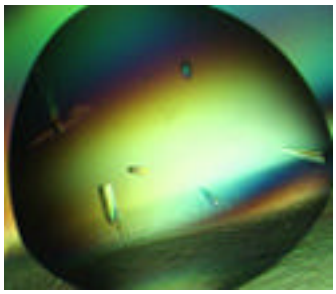
Lane 1 – molecular weight marker for native electrophoresis, lane 2 , 3 – Ets1/Ets1/27TA DNA - washing of the crystals, lane 4 - Ets1/Ets1/27TA DNA – dissolved crystals, lane 5 - Ets1/Ets1/22TA DNA – 1st wash and partially dissolved crystals, lane 6 - Ets1/Ets1/22TA DNA – dissolved crystals

On lane 4 one can see dissolved crystals of Ets1/Ets1/27 bp with TA-overhang DNA complex, the complex was partially degraded. Lane 5 corresponds to the first wash of the crystals of Ets1 dimer on 22 bp DNA but at this stage the crystals were partially dissolved. These crystals were small and thin needles making bundles and they are difficult to manipulate. As a result many of the crystals were left in the

washing solution and dissolved and only some of them were transferred to a fresh drop with the buffer solution used for protein purification and dissolved there. The manner in which the crystals were treated explains why the band on the lane 5 is thicker than on the lane 6.

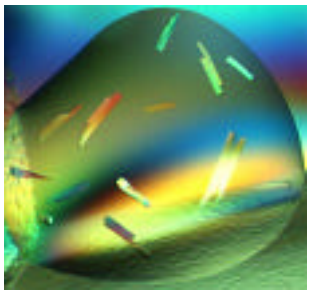
After all the trials to work with the needles produced from the Ets1 dimer and 22 bp DNA with TA-overhang complex which did not give satisfactory results it was decided to try 23 bp DNA with TA-overhang for crystallization. The initial crystal screens were done at 20º C in EMBL high thought put crystallization facility in sitting drops. Crystals from 5 different conditions were obtained. The crystal pictures and crystallization conditions are shown on fig. 30.

A



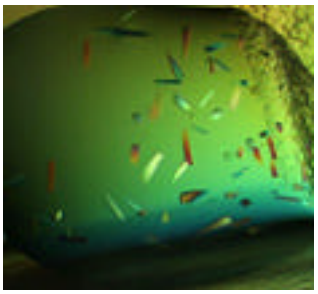
0.2 M ammonium acetate
0.1 M sodium citrate 5.6
30% PEG 4000

B



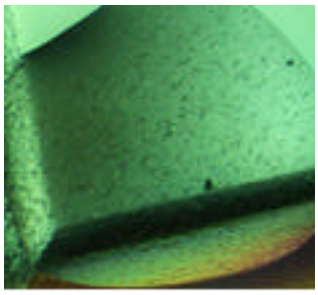
0.2 M ammonium
0.1 M sodium cacc
30% PEG 8000

C



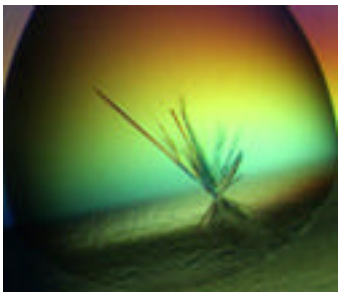
0.2 M ammonium sulfate
0.1 M MES 6.5
30% PEG MME 5000

D



0.1 M $CaCl_2$
0.1 M MES 6.5
15% PEG 4000

E



0.1 M Li_2SO_4
0.1 M MES 6.0
20% MPEG 2000

Fig. 30 **Crystal pictures and crystallization conditions for the complex of Ets1 dimer and 23bp DNA with TA-overhang**

A-E Crystal pictures

The crystals presented on fig. 30 A diffracted to a resolution of 3.1 Å and the full dataset was collected. The resolution obtained for the other conditions is summarized in table 1. Needles were grown for the complex of Ets1 dimer and 23 bp DNA with TA-overhang under similar conditions as for the complex of Ets1 dimer and 22 bp DNA with TA-overhang.

In order to prove that the crystals grown under conditions illustrated on fig. 30A indeed are the crystals of Ets1 dimer and 23 bp DNA with TA-overhang complex, the crystals were dissolved and analyzed on native gel (fig. 31). One can see that dissolved crystals are the crystals of Ets1 dimer and 23 bp DNA with TA-overhang complex.

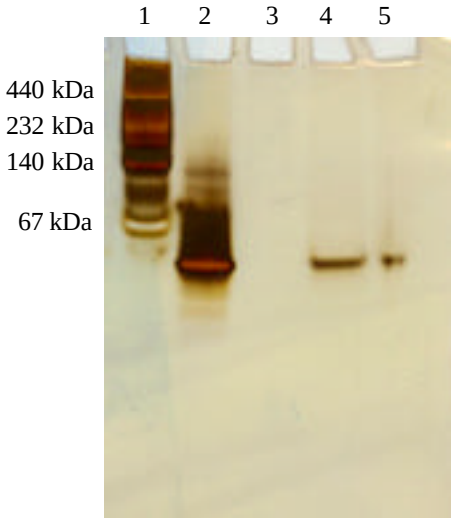


Fig. 31 **Native gel of dissolved crystals of Ets1 dimer and 23 bp DNA with TA-overhang complex (silver staining)**

Lane 1 - molecular weight marker for native electrophoresis, lane 2 – Ets1/Ets1/23 bp DNA with TA-overhang (positive control), lane 4 – dissolved crystals, lane 5 – crystal wash

Table 1 – **summary of crystallization conditions**

DNA length	crystals	conditions				resolution
		salt	buffer	precipitant	additive	
21 TA;pl	No	-				-
22 TA	Very small needles	0.2 M KCl	0.05 M MES 5.6	5% PEG 8000	0.01 M $MgCl_2$	to small for the beam-line
	needles	0.2 M $CaCl_2$	0.05 M MES 5.6	5% PEG 8000	0.01 M $MgCl_2$	Up to 5-6 Å
	needles	0.2 M $CaCl_2$	0.05 M MES 5.6	5% PEG 8000	0.09 M $MgCl_2$	3.5 Å very weak intensities
23 TA	small needles Sigma/PEG A3	0.1 M $CaCl_2$	0.1 M MES 6.5	15% PEG 4000	-	to small for the beam-line
	needles Sigma/PEG H1	0.1 M Li_2SO_4	0.1 M MES 6.0	20% MPEG 2000	-	not tried yet
	Nextal F9	0.2 M ammonium sulfate	0.1 M sodium cacodylate 6.5	30% PEG 8000	-	7 Å
	Nextal G11	0.2 M ammonium acetate	0.1 M tri-sodium citrate 5.6	30% PEG 4000	-	7 Å
	PEG/Ion A5	0.2 M sodium sulfate	-	20% PEG 3350	-	6 Å
	Hampton G2	0.2 M ammonium sulfate	0.1 M sodium cacodylate 6.5	30% PEG 8000	-	4.1 Å
	Hampton B10	0.2 M ammonium sulfate	0.1 M MES 6.5	30% PEG MME 5000	-	5 Å
	Hampton A2	0.2 M ammonium acetate	0.1 M sodium citrate 5.6	30% PEG 4000		3.1 Å
	optimization	0.2 M ammonium acetate	0.1 M sodium citrate 5.6	28% PEG 2000		2.58 Å
25 TA	No	-				-
26 TA		0.2 M $CaCl_2$	0.1 M HEPES-NaOH 7.0	25% PEG 4000	-	6 Å
		0.2M Mg Acetate	0.1 M MES 6.4	20% PEG 8000	-	6 Å
		0.2 M Ca Acetate	0.1 M MES 6.5	18% PEG 8000	-	5.5 Å
	optimization	0.2 M $(NH_4)H_2PO_4$	0.1 M MES 6.4	25% PEG 8000	-	7 Å
	Very small needles	0.2 M $CaCl_2$	0.1 M HEPES-NaOH 7.0	20% PEG 1500	-	to small for the beam-line
27 TA		0.2 M Ca Acetate	0.1 M MES 6.5	16% PEG 8000	-	7 Å
		0.2 M Ca Acetate	0.1 M MES 6.5	16% PEG 4000	-	7 Å
		0.15 M $(NH_4)_2SO_4$	0.1 M sodium citrate 6.1	30% PEG 8000	-	3.7 Å
28 TA	-	-				-
30	-	-				-

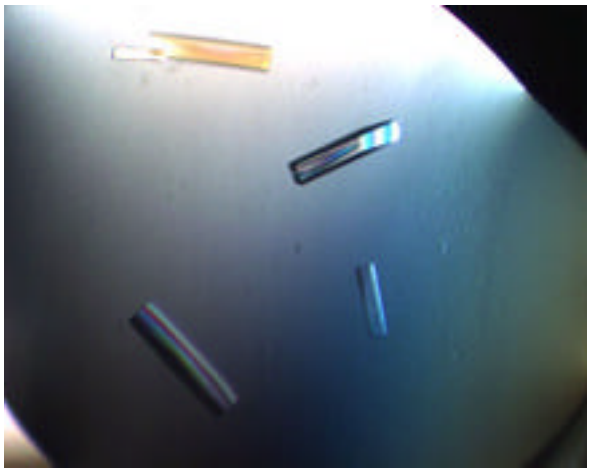
Conclusions on the crystallization conditions of Ets1 dimer and DNA complex:

1. All the crystals were obtained in PEG in pH range from 5.6 to 6.5.
2. The crystals usually required 200 mM salt which could be decreased to 150 mM or even 100 mM. As cations Ca, Mg, Li and ammonim were used as anions acetate, sulfate, chloride or phosphate were used. Sometimes $MgCl_2$ was used as an additive.
3. All the crystals were grown at 20º C. Decrease of the temperature to 4º C leads to formation of very small crystals or even crystalline precipitate. Increase of the temperature to 25º C gave smaller crystals that obtained at 20º C and decrease their diffraction properties.
4. Annealing of the crystals did not help to improve diffraction properties for any type of crystals.
5. DNA used for crystallization is an important parameter. For instance, only DNA length 22 bp with TA-overhang , 23 bp with TA-overhang, 26 bp with TA-overhang and 27 bp with TA-overhang leads to the formation of the crystals when the DNA length 21 bp with TA-overhang, 25 bp with TA-overhang, 28 bp with TA-overhang and 30 does not. However, the DNA length in terms of helical turns is not the only parameter. One helical turn is 10.5 bp that 21 bp corresponds exactly to two helical turns when 23 bp does not but the complex with the DNA 23 bp gives crystals. However, the solvent used, DNA sequence and the way how protein and DNA are interacting, etc are important for the crystallization of protein-DNA complexes. Different conditions and different DNA length have to be tested in order to obtain good quality crystals.

2.7 Ets1/Ets1/DNA structure determination

Crystals of Ets1 dimer complex formed with DNA length 23 bp with TA-overhang grown under the following conditions: 0.2 M ammonium acetate, 0.1 M sodium citrate 5.6, 28% PEG 2000. These crystals did not require cryo-protection. Crystals were mounted in a loop and transferred into a stream of nitrogen at 100 K. A native data set was collected on beam line ID 29 ESRF Grenoble at 0.98 Å wavelength. The crystal picture and their diffraction pattern shown on fig. 32.

A



B

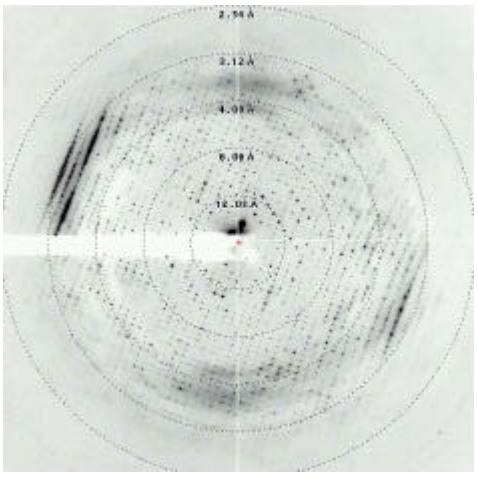


Fig. 32 **Crystals obtained for the complex of Ets1 dimer on 23 bp DNA with TA-overhang and diffraction pattern for one of them**

A. Crystal picture

B. Diffraction pattern of the crystal

The crystals were obtained for the complex formed by two Ets1 molecules (residues 280-441) and 23 bp dsDNA fragment, which had two Ets1 binding sites. The DNA sequence is shown below.

TagacaGGAAgcacTTCCtggag
tctgt CCTT cgtgAAGG acctcA

Ets1 binding sites are shown in red and TA-overhangs are shown in green.

Data set was autoindexed, reduced and scaled with HLK suite of programs (DENZO and SCALEPACK). An overview of data and processing statistics for Ets1/Ets1/DNA complex is given in Table 2.

Table 2. **Data collection statistics for Ets1 dimer bound to 23 bp DNA**

Wavelenght (Å)	0.9756
Resolution range (Å)	20.0-2.58
N_0 of observations	56173
N_0 of reflections	20507
Completeness (%)	95.8 (75.4)
Redundancy	2.7
<I/sI>	20.45 (2.12)
R_{merge}	0.08 (0.66)
Space group	$P2_12_12$
? (Å)	93.556
b (Å)	100.838
c (Å)	69.810

Systematic absences of axial reflection uniquely assign the crystal to the orthorhombic space group $P2_12_12$ with unit cell parameters shown in Table 2. The solvent content amount is estimated as 58.2% and the Matthews coefficient is 3.0 $Å^3$/Da. The crystals contain one complex in the asymmetric unit.

The structure was solved by molecular replacement using MOLREP program. The solutions are presented in Table 3.

As a search model the structure of Ets1 monomer (residues 333-436) bound to DNA fragment with one Ets1 binding site was used. The DNA in the model was cut to 11 bp. The DNA sequence is shown below. Ets1 binding site is shown in red.

gtgccGGAAat
cacggCCTTta

The structure of Ets1 monomer bound to 11 bp DNA fragment with one Ets1 binding site in the text below is called “monomer”.

The complex of Ets1 dimer on 23 bp DNA fragment has non-crystallographic symmetry. In the MOLREP program it was searched for two “monomers” in asymmetric unit (which would correspond to one complex of Ets1 dimer on 23 bp DNA) because it was expected to have one complex according to Matthews coefficient determination (see above).

Table 3. **Solutions for two “monomers” found by MOLREP program**

Solutions	a	ß	?	tx	ty	tz	Rfac	Corr
Solution 1	120.34	-155.55	-170.00	-0.048	0.249	-0.250	0.575	0.437
Solution 2	120.34	155.55	10.00	-0.390	0.254	-0.310	0.551	0.521

Two “monomers” in the asymmetric unit were found by the program MOLREP. However, they belonged to two different complexes. This happened because the program was trying to find the closest position for the protein but not for DNA. In order to check that the solution found by the program MOLREP makes sense contacts between the complexes in the crystal were investigated. It was done in the program O where the complexes were generated according to the symmetry operators. It was found that there was no overlapping between the complexes. The DNA was making layers, which is typical for protein-DNA complexes. When the contact between two “monomers” within one complex occurred the DNA was not disrupted but was making continuous chains of dsDNA from two “monomers”. The real complex is Ets1 dimer on 23 bp DNA. The model of complete complex was constructed in program O. For this purpose, symmetry-related “monomers” were generated, two “monomers” belonging to the same molecule and making continuous DNA were selected and new model (pdb file) was created. This file containing coordinates of Ets1 dimer on 22 bp DNA was used for further studies. Four complexes were found in the unit cell. The position of the complexes in the unit cell is illustrated on fig. 33.

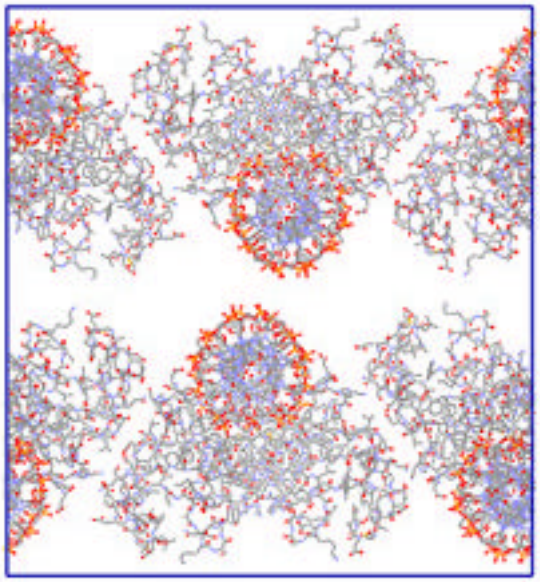


Fig. 33 **Unit cell**

The next step was refinement. First, rigid body refinement was done treating each protein and each DNA chain as a single rigid body. Second, restrained refinement was done. The results are shown in Table 4.

Table 4. **Refinement statistics**

Resolution limits	=	19.964 2.580
Number of used reflections	=	19446
Overall R factor	=	0.3196
Free R factor	=	0.4002
Overall correlation coefficient	=	0.8828
Free correlation coefficient	=	0.8303

R-factor decreased and correlation coefficient increased, but still the value of R-factor is too high. This is due to the fact that in initial model found by MOLREP sixty N-terminal amino acid residues are missing for Ets1 protein and it is necessary to built them.

After initial refinement electron density maps ($2F_0$-F_{calc} and F_0-F_{calc}) were calculated. The model after molecular replacement and refinement and electron density maps are illustrated on fig. 34 and 35.

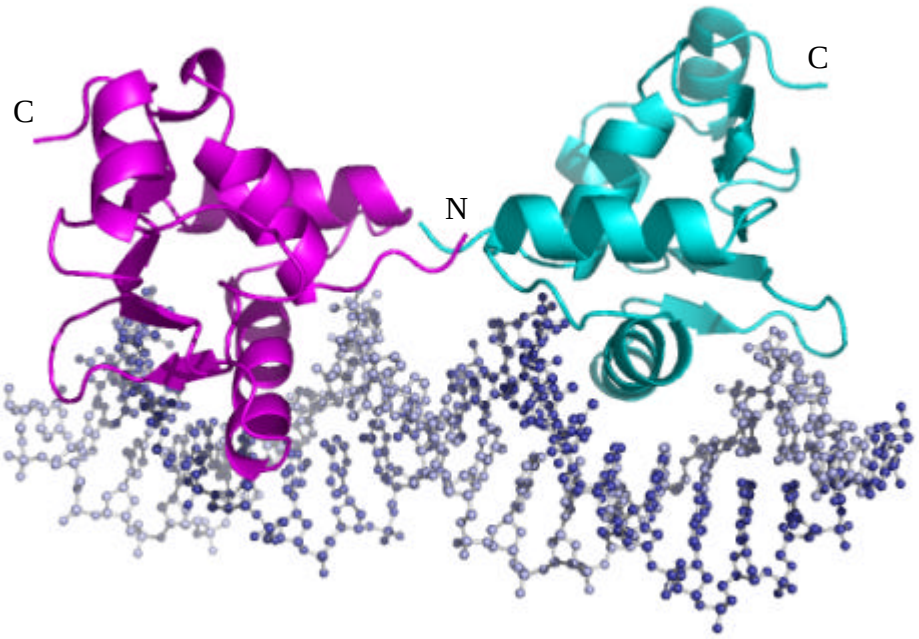
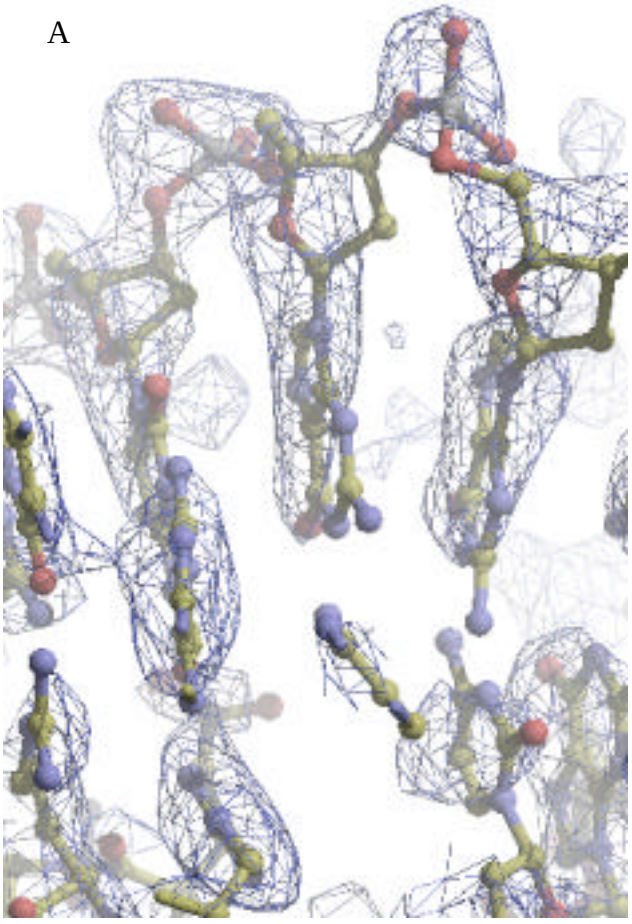


Fig. 34 **The model of the complex of Ets1 dimer on 23 bp DNA**

This work is in progress now. The missing N-terminal part of Ets1 protein has to be manually built. Some bases for DNA have to be modified.

A



B

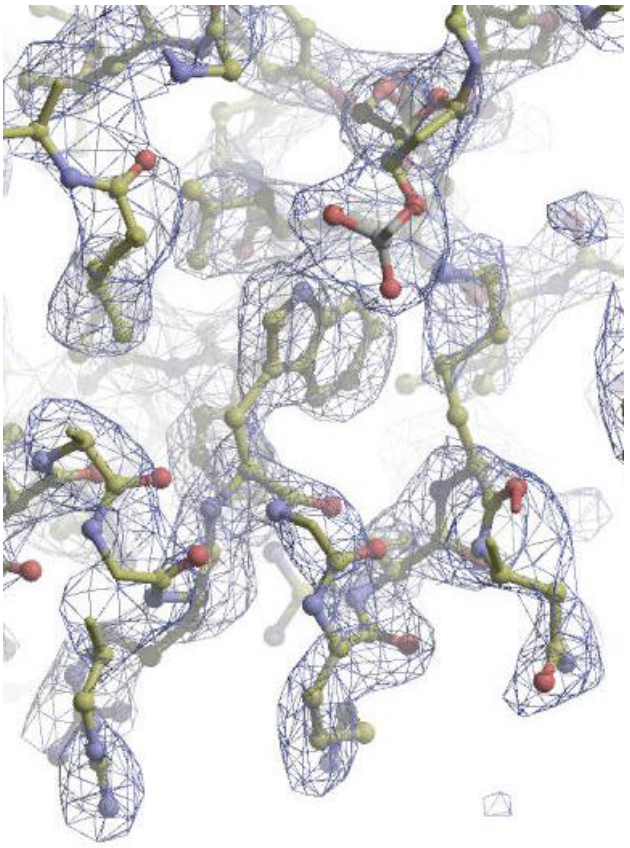


Fig. 35 **Electron density map $2F_0$-F_{calc} with s level 1.5**

A. DNA

B. Ets1 protein

2.8 Comparison of SAXS model and crystallographic model

The models for Ets1 dimer bound to DNA obtained from SAXS and X-ray crystallography were compared by LSQ program. The results are illustrated on fig. 36.

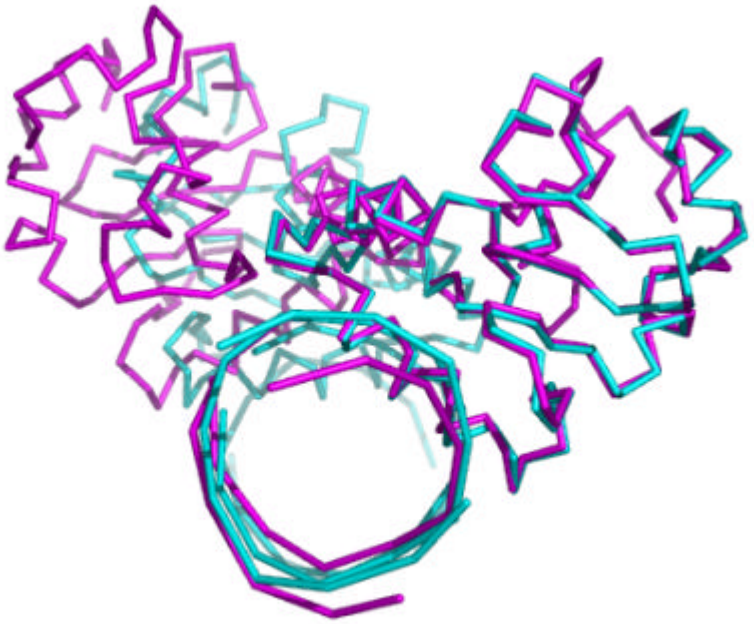


Fig. 36 **Overlay of the tentative model of Ets1/Ets1/DNA complex from SAXS and the crystal structure of Ets1/Ets1/DNA complex**
The crystal structure of Ets1/Ets1/DNA complex is colored in cyan, the tentative model of Ets1/Ets1/DNA complex from SAXS is colored in magenta.

Figure 72 demonstrated that the X-ray crystallography model and tentative SAXS model are in a good agreement. The angle between Ets1 molecules in the tentative SAXS model is slightly larger than in X-ray crystallography model, but the shape is correct.

2.9 Comparison of crystal structure of Ets1/Ets1/DNA complex and Ets1 dimer

The models for Ets1 dimer bound to DNA and Ets1 dimer without DNA (Tahirov et. al., 2002) were compared by LSQ program. The results are illustrated on fig. 37.

Fig. 37 demonstrated that Ets1 conformation in dimer when Ets1 is bound to DNA and when Ets1 is without DNA is different.

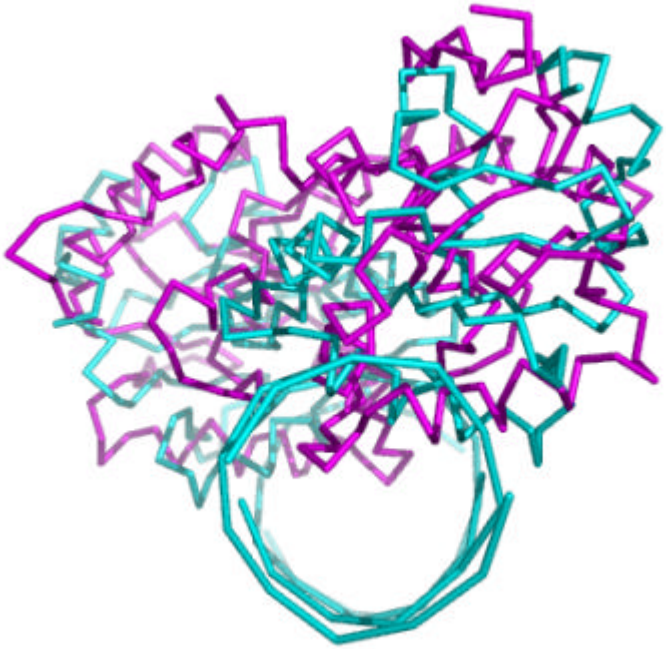


Fig. 37 **Overlay of the crystal structure of Ets1/Ets1/DNA complex and Ets1 dimer** (Tahirov et. al., 2002)
The crystal structure of Ets1/Ets1/DNA complex is colored in cyan, the crystal structure of Ets1 dimer is colored in magenta.

2.10 Preliminary conclusions

The structure for Ets1 dimer and DNA complex was solved to a resolution of 2.58 Å by molecular replacement. In the current structure N-terminal parts of Ets1 protein is missing and has to be built manually. However, according to the electron density maps, preliminary conclusion can be made that protein-protein interaction could take place.

The tentative SAXS model of Ets1/Ets1/DNA complex is in a good agreement with the crystallographic model.

Ets1 conformation in dimer when Ets1 is bound to DNA and when Ets1 is without DNA is different.

2.11 Future perspectives

Future perspectives for the project focused on structure determination of Ets1/Ets1/DNA complex would be to finish the structure and to understand the mechanism of Ets1/Ets1/DNA complex formation.

2. 12 USF1/DNA complex formation and purification

For the DNA binding experiment a dsDNA fragment was used:

tcatCACGTGgcccg
agtaGTGCACcgggc

The core binding sequence is a palindrome of 6 bases (colored in red). USF1/DNA complex was formed with molar ratio 2:1 (protein to DNA) and purified using gel filtration. Complex formation was established by "band shift assay". Identical bands for protein/DNA complexes were observed on these gels. Two bands corresponding to the protein/DNA complexes were found on the native gel (fig. 38 A, B), whereas only one band was found in SDS-PAGE (fig. 38 C), suggesting that USF1 binds to the DNA as a homodimer and as a homotetramer.

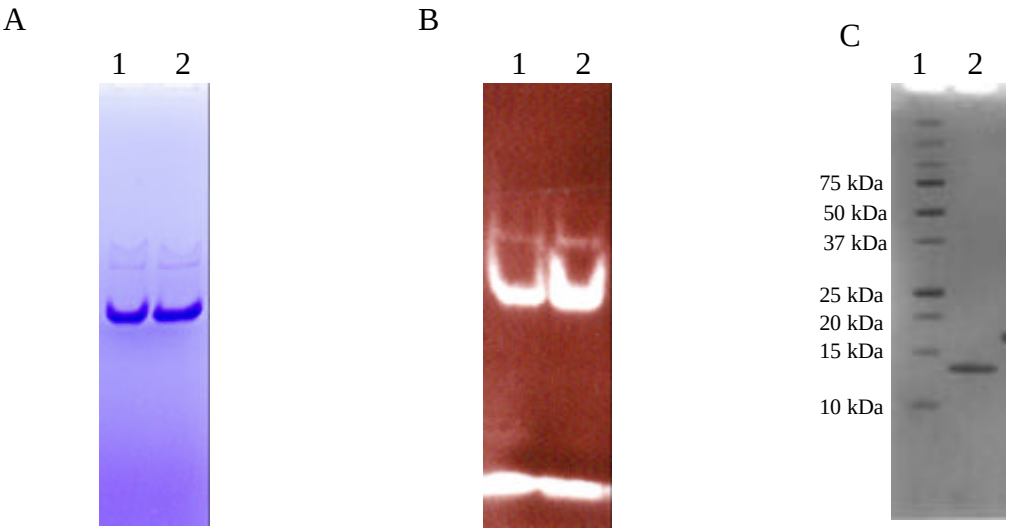


Fig. 38 **Native gel and SDS-PAGE of USF1/DNA complex**

A. Gel stained with Coomassie blue

B. Gel stained with Ethidium Bromide

Lanes 1, 2 – USF1/DNA complex, upper bands correspond to homotetramer, lower bands to dimer. On the ethidium bromide stained gel the bottom band corresponds to free DNA.

C. SDS-PAGE of USF1/DNA complex

Lane 1 – molecular weight marker, lane 2 – USF1

In order to check that USF1 bound to DNA forms bivalent homotetramers and to show that the upper band in fig. 38 indeed corresponds to bivalent tetramers and not to a contamination 2D gel was run. A native gel was run first and subsequently a lane containing USF1/DNA complexes was applied on the SDS PAGE. The result demonstrated that the upper band indeed corresponds to USF1/DNA bivalent homotetramer (data not shown).

The USF1/DNA complex was purified by gel filtration to eliminate free DNA. USF1/DNA dimers and tetramers elute in the same broad peak and all attempts to separate them or to shift the equilibrium in the direction of dimers or tetramers failed.

2.13 SAXS experiment on USF1/DNA complexes

After purification by gel filtration samples of USF1/DNA complexes were concentrated and used for SAXS measurements in the concentration range of 1 mg/ml to 8 mg/ml. Higher concentrations could not be measured due to aggregation. The experimental curve of the sample with the concentration 6 mg/ml is shown on fig. 39.

The samples were polydisperse and USF1/DNA dimer, USF1/DNA tetramer and free DNA, were detected by native gel electrophoresis.

The crystal structure of USF1 (without leucine zipper) dimer/DNA complex (Ferre-D'Amare et. al., 1994) and of the homologous protein Max (with leucine zipper) on its cognate DNA are known (Ferre-D'Amare et. al., 1993). USF1/DNA structure was complemented using the leucine zipper fragment taken from a Max/DNA complex (fig. 40). This was done using program MASSHA (Konarev et. al., 2001), which allows the subunits to be translated and rotated as rigid bodies while observing corresponding changes in the fit to the experimental data.

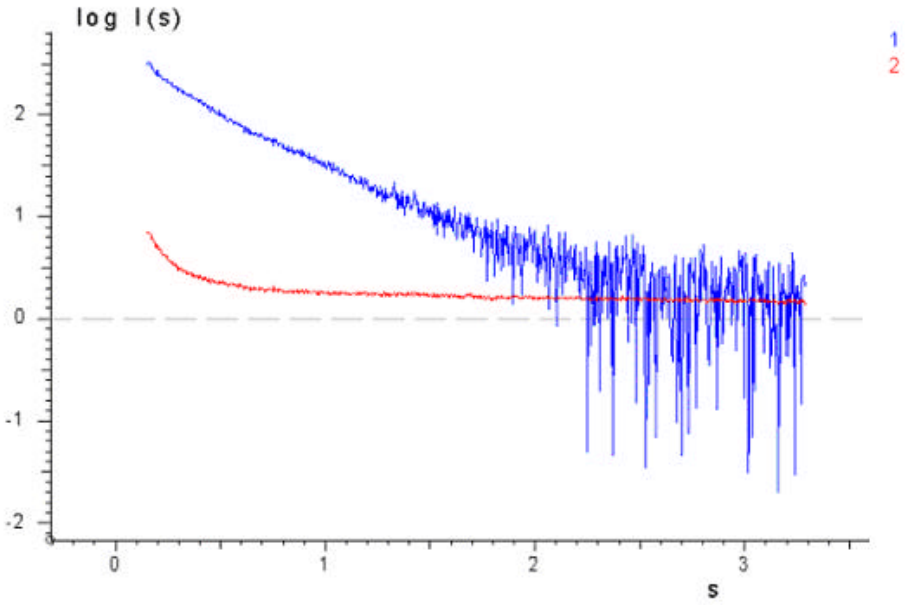


Fig. 39 **Experimental SAXS data from USF1/DNA complex at a concentration of 6 mg/ml**

Curve 1 – experimental data

Curve 2 – errors

This graphic demonstrates dependence of scattering intensity logarithm on absolute value of scattering vector.

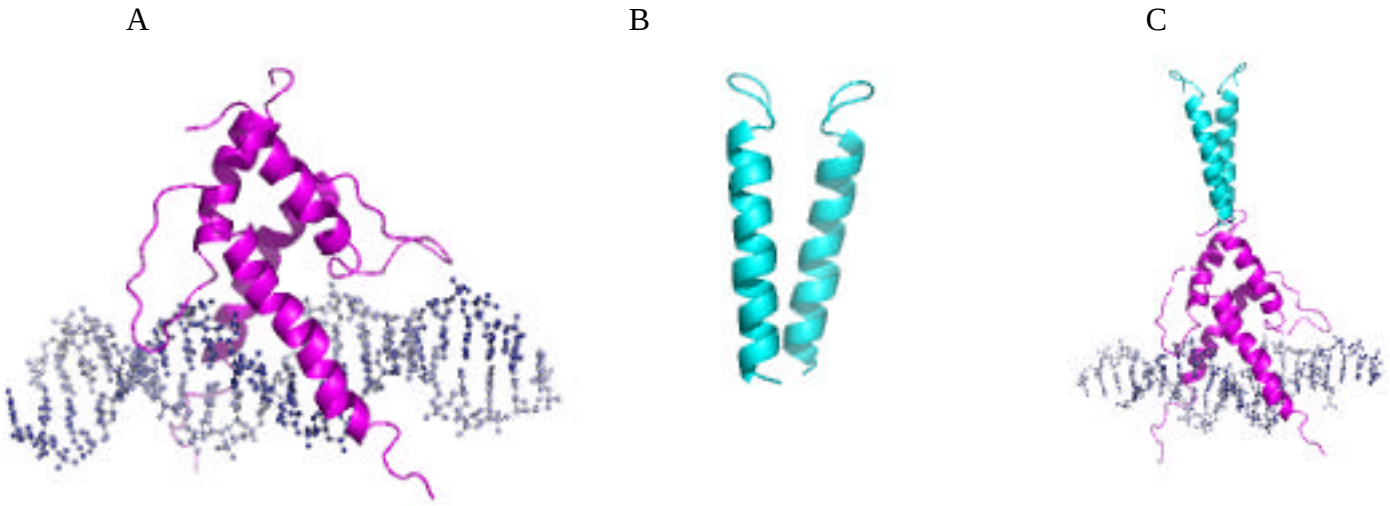


Fig. 40 **Modeling of USF1/DNA dimer**

A. USF1/DNA - crystallographic data
B. Leucine zipper Max/DNA - crystallographic data
C. Model of USF1/DNA with leucine zipper

Based on the USF1/DNA dimer model, a set of different models for USF1 bivalent homotetramer were constructed and analyzed using the program OLIGOMER (Konarev et. al., 2003), which is designed to analyze the experimental scattering curves from polydisperse solutions. The program determines volume fractions of the mixture components. USF1/DNA tetramer, USF1/DNA dimer and free DNA were taken as components of the mixture. Different USF1/DNA tetramer models were investigated. The Oligomer data, which is giving the best results, is shown on fig. 41.

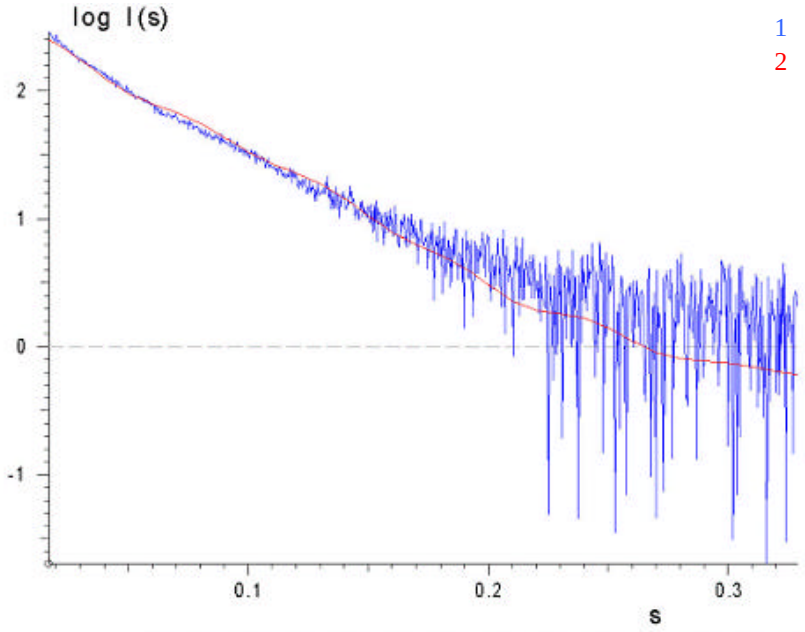


Fig. 41 **Data obtained by OLIGOMER program**

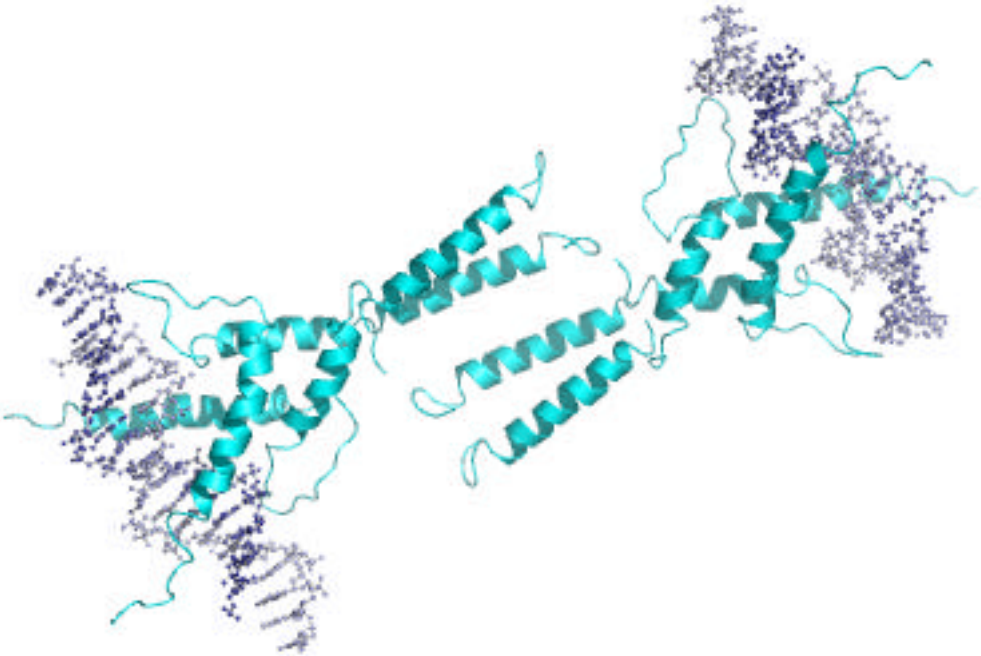
Curve 1 – experimental data from USF1/DNA complex at a concentration of 6 mg/ml

Curve 2 – theoretical curve calculated for the mixture of USF1/DNA tetramer, USF1/DNA dimmer and free DNA according to OLIGOMER program

This graphic demonstrates dependence of scattering intensity logarithm on absolute value of scattering vector.

The best tentative model of the USF1 bivalent homotetramer displays a dimer arrangement similar to the crystallographic structure of Myc-Max heterotetramer (Nair et. al., 2003) (fig. 42). Both Myc and Max proteins are homologues of USF1.

A



B

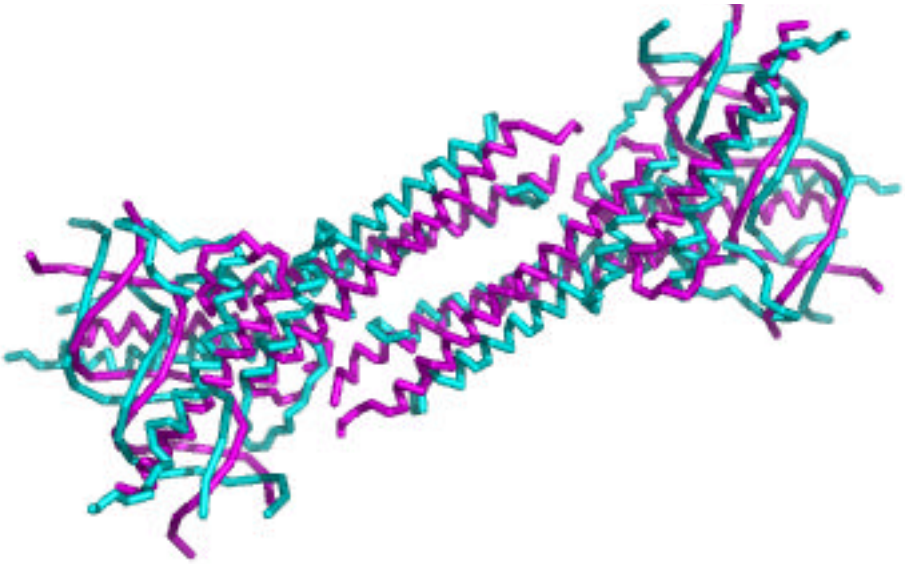


Fig. 42 **Comparison of USF1/DNA tetramer model from SAXS and Myc-Max tetramer crystal structure**

A. The best tentative model of USF1/DNA tetramer obtained from SAXS data

B. Comparison of the best tentative model of USF1/DNA tetramer obtained from SAXS data and Myc-Mac tetramer structrure done by SUBCOMB program. Tentative model of USF1/DNA tetramer from SAXS datais coloured in magenta, Myc-Mac tetramer crystal structure is coloured in green.

2.14 USF1 without DNA - SAXS model

USF1 was concentrated and used for SAXS measurements without DNA. In some samples protein was found to be polydisperse and in some monodisperse. The work was continued with monodisperse samples. The experimental curve for the USF1 monomer at a concentration 3 mg/ml is shown in fig. 43.

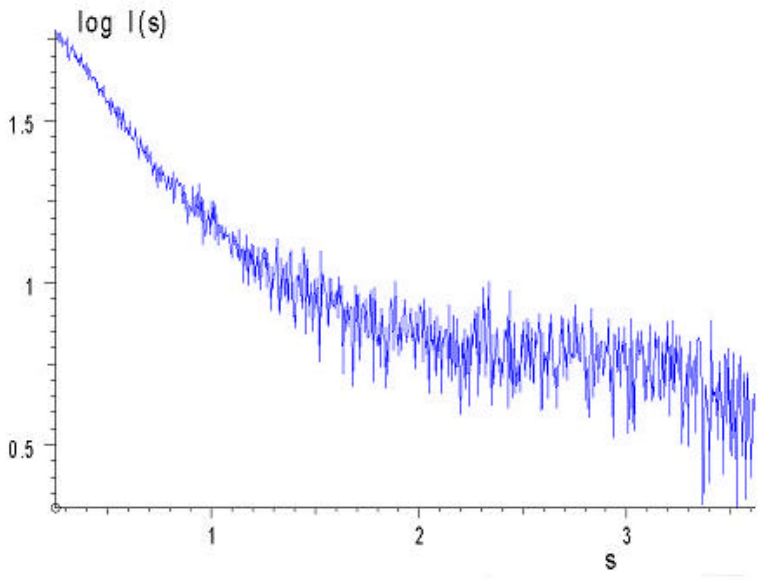


Fig. 43 **Experimental curve from USF1 monomer at a concentration of 3 mg/ml** The graphic shows dependence of scattering intensity logarithm on scattering vector

Twenty models of USF1 were created with the GASBOR program (Svergun et. al., 2001). Two representative models are shown on fig. 44.

A B

Fig. 44 **Two representative models of USF1 calculated by GASBOR program**

The average model was calculated by the program SUPCOMB and compared with a model of USF1 monomer created by extracting one USF1 molecule from USF1/DNA dimeric model built by the program MASSHA. The comparison is shown in figure 45.

A B

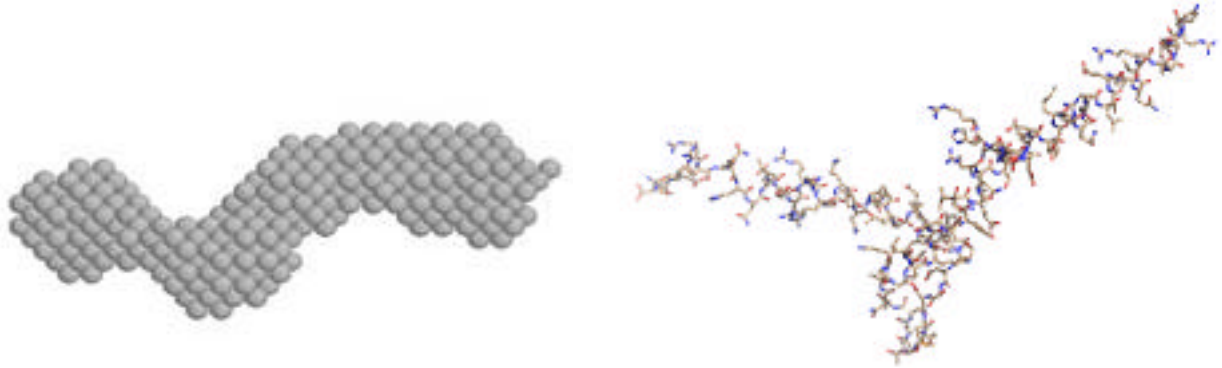


Fig. 45 **Comparison of the average USF1 model with the crystallographic model**
A. average model of USF1 monomer
B. USF1 monomer – crystallographic model

Thus, models for USF1 monomer without DNA, USF1/DNA dimer and USF1/DNA tetramer were obtained from SAXS. Biologically, the most interesting is

the model of USF1/DNA tetramer because it could potentially explain the regulation of HIV1 promoter by transcription factors binding to distal enhancer region. It could be that initially DNA is bent by LEF1 transcription factor and then, due to USF1 tetramerization, the loop is stabilized. However, SAXS experiment was done using high concentrations of USF1/DNA complex and tetramerization could occur due to high concentration. In the SAXS study polydisperse solution was used and it made data analysis less trivial.

In order to confirm USF1 tetramerization two other methods were used: fluorescence resonance energy transfer (FRET) and rotary shadowing electron microscopy (EM). This was done for DNA fragment from HIV1 LTR containing two and three USF1 binding sites. The data is shown for the studies with DNA fragment containing two USF1 binding sites.

2.15 FRET experiment

In order to validate the USF1 tetramer model from the SAXS data and to demonstrate USF1 bivalent tetramer formation at physiological concentration a FRET experiment was performed. FRET was chosen because this method allows distances to be measured. The aim was to measure the distances between labeled DNA in the presence of USF1 (protein could not be labeled because USF1 exists in different oligomerization stages). In the case of USF1 tetramer formation, two distant fragments of DNA coming close to each other and FRET will occurs, when the tetramer does not form FRET can not be detected because the distances will be too large (fig. 46). Two different experiments were performed: one with two short fragments of DNA and another one with long fragment of DNA, containing two USF1 binding site. The experiment with the long fragment of DNA is more physiological, but it has some disadvantages because the loop formation could potentially be unstable *in vitro* conditions (DNA fragments used in this study are listed in "Materials and methods" chapter).

In the nucleus USF1 concentration was estimated as 0.5 µM, assuming homogeneous protein distribution (Sawadogo et. al., 1988), as a result the minimal concentration used in FRET experiment was 0.5 µM. In order to perform FRET

experiment with two short fragments of DNA, DNA was labeled with fluorescent donor fluorescein and fluorescent acceptor CY3. In case of the long fragment of DNA containing two USF1 binding sites, PCR using fluorescent labeled primers (fluorescein and CY3) was done. The majority of experiments were performed with short fragments of DNA. Initially, the emission spectrum of dsDNA labeled with donor was measured. Then dsDNA labeled with donor was mixed with dsDNA labeled with acceptor and emission spectrum was measured. This was done ensure that the dyes do not interact. USF1 was then titrated in equimolar concentrations. As a negative controls buffer titration and BSA (bovine serum albumin) titration were performed.

In order to check whether FRET happens, the ratio F_d/F_a for each USF1 concentration was calculated, where F_d is the intensity of donor in the presence of acceptor and USF1 and F_a is the intensity of acceptor in the presence of donor and USF1. If FRET occurs, the ratio F_d/F_a should decrease upon increasing USF1 concentration during titration and then reach equilibrium. If FRET does not occur, then the ratio F_d/F_a does not change for the different USF1 concentrations and the changes of the emission spectra happens due to dilution.

No reproducible FRET signal was observed. Similar experiment was performed for long fragment of DNA with two USF1 binding sites, and again no FRET signal was detected.

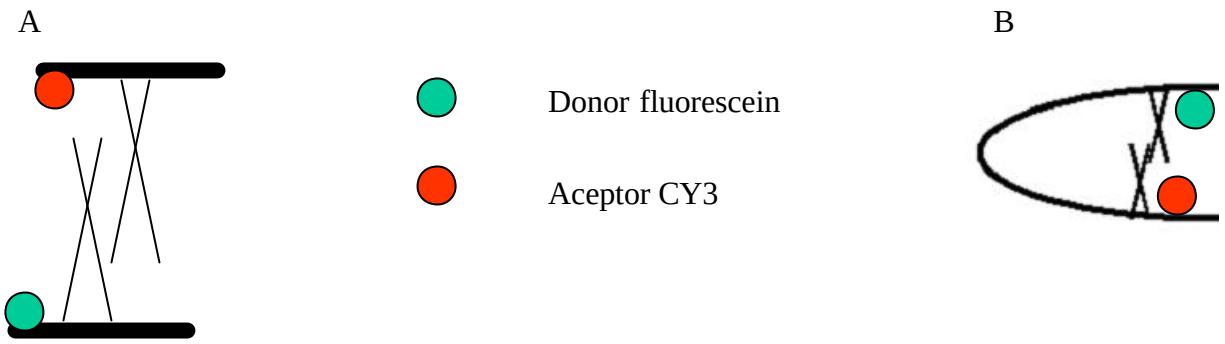


Fig. 46 **Scheme of FRET experiment**

A. with two short pieces of DNA

B. with long piece of DNA containing two USF1 binding sites

FRET can be applied when the distance between donor and acceptor is in the range of 10-100 Å. The distance estimation from USF1 tetramer SAXS model was 80-90 Å. The most probable explanation why no FRET signal was obtain would be a large distance between donor and acceptor (as labels were on the DNA, the distance between them could be larger then the distances estimated as a size of USF1 tetramer).

2.16 Rotary shadowing electron microscopy

A number of techniques are available for visualizing isolated proteins and protein-DNA complexes in the electron microscope. The most commonly used are negative staining and metal shadowing. But in spite of the fact that negative stating could potentially be used for protein-DNA complexes, this method has disadvantages. The most important of them is aggregation of protein-DNA complexes or DNA in the present of heavy metals. That's why the method of choice for protein-DNA complexes investigation by electron microscopy should be metal shadowing.

Formation of the bivalent homotetramer may lead to the DNA looping recruiting USF1 from the distal region of the promoter to the initiator element (Inr) (Ferre D'Amare et. al., 1994). In order to investigate the DNA loop formation electron microscopy experiment was performed. The DNA fragment containing E-box and Inr element was amplified via PCR and had a length of 525 bp which corresponds to approximately 300 kDa and can be visualized by EM. The smaller fragment which would corresponds to HIV1 promoter region and distal initiator element (100 kDa) could be difficult to visualize. The DNA fragment used in EM studies had USF1 promoter binding site on one end and USF1 binding site in the middle from distal enhancer region (fig. 47). The complex between DNA and USF1 was formed and the complex formation was demonstrated by "band-shift assay" (fig. 48 illustrating only Ethidium Bromide staining). USF1 was titrated and it can be seen that molecular weight of the complexes formed between USF1 and DNA is increased, but it is difficult to say what kind of complexes were formed because migration of the molecules on native gel not only depends on the molecular weight but also on the shape and charge of the molecule. Different types of the complexes which could be

formed between USF1 and DNA are illustrated on fig. 49. It is known from the literuture (Du et. al., 1993) that USF1 binds to E-box situated in distal enhancer region of HIV1 LTR with a higher affinity than to Inr element, but the binding constants were not determined. Fig. 49 demonstrates only some of the complexes which could be formed between USF1 and DNA fragment containing E-box and Inr element.

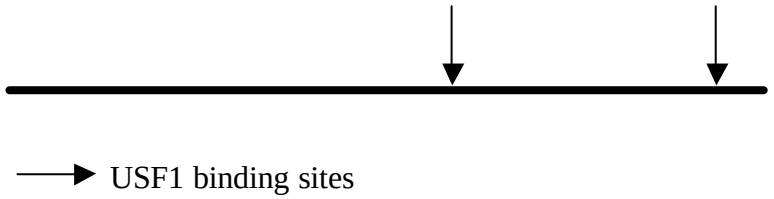


Fig. 47 **The DNA fragment used in EM studies**
USF1 binding sites are indicated by arrows

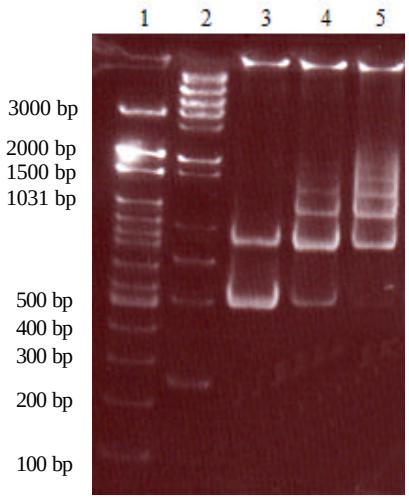


Fig. 48 **Native gel of USF1 and DNA complexes stained with ethidium bromide**
Lane 1- 100 bp DNA ladder plus, lane 2 - 1 kb DNA ladder, lanes 3-5 - PCR (HIV LTR) + USF1
(with increased protein concentration). The lower band on the lane 3 corresponds to free DNA fragment, all the upper bands on lanes 3-5 corresponds to different protein-DNA complexes.

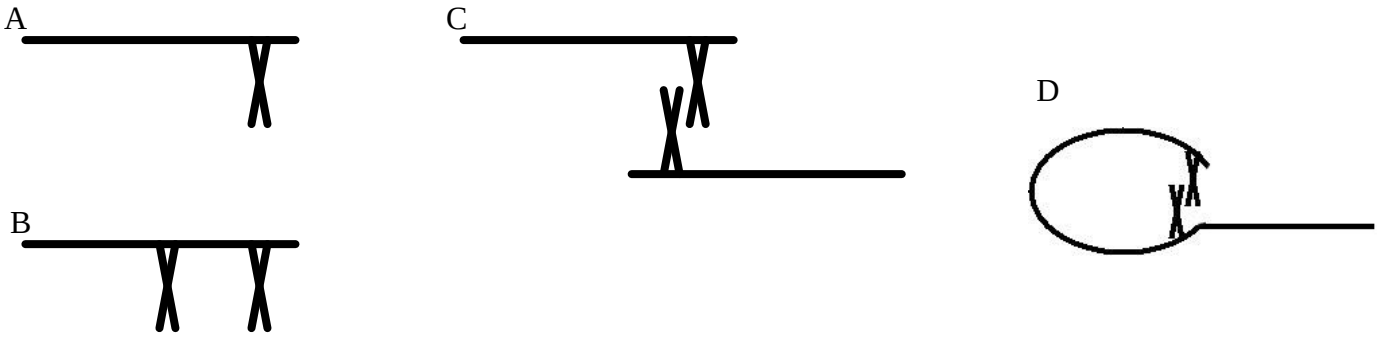


Fig. 49 **Different complexes which could be formed between USF1 and DNA**

A. USF1 bound to E-box

B. USF1 bound to E-box and Inr element

C. USF1 formed tetramer and brought to DNA fragments together

D. USF1 formed tetramer leading to loop formation

After the complexes were formed and analyzed on native gel they were prepared for rotary shadowing electron microscopy. As a negative control free DNA was used. The results for USF1/DNA complexes are shown on fig. 50.

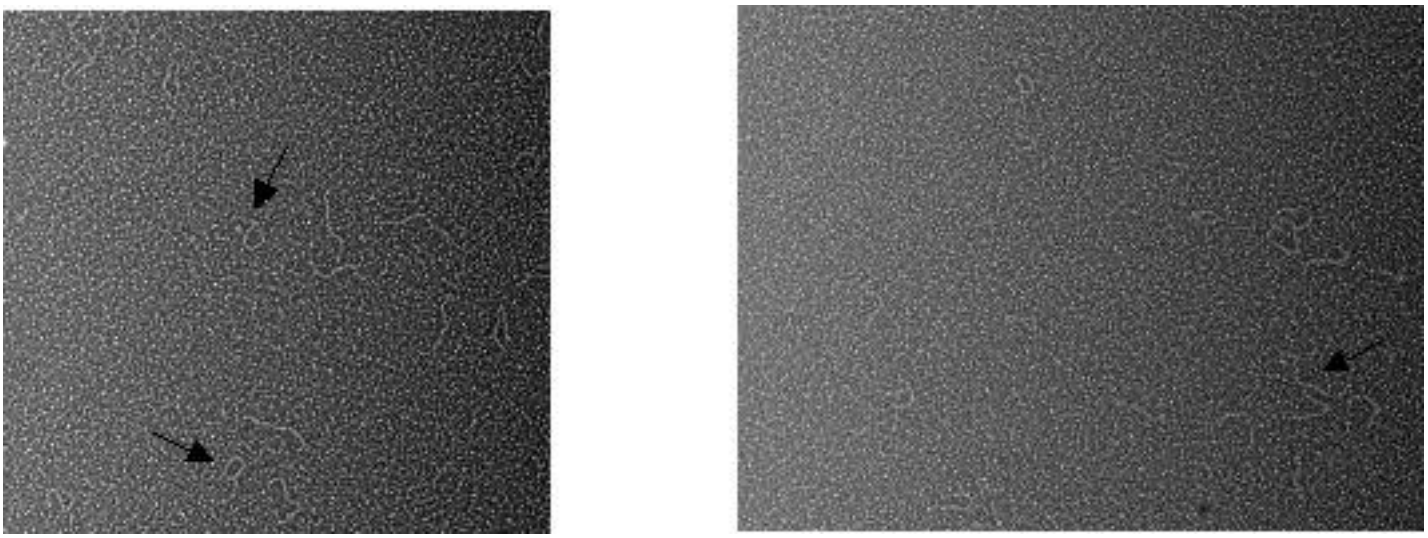


Fig. 50 **Rotary shadowing EM images for USF1/DNA complex**

A. two loops indicated by arrows (corresponds to fig. 49 D)

B. two DNA fragments bound together indicated by arrow (corresponds to fig. 49 C)

No loops or other interesting formations were observed for the control samples (free DNA). However, for USF1/DNA complexes the loops were a minor product and they could be artifacts (because the DNA could bend on its own without protein). It is difficult to draw conclusions and to distinguish between real loop formation and artifact.

2.17 Crystallization of USF1/DNA complex

The attempts to crystallize USF1/DNA complex were made. Only tiny needles were obtained so far which could not be optimized.

2.18 Conclusions

A SAXS model for USF1-monomer bHLHZip region without DNA was built. It was compared with the crystallographic model. There is no crystal structure for USF1 without DNA. The structural model for USF1-monomer bHLHZip region was made by extraction of the USF1 monomer from the USF1/DNA dimer model from SAXS containing Zip region. The crystallographic model is similar to the SAXS model calculated by GASBOR (fig. 45).

SAXS models of USF1 containing bHLHZip region bound to the E-box DNA fragment, of the USF1/DNA dimer and of the USF1 tetramer were built. USF1/DNA tetramer showed similar dimers arrangement as in Myc-Max tetramer structure (fig. 42). But tetramer formation could be an artifact due to the high concentrations used in SAXS experiments. USF1 tetramer formation was not demonstrated by FRET experiment or by rotary shadowing EM.

2.19 Future perspectives

Since, no good quality crystals were obtained up to now for USF1/DNA complex it is reasonable to continue crystallization trials of the USF1/DNA complex.

2.20 Ets1/USF1/DNA complex formation and purification

For the DNA binding experiment a dsDNA fragment was used:

tcatCACGTGgcccgagagctgCATCCGgagta
AgtaGTGCACcgggctctcgacGTAGGCctcat

The core binding sequences for Ets1 and for USF1 are colored in red. USF1/Ets1/DNA complex was formed with molar ratio 2:1:1 (USF1 : Ets1 : DNA). Complex formation was established by "band shift assay" (fig. 51). Fig. 52 demonstrates SDS-PAGE of USF1, Ets1 (301-441 and 280-441) and ternary complexes.

Ternary complex purification protocol provided by Fransisco Fernandez was modified in the following way: native protein Ets1 was used for the complex formation and purification, ternary complex was formed by mixing Ets1, USF1 and DNA in equimolar ratio in contrast to the protocol provided by Fransisco Fernandez where denatured Ets1 protein was refolded by dilution in the presence of USF1/DNA complex.

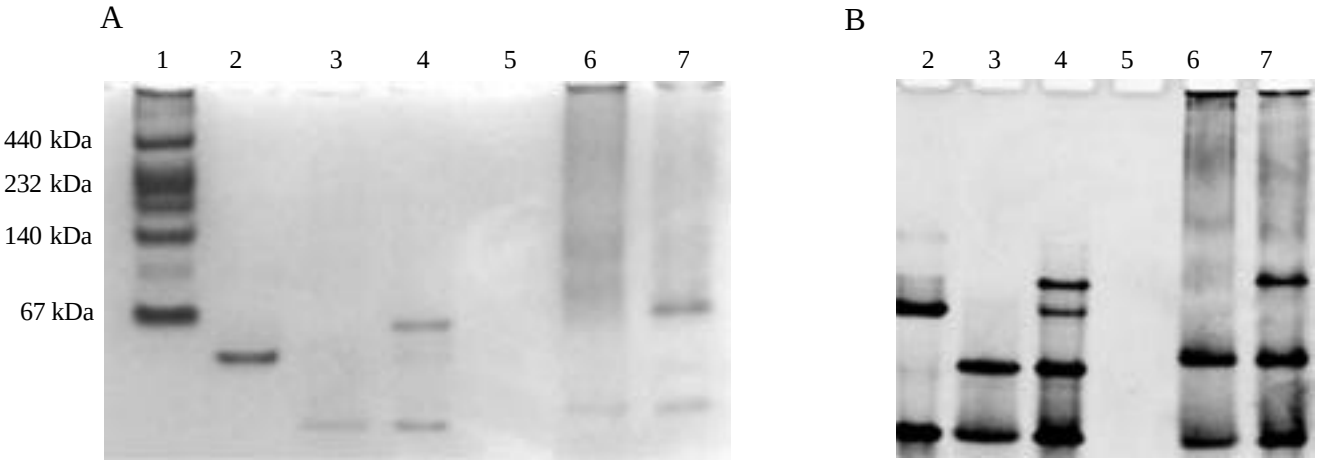


Fig. 51 **Native gel of USF1, Ets1 and DNA complexes**

A. gel stained with coomassie blue

B. gel stained with Ethidium Bromide

Lane 1 – molecular weight marker, lane 2 – USF1/DNA, lane 3 – Ets1 (301-441)/DNA, lane 4 – ternary complex Ets1 (301-441)/USF1/DNA, lane 6 – Ets1 (280-441)/DNA, lane 7 – ternary complex Ets1 (280-441)/USF1/DNA

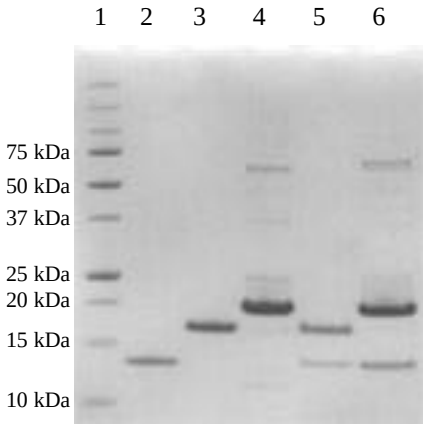


Fig. 52 **SDS-PAGE of USF1, Ets1 and ternary complexes**
Lane 1 – molecular weight marker, lane 2 – USF1, lane 3 – Ets1 (301-441), lane 4 – Ets1 (280-441), lane 5 – ternary complex Ets1 (301-441)/USF1/DNA, lane 6 – ternary complex Ets1 (280-441)/USF1/DNA

As illustrated in fig. 51, binary complexes Ets1/DNA and USF1/DNA were always present in addition to ternary complex. In order to purify the ternary complex and eliminate binary complexes gel filtration was used. However, it was not possible to separate ternary complex from binary complexes on gel filtration column. This was due to the fact that USF1/DNA complex has molecular weight 48 kDa and elongated shape and migrate faster on gel filtration column. The molecular weight for ternary complex formed with Ets1 (301-441) is 64 kDa and with Ets1 (280-441) is 68 kDa.

2.21 Crystallization of Ets1/USF1/DNA ternary complex

The crystallization attempts were mostly focused on the ternary complex formed with Ets1 (280-441). In crystallization screens two temperatures were tested (room temperature and 4°C) and 960 different conditions were tried. For the ternary complex formed with Ets1 (301-441) two temperatures were tested (room temperature and 4°C) and 480 different conditions were tried.

Attempts to crystallize the ternary complex only yielded DNA crystals, indicating that the ternary complex was of limited stability. The fact that no crystals for the ternary complex were obtained can have several explanations:

1. The DNA used for crystallization has length 32 bp which is quite long and can be flexible.

2. The ternary complex used for crystallization was not pure, it always has USF1/DNA complex in addition.

2.22 Conclusions

The ternary complex Ets1/USF1/DNA was formed and purified. All the attempts to crystallize the ternary complex did not give crystals.

3. Materials and Methods

3.1 Materials

3.1.1 Chemicals

All commonly used chemicals were purchased from either Mersk, Sigma, Roth, Fluka, Qiagen, Amersham-Biotech or BioRad.

3.1.2 Buffers

A list of most commonly used buffers during this experimental work is shown below.

10X ligase buffer (NEB)
100 mM Tris-HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP, 250 mg/ml BSA

TAE-buffer (running buffer for agarose gel electrophoresis)
50X TAE buffer contained 242 g Tris base, 57.1 ml of glacial acetic acid and 18.61 g of EDTA, and adjusted to 1 L solution with Milli Q H_2O.

Running buffer for SDS-PAGE
20X MES SDS-PAGE running buffer contained 390 g MES, 242 g Tris Base, 10% (w/v) SDS, 12 g EDTA and adjusted to 2 L solution with Milli Q H_2O.

Running buffer for native gel (TBE buffer)
10X TBE buffer contained 108 g Tris Base, 55 g boric acid, 20 ml of 0.5 M EDTA pH 8 and adjusted to 1 L solution with Milli Q H_2O.

UV-cocktail
6 M guanidine hydrochloride, 0.02 M phosphate buffer pH 7.5

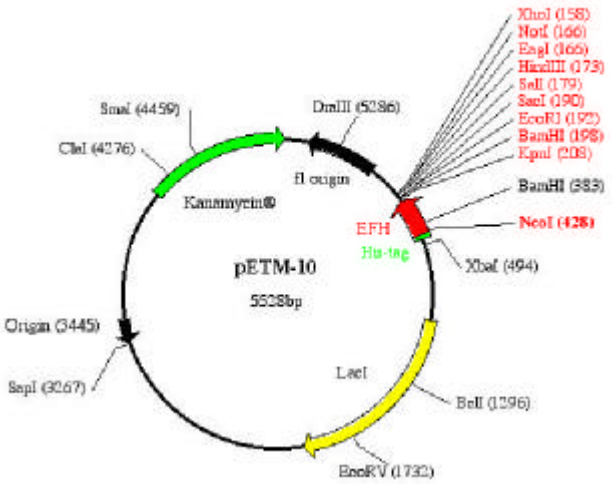
3.1.3 Media

SOB media
20 g/L bacto-tryptone, 5 g/L yeast extract, 0.5 g/L mM NaCl, 0.186 g/L KCl, 10 mM $MgCl_2^*$ were adjusted to 1 L solution with Milli Q H_2O.
* was added just before use

SOC media
SOB medium, containing 20 mM glucose
20 ml of sterilized 1 M glucose to 1 L of SOB media added just before use

Luria-Bertani media (LB)
10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl were adjusted to 1 L solution with Milli Q H_2O. pH was adjusted to 7.4

3.1.4 Expression vectors



pETM-10

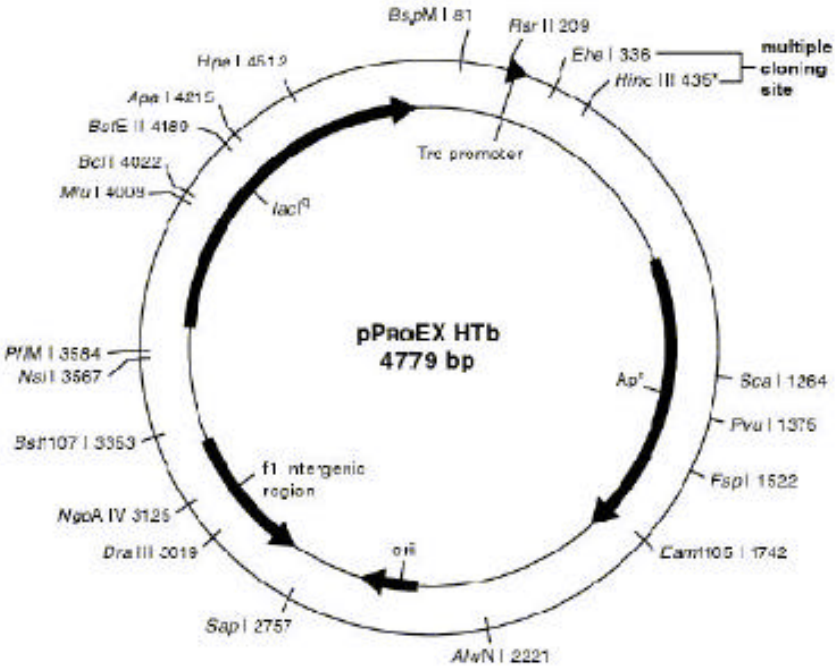
```
     T7 promoter  -->            Lac operator     XbaI
CGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAA
GCTTTAATTATGCTGAGTGATATCCCCTTAACACTCGCCTATTGTTAAGGGGAGATCTTT

                    rbs                      His-tag
TAATTTTGTTTAACTTTAAGAAGGAGATATACCATGAAACATCACCATCACCATCACCCC
ATTAAAACAAATTGAAATTCTTCCTCTATATGGTACTTTGTAGTGGTAGTGGTAGTGGGG
                                  METLysHisHisHisHisHisHisPro

NcoI
ATGGCTGACACCGACACTGCCGAG..171bp..GCACTCTACGGGGAGAGCGATCTGTGA
TACCGACTGTGGCTGTGACGGCTC...EFH...CGTGAGATGCCCCTCTCGCTAGACACT
METAlaAspThrAspThrAlaGlu...57aa..AlaLeuTyrGlyGluSerAspLeu***

                                        NotI
       BamHI        SacI                EagI
 KpnI        EcoRI          SalI HindIII        XhoI  C-His-tag
GGTACCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCAC
CCATGGCCTAGGCTTAAGCTCGAGGCAGCTGTTCGAACGCCGGCGTGAGCTCGTGGTGGTG
                                                  HisHisHis
```

pProExHTb expression vector



RBS HIS-TAG

```
ACAGGAAACAGACCATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCA
TGTCCTTTGTCTGGTACCACATGATGGTAGTGGTAGTGGTAGTGCTAATGCTATAGGGT
            MetSerTyrTyrHisHisHisHisHisHisAspTyrAspIlePro
```

EheI BamHI

TEV-site NcoI EcoRI StuI SalI

```
ACGACCGAAAACCTGTATTTTCAG GGCGCCATGGGATCCGGAATTCAAAGGCCTACGTCG
TGCTGGCTTTTGGACATAAAAGTC CCGCGGTACCCTAGGCCTTAAGTTTCCGGATGCAGC
ThrThrGluAsnLeuTyrPheGln|GlyAlaMetGlySerGlyIleGlnArgProThrSer
```

SstI SpeI NotI NspV XbaI PstI XhoI KpnI

```
ACGAGCTCACTAGTCGCGGCCGCTTTCGAATCTAGAGCCTGCAGTCTCGAGGCATGCGGT
TGCTCGAGTGATCAGCGCCGGCGAAAGCTTAGATCTCGGACGTCAGAGCTCCGTACGCCA
ThrSerSerLeuValAlaAlaAlaPheGluSerArgAlaCysSerLeuGluAlaCysGly
```

HindIII

```
ACCAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGA
TGGTTCGAACCGACAAAACCGCCTACTCTCTTCTAAAAGTCGGACTATGTCT
ThrLysLeuGlyCysPheGlyGly***
```

3.1.5 Oligonucleotides

Oligonucleotides used for FRET

E-box CY3 (forward primer)	5'- CY3-TTCATCACGTGGCCCGAGAG-3'
compl E-box CY3	5'- CTCTCGGGCCACGTGATGAA-3'
E-box fluorescein	5' – Fluorescein-TCATCACGTGGCCCG – 3'
compl E-box fluorescein	5' – CGGGCCACGTGATGA – 3'
Inr-Oregon green (reverse primer)	5' – Oregon green – ACCAGAGAGACCCAGTACAGGC – 3'

Oligonucleotides used for rotary shadowing EM

forward primer	5' – ATCCTTGATCTGTGGATCTA – 3'
reverse primer	5' – TTTACCAACAGTACCGGAAT – 3'

3.2 Methods

3.2.1 Sub-cloning

Ets1 constructs (residues 335-441, 301-441 and 280-441) were subcloned into pETM10 expression vector, which has a noncleavable N-terminal His-tag. For this purpose, pETM10 vector and vector containing Ets1 DNA sequence were cut by NcoI and KpnI restriction enzymes, digested insert DNA and digested vector were purified and ligated together. The detailed procedure is described below (the protocols are taken from EMBL-Hamburg web-page and made by Arie Geerlof).

3.2.1.1 Digestion of insert or vector DNA

Mix in 500-µl microfuge tubes:

5 µl	10X restriction enzyme buffer 1 (NEB)
0.5 µl	100X BSA
	DNA (7.5 µg of plasmid DNA containing the fragment of interest or of expression vector)
2.5 µl	restriction enzyme KpnI (10-20 units/µl)
2.5 µl	restriction enzyme NcoI (10-20 units/µl)
	add sterile water to a volume of 50 µl

1. Add the restriction enzymes last.
2. Mix gently by tapping the tube or pipetting the solution up and down.
3. Incubate the reaction mixtures at 37°C for 2-4 h.

3.2.1.2 Purification of digested insert DNA or digested vector

To separate the digested vector or the insert DNA from other DNA fragments generated by digestion, the samples is purified using preparative agarose gel electrophoresis. The vector DNA or insert DNA is extracted from the agarose gel using a commercial gel extraction kit from QIAGEN.

3.2.1.3 Ligation of DNA fragments with sticky ends

For a standard ligation reaction of DNA fragment with 2-4 base sticky ends, we set-up reactions under the following conditions:

- 100 ng of digested vector DNA.
- Digested insert to vector DNA in a molar ratio of 1:1, 2:1, and 3:1.
- Two control reactions with only digested vector or insert DNA to determine the background of non-recombinants.

The concentration of insert and vector DNA can be determined by measuring the absorbance at 260 nm, assuming that a solution of 50 ng/µl gives an A_{260} of 1.

Ligation reactions

Mix in 500-µl microfuge tubes:

2 µl	10X ligase buffer
2 µl	digested vector DNA (50 ng/µl)
	the appropriate amount of digested insert DNA (see above)
1 µl	T4 DNA ligase (20 NEB units/µl)
	add sterile water to a volume of 20 µl

1. Add the ligase last.
2. Mix gently by pipetting the solution up and down.
3. Incubate the reaction mixtures at 16°C for 2 h to overnight.

3.2.1.4 Transformation of plasmid DNA to chemically competent E. coli cells

1. Thaw the appropriate amount of competent cells on ice. Also pre-chill the required number of empty 1.5 ml microcentrifuge tubes.
2. Pipet 200 µl aliquots of cells into the pre-chilled tubes.
3. Add max. 20 µl of a ligation reaction mix or 5 ng of pure plasmid DNA to each tube. Mix gently.
4. Incube the tubes of ice for 30 min.
5. Heat shock the cells for 60 sec at 42°C.
6. Place the tubes immediately on ice for at least 2 min.
7. Add 800 µl of SOC medium to each tube, transfer the suspensions to 15 ml Falcon tubes, and incubate for 1 hour at 37°C. When pure plasmid DNA was used for the transformation, plate out 1, 10 and 100 µl of the suspensions directly on LB agar plates containing the appropriate antibiotic. To facilitate the speading of the sample, add the 1 and 10 µl aliquots to 100 µl of SOC medium.
8. Transfer the cultures to 1.5 ml microcentrifuge tubes and spin for 1 min at 6000 rpm.
9. Remove 800 µl of the supernatant and resuspend the pellet.
10. Plate out the suspension on a LB agar plate containing the appropriate antibiotic.
11. Incubate the plates overnight at 37°C.

3.2.1.5 Colonies selection

Agar plates should be used with appropriate antibiotic that's why only colonies containing the vector with the insert should grow. The DNA is isolated from colonies and analyzed by restriction to prove that the insert is present and sequenced.

3.2.2 Expression and Solubility studies

Preparation of soluble/insoluble protein from cells

Jeanne Perry (Molecular Biology Institute, UCLA, Los Angeles, USA)

The solubility of a protein depends strongly on the composition of the lysis buffer. Using the procedure described below the solubility of a specific protein can be tested under many different conditions.

1. Grow the cells and induce protein expression using 1 mM IPTG
2. Take so many 1 ml-samples from the cell culture at OD_{600} of 1 as experiments you would like to carry out.
3. Spin down the cells for 5 min at 6000 rpm in a microfuge.
4. To each cell pellet, add 100 µl of the appropriate buffer (see below)

For a first screen use the buffer described in below:

7.5N	50 mM Tris pH 7.5, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme
2S	50 mM Tris pH 7.5, 2 M NaCl, 5 mM EDTA, 1 mg/ml lysozyme
0.5U	50 mM Tris pH 7.5, 50 mM NaCl, 5 mM EDTA, 0.5 M urea, 1 mg/ml lysozyme
D	20 mM Tris pH 7.5, 50 mM NaCl, 0.2% NP 40, 1 mg/ml lysozyme

5. Vortex to resuspend the cells.
6. Sonicate (using a microtip) or Freeze-thaw (see below for protocol) to lyse cells.

   Freeze-thaw protocol:

   - freeze quickly on dry ice and leave for 3 min.
   - thaw immediately at 42 °C. Vortex vigorously to mix well.
   - Repeat the two previous steps for three more times (4 freeze-thaw-vortex cycles in all).

7. Spin the tubes for 5 min at maximum speed in a microfuge.
8. Separate the supernatant (contains soluble protein) from the pellet (contains insoluble protein) by pipetting out the supernatant to a clean tube.
9. Label the supernatant "S__" (fill in with the buffer name, *e.g.* "S7N").

10. Label the pellet "P__" (fill in with the buffer name, *e.g.* "P7N").

11. To each supernatant, add 1 ml acetone and vortex. Freeze or leave on ice for 15 min. Spin 5 min at maximum speed. Remove the acetone by pipetting it out, being careful not to disturb the pellet.

12. Dry at 37 °C.

13. To the acetone-treated pellet, add 15 µl SDS-Page loading buffer.

14. To the cell pellet, add 25 µl SDS-Page loading buffer.

15. Heat all samples to 95 °C (or greater) for 5 minutes.

16. Vortex and centrifuge 5 min at maximum speed.

17. Load 10 µl on a SDS-PAGE gel, taking sample from the top and avoiding any pellet.

Based on results, follow up by trying another set of screens and then continue with combinations.

pH solubility screens

5N	50 mM Na Acetate pH 5, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme
6N	50 mM MES pH 6, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme
7N	50 mM Tris pH 7, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme
8N	50 mM Tris pH 8, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme
9N	50 mM Tris pH 9, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme

Salt solubility screen

0.1S	50 mM Tris pH 7.5, 0.1 M NaCl, 5 mM EDTA, 1mg/ml lysozyme
0.5S	50 mM Tris pH 7.5, 0.5 M NaCl, 5 mM EDTA, 1mg/ml lysozyme
1S	50 mM Tris pH 7.5, 1 M NaCl, 5 mM EDTA, 1mg/ml lysozyme
0.1KS	50 mM Tris pH 7.5, 0.1 M KCl, 5 mM EDTA, 1mg/ml lysozyme
1KS	50 mM Tris pH 7.5, 1 M KCl, 5 mM EDTA, 1mg/ml lysozyme

Urea solubility screen

0.5U	50 mM Tris pH 7.5, 0.5 M Urea, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme
1U	50 mM Tris pH 7.5, 1 M Urea, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme
2U	50 mM Tris pH 7.5, 2 M Urea, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme
3U	50 mM Tris pH 7.5, 3 M Urea, 50 mM NaCl, 5 mM EDTA, 1 mg/ml lysozyme
4U	50 mM Tris pH 7.5, 4 M Urea, 50 mM NaCl, 5 mM EDTA
5U	50 mM Tris pH 7.5, 5 M Urea, 50 mM NaCl, 5 mM EDTA
6U	50 mM Tris pH 7.5, 6 M Urea, 50 mM NaCl, 5 mM EDTA

Detergents solubility screen

D	20mM Tris HCl pH 7.5, 50 mM NaCl, 0.2% NP 40, 1 mg/ml lysozyme
X	20mM Tris HCl pH 7.5, 50 mM NaCl, 0.2% triton X-100, 1 mg/ml lysozyme
T	20mM Tris HCl pH 7.5, 50 mM NaCl, 0.2% Tween-20, 1 mg/ml lysozyme
M	20mM Tris HCl pH 7.5, 50 mM NaCl, 0.2% dodecylmaltoside, 1 mg/ml lysozyme

Stabilizer

10G	20mM Tris HCl pH7.5, 50 mM NaCl, 10% glycerol, 1 mg/ml lysozyme
50G	20mM Tris HCl pH7.5, 50 mM NaCl, 50% glycerol, 1 mg/ml lysozyme

The procedure for expression and solubility study was the following: first expression test was curried out (for this purpose the protocol for solubility study was used with one exception: in N4 the most common buffer (20 mM Tris pH 8, 0.3 M NaCl, 5 mM Imidazole, 1mg/ml lysozyme) and different temperatures and different times after induction were used). If after expression study the protein was found to be expressed then solubility test was curried out if it was necessary.

Based on the results of the solubility study lysis buffers for Ets1 and USF1 were chosen. Expression and initials lysis conditions for Ets1 and USF1 are described in “Results and discussions” chapter.

3.2.3 Protein purification

The way of purification varies for different proteins and based on their properties. For recombinant proteins, which have His-tag, the first step of purification is affinity chromatography Ni-NTA, which normally gives relatively pure protein. If His-tag can be removed it is done using certain enzyme (quite often TEV-protease is used). Further purification procedure can be presented by ion exchange chromatography, and the last step is gel filtration, which allows to get pure protein suitable for crystallization trials.

3.2.4 USF1 expression and purification

Bacterial culture was grown at 37°C up to an OD_{600} = 0.6 and then induced with 1mM IPTG for 1.5 hours. The cells were harvested by centrifugation at 6000 rpm for 30 mins. The bacterial pellet was resuspended in lysis buffer (20 mM Tris-HCl, pH 8, 300 mM NaCl, 5 mM imidazole) by addition of 30 ml buffer for each 4g of pellet. EDTA-free protease inhibitor mix (Roche), 1mg of lysosyme and 100µg Dnase I were added for each 30 ml of lysis buffer and the reaction was incubated on ice for 30 mins. The sample was sonicated and centrifuged for 1 hour at 18000 rpm in order to separate the soluble and unsoluble fractions. The soluble fraction containing USF1 protein (10-15mg protein for 1 liter of bacterial culture) was filtered and applied to a Ni-NTA column equilibrated in lysis buffer. The column was washed with lysis buffer and the protein was eluted with lysis buffer containing an additional 400mM imidazole (fig. 53).

After Ni-NTA affinity chromatography TEV protease was added in the ratio w/w 1:50 (TEV protease: protein) and the sample was incubated overnight at 4°C. Cleavage was checked by SDS-PAGE (fig. 54).

The next purification step was cation exchange chromatography using a Mono S column (fig. 55). USF1 protein was dialyzed against buffer (50mM NaCl, Tris-HCl pH=7.5, 1mM EDTA) and eluted on a gradient of 0.05 - 1M NaCl. USF1 eluted between 0.4-0.5M NaCl.

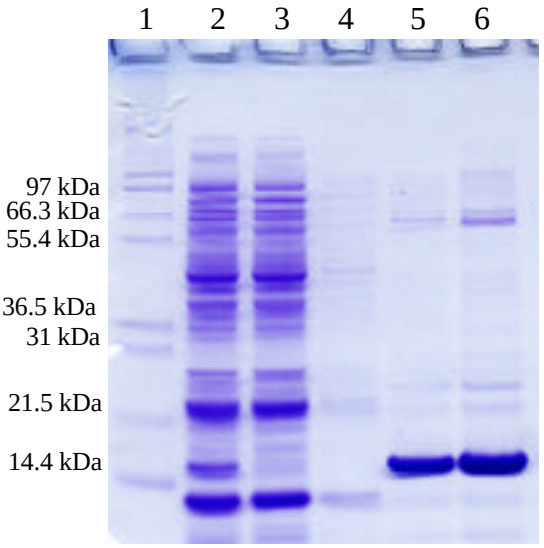


Fig. 53 **Metal affinity chromatography (Ni-NTA) of USF1 protein**

Lane 1 – protein marker 12, lane 2 – lysate, lane 3 – flow through, lane 4 – wash, lanes 5 and 6 – elution fractions

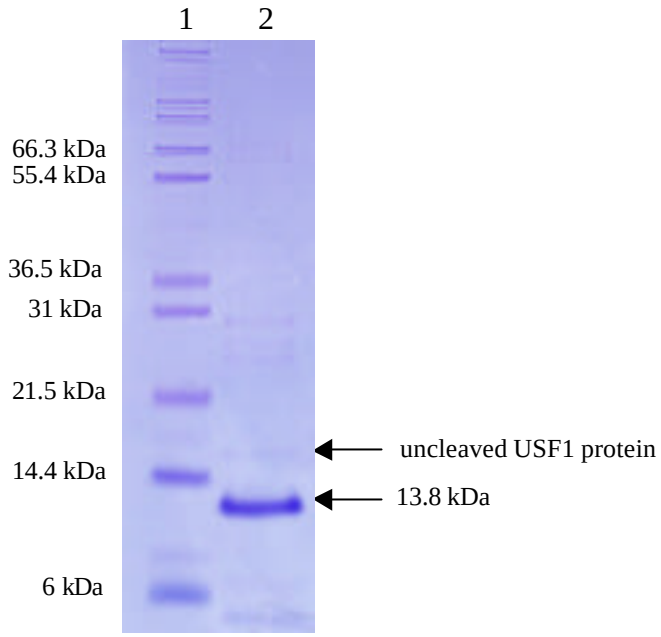


Fig. 54 **SDS-PAGE of USF1 cleaved by TEV-protease**

Lane 1 – protein marker 12, lane 2 – USF1 cleaved by TEV protease

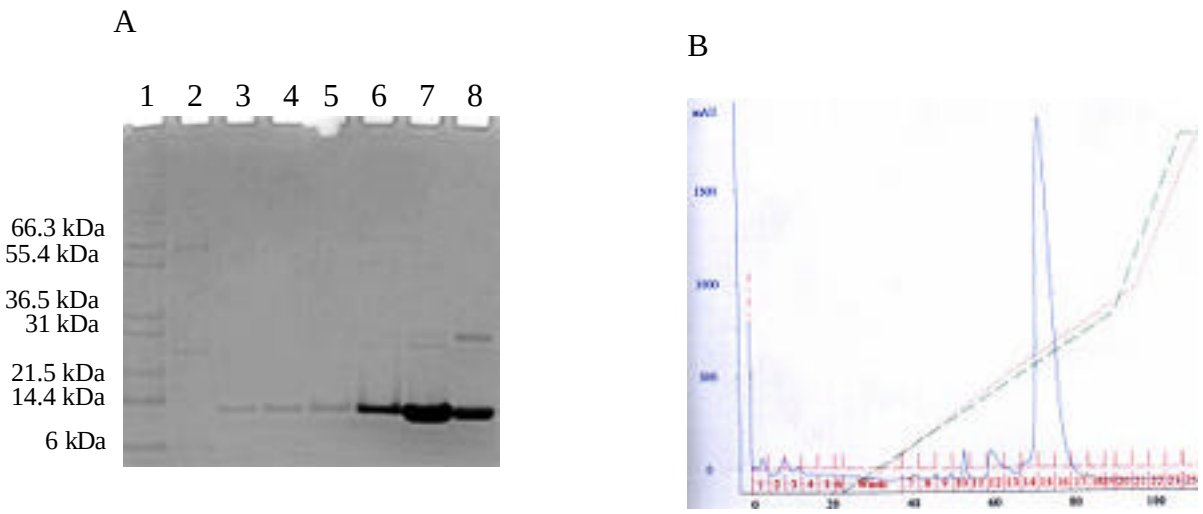


Fig. 55 **Cation exchange chromatography of USF1**

A. SDS-PAGE of USF1 after cation exchange chromatography

Lane 1 – protein marker 12, lanes 2-8 – elution fractions

B – chromatogram of USF1 purification by cation exchange chromatography

The last purification step by size-exclusion chromatography – gel filtration in a buffer containing 200mM NaCl, 20mM Tris-HCl pH=8, 2mM EDTA (fig. 56), yielded pure protein.

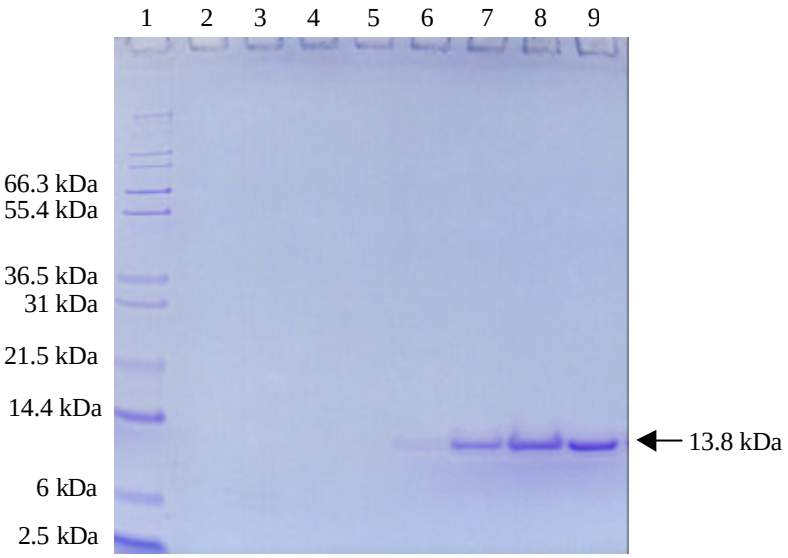


Fig. 56 **Gel-filtration of USF1 (Superdex 75 16/60)**

Lane 1 – molecular weight marker, lanes 6-9 – USF1 protein after gel filtration

3.2.5 Ets1 expression and purification

For Ets1 protein from the constructs corresponding to residues 301-441 and 280-441 the same expression and purification procedure were used. The experimental data is shown for Ets1 construct corresponding to the residues 280-441.

Bacterial culture was grown at 25°C up to OD_{600} = 0.6-0.8 and then induced with 1mM IPTG overnight. The cells were harvested by centrifugation at 6000 rpm during 30 mins. The bacterial pellet was resuspended in lysis buffer (20 mM Tris-HCL, pH 8, 300 mM NaCl, 5 mM imidazole) by addition of 30 ml buffer for each 4g of pellet. EDTA-free protease inhibitor mix (Roche), 1mg lysosyme and 100µg Dnase I were added for each 30 ml of lysis buffer and the reaction was incubated on ice for 30 mins. The sample was sonicated and centrifuged for 1 hour at 18000 rpm in order to separate the soluble and insoluble fractions. The soluble fraction containing Ets1 protein (circa 10 mg protein for 1 liter of bacterial culture) was filtered and applied to a Ni-NTA column previously equilibrated with lysis buffer. The column was washed with lysis buffer and the protein was eluted with lysis buffer containing 400mM imidazole (fig. 57).

The next purification step was gel filtration. Ets1 protein was dialyzed against buffer (200 mM NaCl, 20 mMTris-HCl pH=8.0), after dialysis the protein solution was diluted one to one with the buffer containing 20 mM Tris-HCl pH=8 and 20% glycerol (in order to prevent protein aggregation), the sample was concentrated and applied on the column. The gel filtration profile for Ets1 from the construct corresponding to residues 280-441 and the standard are illustrated on the fig. 58. After gel filtration the samples were analyzed on SDS gel and the protein was found to be more than 99% pure as examined by SDS-PAGE (fig. 59).

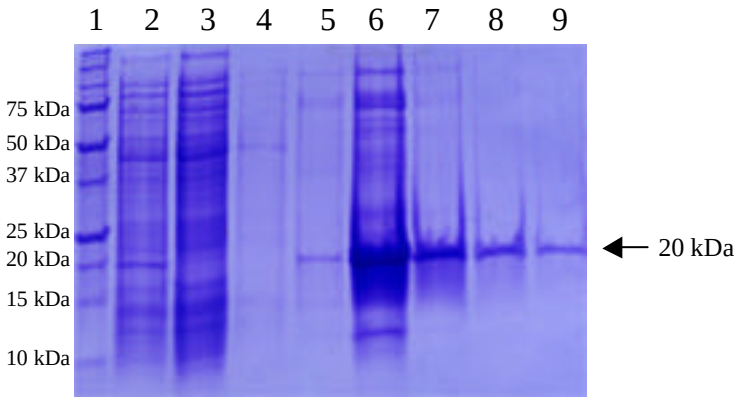


Fig. 57. **Ni-NTA affinity chromatography for Ets1 from the construct corresponding to the residues 280-441**

Lane 1 – molecular weight marker, lane 2 – before column, lane 3 – flow though, lane 4 – wash, lane 5-9 – elution fractions

A B

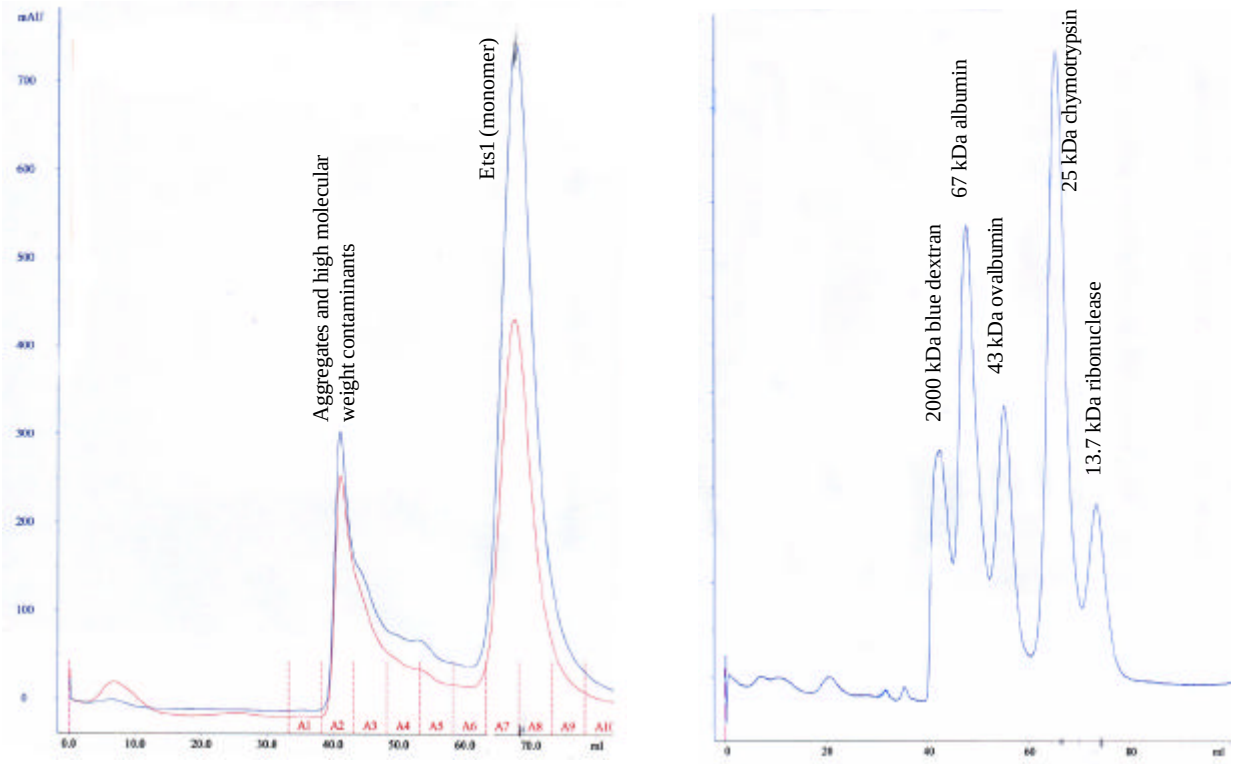


Fig. 58 A. Gel filtration of Ets1 from the construct corresponding to residues 280-441

B. Low molecular weight gel filtration standard (Amersham)

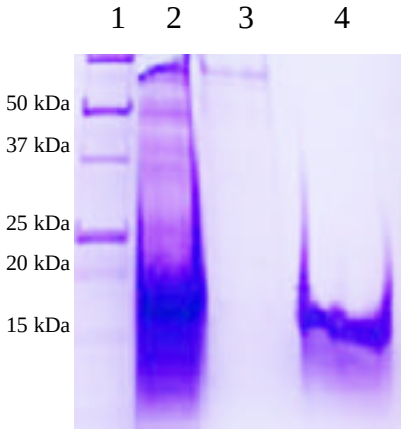


Fig. 59 **SDS gel of Ets1 (after gel filtration)**
Lane 1 – molecular weight marker, lane 7 – Ets1 280-441 before gel filtration, lane 8 - Ets 280-441 after gel filtration (peak 1), lane 9 - Ets 280-441 after gel filtration (peak2)

3.2.6 Ets1/Ets1/DNA complex formation and purification by gel filtration

The procedure for Ets1/Ets1/DNA complex formation was as followings: after Ni-NTA column, Ets1 protein was dialyzed against buffer containing 200 mM NaCl, 20 mM Tris-HCl pH=8. After dialysis the protein was diluted one to one with the buffer containing 20 mM Tris-HCl pH=8 and 20% glycerol. Glycerol was added to prevent aggregation and the protein was diluted to decrease the salt concentration, which may influence protein/DNA complex formation. DNA was added to the diluted protein and the reaction was incubated on ice for 1 hour. The complex was concentrated and applied to a gel filtration column. The following buffer was used: 20 mM Tris-HCl pH=8, 100 NaCl and 10% glycerol.

3.2.7 SDS-PAGE

Gels for SDS-PAGE were prepared according to the guidelines of Molecular Cloning. 10% and 12% separation gels were used. After electrophoresis, the gels were stained with Commassie blue or silver stating was used (silver staining kit is available from Novex).

3.2.8 Native gels

Commercially available 4-20 % gradient native gels in TBE buffer were used (Novagen). They were run in 1XTBE buffer.

3.2.9 Protein concentration

Protein samples were concentrated in an different devices. For USF1 protein and USF1/DNA complex a Centriprep device (20 ml) with membrane 10 kDa was used. For Ets1 and Ets1/Ets1/DNA complex a Vivaspin device (20 ml) with membrane 10 kDa was used. Protein concentration was performed at 4º C.

3.2.10 Protein or protein/DNA complex concentration determination

Absorbance assay 280 nm

Protein concentrations were measured according to absorbance assay at 280 nm (Layne et. al., 1957). Proteins in solution absorb ultraviolet light with absorbance maxima at 280. Amino acids with aromatic rings are the primary reason for the absorbance peak at 280 nm. Secondary, tertiary, and quaternary structure all affect absorbance, because of this reason the protein concentration was determined for denatured protein. The spectrophotometer was calibrated to zero absorbance with UV-cockail only. 10 µl of protein sample was diluted with 90 µl of UV-coctail and the absorbance at 280 nm was measured. The protein concentration was calculated according to extinction coefficient.

Bradford assay

For protein/DNA complexes the concentrations can not be measured by absorbance assay (280 nm) because DNA absorbs at this wavelength as well. Since there is no appropriate method for protein/DNA complexes concentration determination, the concentrations were determined by Bradford assay (Bradford 1976), which allows to measure protein concentration only. 1 µl of protein sample was added to 999 µl of diluted Bradford solution (Bradford solution was diluted with water four times). The reaction was left for 5 minutes, followed by measurement of

absorbance at 595 nm. Protein concentration was determined according to the graphic showing dependence of BSA absorbance on the BSA concentration. BSA protein was used as a standart for protein concentration determination. Set of different BSA concentrations was measured in order to draw the reference graphic (1 mg/ml, 2.5 mg/ml, 5mg/ml, 7.5 mg/ml, 10 mg/ml, 15 mg/ml).

3.2.11 Fluorescence resonance energy transfer (FRET)

The long fragment of DNA containing two USF1 binding sites was amplified via PCR using the primers listed above.

The principle of FRET is explained in Appendix. For the FRET experiment %E (efficiency of energy transfer) was planned to be measured by the approach based on the decrease in the fluorescence intensity of the donor (the first approach explained in Appendix). Emission spectra for buffer, donor labeled dsDNA, acceptor labeled dsDNA, mixture of donor and acceptor labeled dsDNA (to make sure that FRET does not happen) were measured. Then USF1 was titrated to the mixture of donor and acceptor labeled dsDNA. As a negative controls buffer titration and BSA titration were used.

3.2.12 Rotary Shadowing Electron Microscopy

The DNA was amplified via PCR using primers listed above. After PCR DNA was purified using PCR purification kit (QIAGEN), concentrated by ethanol precipitation and solubilised. USF1 and DNA were mixed in equimolar ratio and the complexes were analyzed on the native gel and stained with Ethidium bromide.

The basic technique of metal shadowing involves: preparing a protein solution in high concentration of glycerol; spraying the solution onto mica; evaporating the solvent under vacuum; rotary shadowing by electron bombardment of tantalum/tungsten; cabon coating the shadowed preparation. A basic description of metal shadowing is given in ref. Sommerville 1987.

List of references

1. Ayer, D. E., Kretzner, L. & Eisenman, R. N. (1993). Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell* 72, 211-22.
2. Baillat, D., Begue, A., Stehelin, D. & Aumercier, M. (2002). ETS-1 transcription factor binds cooperatively to the palindromic head to head ETS-binding sites of the stromelysin-1 promoter by counteracting autoinhibition. *J Biol Chem* 277, 29386-98.
3. Beckmann, H., Su, L. K. & Kadesch, T. (1990). TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer muE3 motif. *Genes Dev* 4, 167-79.
4. Blackwood, E. M. & Eisenman, R. N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* 251, 1211-7.
5. Boulukos, K. E., Pognonec, P., Rabault, B., Begue, A. & Ghysdael, J. (1989). Definition of an Ets1 protein domain required for nuclear localization in cells and DNA-binding activity in vitro. *Mol Cell Biol* 9, 5718-21.
6. Bungert, J., Kober, I., During, F. & Seifart, K. H. (1992). Transcription factor eUSF is an essential component of isolated transcription complexes on the duck histone H5 gene and it mediates the interaction of TFIID with a TATA-deficient promoter. *J Mol Biol* 223, 885-98.
7. Cai, M. & Davis, R. W. (1990). Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. *Cell* 61, 437-46.
8. Carr, C. S. & Sharp, P. A. (1990). A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. *Mol Cell Biol* 10, 4384-8.
9. Carthew, R. W., Chodosh, L. A. & Sharp, P. A. (1985). An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. *Cell* 43, 439-48.
10. Du, H., Roy, A. L. & Roeder, R. G. (1993). Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the Ad-ML promoters. *Embo J* 12, 501-11.
11. Duh, E. J., Maury, W. J., Folks, T. M., Fauci, A. S. & Rabson, A. B. (1989). Tumor necrosis factor alpha activates human immunodeficiency virus type 1 through induction of nuclear factor binding to the NF-kappa B sites in the long terminal repeat. *Proc Natl Acad Sci U S A* 86, 5974-8.
12. Ferre-D'Amare, A. R., Pognonec, P., Roeder, R. G. & Burley, S. K. (1994). Structure and function of the b/HLH/Z domain of USF. *Embo J* 13, 180-9.
13. Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 363, 38-45.
14. Flannery, C. R., Lark, M. W. & Sandy, J. D. (1992). Identification of a stromelysin cleavage site within the interglobular domain of human aggrecan. Evidence for proteolysis at this site in vivo in human articular cartilage. *J Biol Chem* 267, 1008-14.
15. Garvie, C. W., Pufall, M. A., Graves, B. J. & Wolberger, C. (2002). Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. *J Biol Chem* 277, 45529-36.

16. Ghysdael, J., Gegonne, A., Pognonec, P., Dernis, D., Leprince, D. & Stehelin, D. (1986). Identification and preferential expression in thymic and bursal lymphocytes of a c-ets oncogene-encoded Mr 54,000 cytoplasmic protein. *Proc Natl Acad Sci U S A* 83, 1714-8.
17. Goueli, B. S. & Janknecht, R. (2003). Regulation of telomerase reverse transcriptase gene activity by upstream stimulatory factor. *Oncogene* 22, 8042-7.
18. Graves, B. J. & Petersen, J. M. (1998). Specificity within the ets family of transcription factors. *Adv Cancer Res* 75, 1-55.
19. Gregor, P. D., Sawadogo, M. & Roeder, R. G. (1990). The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. *Genes Dev* 4, 1730-40.
20. Hu, Y. F., Luscher, B., Admon, A., Mermod, N. & Tjian, R. (1990). Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. *Genes Dev* 4, 1741-52.
21. Israel, N., Hazan, U., Alcami, J., Munier, A., Arenzana-Seisdedos, F., Bachelerie, F., Israel, A. & Virelizier, J. L. (1989). Tumor necrosis factor stimulates transcription of HIV-1 in human T lymphocytes, independently and synergistically with mitogens. *J Immunol* 143, 3956-60.
22. Kim, J. Y., Gonzalez-Scarano, F., Zeichner, S. L. & Alwine, J. C. (1993). Replication of type 1 human immunodeficiency viruses containing linker substitution mutations in the -201 to -130 region of the long terminal repeat. *J Virol* 67, 1658-62.
23. Kodandapani, R., Pio, F., Ni, C. Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R. A. & Ely, K. R. (1996). A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature* 380, 456-60.
24. Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). MASSHA - a graphic system for rigid body modelling of macromolecular complexes against solution scattering data. *J. Appl. Crystallogr.* 34, 527-532.
25. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36, 1277-1282.
26. Lee, G. M., Donaldson, L. W., Pufall, M. A., Kang, H.-S., Pot, I., Graves, B. J., Mcintosh, L. P. (2005). The structural and dynamics Basis of Ets-1 DNA binding autoinhibition. *J. Biol. Chem.* 280, 7088-7099.
27. Liu, Z., Ivanoff, A. & Klominek, J. (2001). Expression and activity of matrix metalloproteases in human malignant mesothelioma cell lines. *Int J Cancer* 91, 638-43.
28. Malemud, C. J. & Goldberg, V. M. (1999). Future directions for research and treatment of osteoarthritis. *Front Biosci* 4, D762-71.
29. Mauviel, A. (1993). Cytokine regulation of metalloproteinase gene expression. *J Cell Biochem* 53, 288-95.
30. Miyamoto, N. G., Moncollin, V., Egly, J. M. & Chambon, P. (1985). Specific interaction between a transcription factor and the upstream element of the adenovirus-2 major late promoter. *Embo J* 4, 3563-70.
31. Murre, C., McCaw, P. S. & Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56, 777-83.

32. Nair, S. K. & Burley, S. K. (2003). X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* 112, 193-205.
33. Nelson, A. R., Fingleton, B., Rothenberg, M. L. & Matrisian, L. M. (2000). Matrix metalloproteinases: biologic activity and clinical implications. *J Clin Oncol* 18, 1135-49.
34. Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D. & Graves, B. J. (1992). Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. *Genes Dev* 6, 975-90.
35. Osborn, L., Kunkel, S. & Nabel, G. J. (1989). Tumor necrosis factor alpha and interleukin 1 stimulate the human immunodeficiency virus enhancer by activation of the nuclear factor kappa B. *Proc Natl Acad Sci U S A* 86, 2336-40.
36. Perkins, N. D., Edwards, N. L., Duckett, C. S., Agranoff, A. B., Schmid, R. M. & Nabel, G. J. (1993). A cooperative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. *Embo J* 12, 3551-8.
37. Petoukhov, M. V. & Svergun, D. I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89, 1237-50.
38. Pognonec, P. & Roeder, R. G. (1991). Recombinant 43-kDa USF binds to DNA and activates transcription in a manner indistinguishable from that of natural 43/44-kDa USF. *Mol Cell Biol* 11, 5125-36.
39. Rekdal, C., Sjottem, E. & Johansen, T. (2000). The nuclear factor SPBP contains different functional domains and stimulates the activity of various transcriptional activators. *J Biol Chem* 275, 40288-300.
40. Sawadogo, M. & Roeder, R. G. (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* 43, 165-75.
41. Sawadogo, M., Van Dyke, M. W., Gregor, P. D. & Roeder, R. G. (1988). Multiple forms of the human gene-specific transcription factor USF. I. Complete purification and identification of USF from HeLa cell nuclei. *J Biol Chem* 263, 11985-93.
42. Seth, A., Robinson, L., Thompson, D. M., Watson, D. K. & Papas, T. S. (1993). Transactivation of GATA-1 promoter with ETS1, ETS2 and ERGB/Hu-FLI-1 proteins: stabilization of the ETS1 protein binding on GATA-1 promoter sequences by monoclonal antibody. *Oncogene* 8, 1783-90.
43. Sha, M., Ferre-D'Amare, A. R., Burley, S. K. & Goss, D. J. (1995). Anti-cooperative biphasic equilibrium binding of transcription factor upstream stimulatory factor to its cognate DNA monitored by protein fluorescence changes. *J Biol Chem* 270, 19325-9.
44. Sheridan, P. L., Sheline, C. T., Cannon, K., Voz, M. L., Pazin, M. J., Kadonaga, J. T. & Jones, K. A. (1995). Activation of the HIV-1 enhancer by the LEF-1 HMG protein on nucleosome-assembled DNA in vitro. *Genes Dev* 9, 2090-104.
45. Sieweke, M. H., Tekotte, H., Frampton, J. & Graf, T. (1996). MafB is an interaction partner and repressor of Ets-1 that inhibits erythroid differentiation. *Cell* 85, 49-60.

46. Sieweke, M. H., Tekotte, H., Jarosch, U. & Graf, T. (1998). Cooperative interaction of ets-1 with USF-1 required for HIV-1 enhancer activity in T cells. *Embo J* 17, 1728-39.
47. Skalicky, J. J., Donaldson, L. W., Petersen, J. M., Graves, B. J. & McIntosh, L. P. (1996). Structural coupling of the inhibitory regions flanking the ETS domain of murine Ets-1. *Protein Sci* 5, 296-309.
48. Sternlicht, M. D., Bissell, M. J. & Werb, Z. (2000). The matrix metalloproteinase stromelysin-1 acts as a natural mammary tumor promoter. *Oncogene* 19, 1102-13.
49. Sternlicht, M. D., Lochter, A., Sympson, C. J., Huey, B., Rougier, J. P., Gray, J. W., Pinkel, D., Bissell, M. J. & Werb, Z. (1999). The stromal proteinase MMP3/stromelysin-1 promotes mammary carcinogenesis. *Cell* 98, 137-46.
50. Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* 28, 768-773.
51. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80, 2946-53.
52. Venanzoni, M. C., Robinson, L. R., Hodge, D. R., Kola, I. & Seth, A. (1996). ETS1 and ETS2 in p53 regulation: spatial separation of ETS binding sites (EBS) modulate protein: DNA interaction. *Oncogene* 12, 1199-1204.
53. Wasylyk, C., Gutman, A., Nicholson, R. & Wasylyk, B. (1991). The c-Ets oncoprotein activates the stromelysin promoter through the same elements as several non-nuclear oncoproteins. *Embo J* 10, 1127-34.
54. Wasylyk, C., Schlumberger, S. E., Criqui-Filipe, P. & Wasylyk, B. (2002). Sp100 interacts with ETS-1 and stimulates its transcriptional activity. *Mol Cell Biol* 22, 2687-702.
55. Watson, D. K., McWilliams, M. J., Lapis, P., Lautenberger, J. A., Schweinfest, C. W. & Papas, T. S. (1988). Mammalian ets-1 and ets-2 genes encode highly conserved proteins. *Proc Natl Acad Sci U S A* 85, 7862-6.
56. Watson, D. K., McWilliams-Smith, M. J., Nunn, M. F., Duesberg, P. H., O'Brien, S. J. & Papas, T. S. (1985). The ets sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21: both loci are transcriptionally active. *Proc Natl Acad Sci U S A* 82, 7294-8.
57. Werner, M. H., Clore, G. M., Fisher, C. L., Fisher, R. J., Trinh, L., Shiloach, J. & Gronenborn, A. M. (1997). Correction of the NMR structure of the ETS1/DNA complex. *J Biomol NMR* 10, 317-28.
58. Ye, S., Whatling, C., Watkins, H. & Henney, A. (1999). Human stromelysin gene promoter activity is modulated by transcription factor ZBP-89. *FEBS Lett* 450, 268-72.
59. Yoshiyama, Y., Asahina, M. & Hattori, T. (2000). Selective distribution of matrix metalloproteinase-3 (MMP-3) in Alzheimer's disease brain. *Acta Neuropathol (Berl)* 99, 91-5.
60. Zeichner, S. L., Kim, J. Y. & Alwine, J. C. (1991). Linker-scanning mutational analysis of the transcriptional activity of the human immunodeficiency virus type 1 long terminal repeat. *J Virol* 65, 2436-44.
61. Zervos, A. S., Gyuris, J. & Brent, R. (1993). Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. *Cell* 72, 223-32.

62. Zucker, S., Cao, J. & Chen, W. T. (2000). Critical appraisal of the use of matrix metalloproteinase inhibitors in cancer treatment. *Oncogene* 19, 6642-50.
63. Kalinin V. L. Introduction to Molecular Virology (in Russian) (2002)
64. Kalinin V. L. Transcription and regulation of gene expression (in Russian) (2001)
65. Sommerville (1987) Electron microscopy in molecular biology. A practical approach
66. Batchelor, A. H., Piper, D. E., Wolberger, C. (2002). Crystallization of protein-nucleic acid complexes. *Encyclopedia of life science (web-page)*

5. Appendix

5.1 Fluorescence resonance energy transfer

Fluorescence resonance energy transfer can be used to determine molecular distances in biological macromolecules the range of 10-100 Å. The technique is based on the theories of Förster who proposed that electronic excitation energy can be efficiently transferred through dipole-dipole interactions from a donor fluorophore to an acceptor fluorophore in a distance dependent manner (Golemis 2002).

Several factors determine how efficiently the energy is transferred. For a given donor-acceptor pair the distance at which energy transfer is 50% is called Ro. When these fluorophores are placed at specific sites in the macromolecule of interest, the distance between them is quantified by measuring the efficiency of energy transfer (%E) which then relates to the distance between the probes (R) by the following equation:

$$\%E = Ro^6/(Ro^6 + R^6)$$

So if the molecular distance was Ro, the equation shows that the %E would be 50%, which is how Ro is defined. The %E can be measured by three slightly different approaches. The first two monitor the quenching of the donor and the third monitors the enhanced fluorescence of the acceptor.

1. Decrease in the fluorescence intensity of the donor. Where F_{da} is the intensity of the donor in the presence of the acceptor and F_d is the intensity of the donor in the absence of the acceptor. ($\%E = 1- F_{da}/F_a$)
2. Decrease in the excited state lifetime of the donor. Where t_{da} is the lifetime of the donor in the presence of the acceptor and td is the lifetime of the donor in the absence of the acceptor. $\%E = 1- t_{da}/t_d$
3. An increase in the fluorescence of the acceptor, if the acceptor happens to be fluorescent. (A fluorescent acceptor is not necessarily a requirement for energy transfer; the acceptor must only be able to absorb the light given off by the donor. There are several different, and rather complex formulations for quantifying %E by this method).

5.2 Small-angle X-ray scattering

Small angle X-ray scattering is used for the analysis of low resolution structure. Given that the resolution of the method is lower than the X-ray wavelength, it is sufficient to record the scattering of small (low degrees) angles. The most useful feature of the method is the possibility to analyze systems, which are not ordered.

Imagine a protein as an assemble of a number of nuclei nd electrons. Both nucleus and electrons can be treated as small scattering centers. Scattering ability of such an object when it is exposed to X-rays can be characterized by scattering density $\varphi(\mathbf{r})$, which is at the same time electron distribution density (fig. 60).

The wave interacts with all scattering centers, which are becoming sources of spherical waves. The superposition of these waves is the first approximation to the real scattering. The scattered wave is scattered further on the scattering centers, which gives the second approximation and so on. When the interaction between the scattering wave and the object is not very strong, the approximations are going to the certain resulting wave. If the interaction is weak, only the first approximation can be taken into account (Svergun, Feigin, 1986).

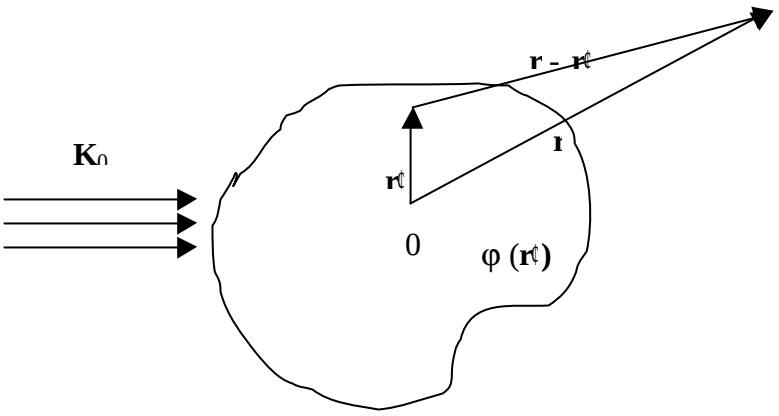


Fig. 60 **Scattering of the object with scattering density j (r)**

The resulting scattered wave is a solution of the wave equation (1),

$$[\Delta + k^2_0 + v\,\varphi(\mathbf{r})]\,\psi(\mathbf{r}) = 0 \quad (1)$$

where $\psi(\mathbf{r})$ is the wave scattered by the field $\varphi(\mathbf{r})$, k_0 is a wave number, Δ is Laplacian, v is a parameter which characterize the interaction intensity of the wave with the matter.

The solution (2) is the sum of the incoming wave and the first approximation to the resulting scattered wave (when the others can be neglected).

$$\varphi^0(\mathbf{r}) + \varphi^1(\mathbf{r}) = A_0 \exp(i\mathbf{k_0 r}) + (A_0\, v \exp(i\mathbf{k_0 r})) / 4\pi r \int \varphi(\mathbf{r'}) \exp(i\mathbf{s r'})\, d\mathbf{r'}, \quad (2)$$

where $\mathbf{s} = \mathbf{k} - \mathbf{k_0}$ is a scattering vector ($|\mathbf{s}| = 4\pi \sin\theta / \lambda$, 2θ is scattering angle).

$f(\mathbf{s})$ can be introduced (3) which is the amplitude of the scattering on the field $\varphi(\mathbf{r})$.

$$f(\mathbf{s}) = v / 4\pi\, r \int \varphi(\mathbf{r}) \exp(i\mathbf{s r})\, d\mathbf{r} \quad (3)$$

The equation (3) is a Fourier integral and $\varphi(\mathbf{r})$ can be written as in (4)

$$\varphi(\mathbf{r}) = 1 / 2\pi^2\, v \int f(\mathbf{s}) \exp(-i\mathbf{s r})\, d\mathbf{s} \quad (4)$$

Thus, Fourier transformation is the basis for the calculation of the amplitudes from the field $\varphi(\mathbf{r})$ and vice versa. Fourier transformation connects the function in the real space ($\mathbf{r}$ – space of the object) with the function in the reciprocal space ($\mathbf{s}$ – space of the reflection).

Another important definition connected to Fourier transformation function convolution, which finds an important application in the diffraction theory. Fourier integral from the scattering intensity gives a convolution of $\varphi(\mathbf{r})$ with inself inverted in the coordinate start point. This function is called autocorrelation function $P(\mathbf{r})$ (5)

$$P(\mathbf{r}) = \int I(\mathbf{s}) \exp(-2\pi i\mathbf{s r})\, d\mathbf{s},$$

$$I(\mathbf{s}) = \int P(\mathbf{r}) \exp(2\pi i\, \mathbf{s r})\, d\mathbf{r} \quad (5)$$

In case of X-ray scattering of atoms, the scattering is mostly due to electrons. Whereas the scattering of the nuclei is negligible. The object is not ordered and in a SAXS experiment the average intensity is measured. That's why it is necessary to use a statistical function of atom distribution in the object.

Let us assume that the system contains N identical molecules with scattering amplitude (form-factor) f(**s**).

$$I(\mathbf{s}) = \sum\sum f_i(\mathbf{s})\, f_j(\mathbf{s}) \exp[i\mathbf{s}(\mathbf{r_i} - \mathbf{r_j})] \quad (6),$$

where $\mathbf{r_i}$ and $\mathbf{r_j}$ are coordinates of the molecules.
The average intensity when $s>2\pi/D$ where $D\approx V^{1/3}$ (V – irradiated sample volume) is (7)

$$\langle I(\mathbf{s})\rangle = \langle N\rangle \{\langle f^2(\mathbf{s})\rangle - \langle f(\mathbf{s})\rangle^2/v_1 \int [1- P(r)] (\sin sr)/sr\, 4\pi r^2\, dr\} \quad (7)$$

If $\langle f^2(\mathbf{s})\rangle = \langle f(\mathbf{s})\rangle^2$, then (7) can be written

$$\langle I(\mathbf{s})\rangle = \langle N\rangle F^2(\mathbf{s}) \{1 - 1/v_1 \int [1- P(r)] (\sin sr)/sr\, 4\pi r^2\, dr\} \quad (8),$$

where $v_1 = V/\langle N\rangle$, F(s) is the average form-factor.

The equation means that when $F^2(s)$ is known one can find P(r) and vice versa. In case of the system containing different types of molecules $F^2(s)$ the scattering is given by the average of form-factors of different types of molecules.

In the above considerations, the scattering properties of the buffer have not been taken into account. If ρ_s is scattering density of the buffer and the scattering density of each molecule is $\rho(\mathbf{r})$ than

$$g(\mathbf{r}) = \rho(\mathbf{r}) - \rho_s \quad (9),$$

where $g(\mathbf{r})$ is effective density, which will be used below to represent of the object.

In the case of a diluted sample the average intensity of the molecule is proportional to the average intensity of the sample.

Important characteristics

One of the most important particle parameters, which can be directly computed from the scattering data, is radius of gyration with respect to its center of electron density.

Radius of gyration is

$$R_g^2 = {}_v\int \rho(\mathbf{r})\, r^2\, d\mathbf{r} \,/\, {}_v\int \rho(\mathbf{r})\, d\mathbf{r} \quad (10)$$

The value of R_g can be calculated using Guinier law, which is valid for the very beginning of the curve I(s) ($s<1/R_g$).

$$I(s) = I(0) \exp(-s^2 R_g^2 / 3) \quad (11)$$

The maximum radius is calculated from the correlation function, which is average selfconvolution of the scattering density. $\gamma(r)$ is correlation function. R_{max} is a maximum distance within the particle.

$$\gamma(r) = 1 / 2\pi^2 \, {}_0\int^{\infty} I(s) (\sin sr) / sr\, s^2\, ds \quad (12)$$

$$\gamma(r) \equiv 0, \text{ when } r > R_{max} \text{ and } R_{max} = 2 / \gamma(0) \, {}_0\int^{D} \gamma(r)\, dr \quad (13)$$

SAXS programs

CRYSOL

CRYSOL (Svergun et. al., 1995) is a program for evaluating the solution scattering from macromolecules with known atomic structure. Given the atomic coordinates it can either predict the solution scattering curve or fit the experimental scattering curve to the solution scattering curve calculated from crystallographic model.

Program MASSHA

A program, MASSHA (Konarev et. al., 2001), is made for three-dimensional rendering and rigid-body refinement. The program allows display and manipulation of high-resolution atomic structures and low-resolution models represented as smooth envelopes or ensembles of beads.

Program PRIMUS

A program suite for one-dimensional small-angle scattering data processing. The main program, PRIMUS (Konarev et. al., 2003), has a menu-driven graphical user interface calling computational modules to perform data manipulation and analysis.

The program allows buffer extraction and data extrapolation to zero speciment concentration in the sample. It allows as well calculations of radius of gyration and Porod volume.

For computation of characteristic functions of dilute monodisperse or polydisperse systems, PRIMUS provides an interface to the indirect transformation program GNOM (Semenyuk & Svergun, 1991; Svergun, 1992).

Program GASBOR

Program GASOR (Svergun et. al., 2001) presents an *ab initio* method for building structural models of proteins from x-ray solution scattering data.

Proteins typically consist of folded polypeptide chains composed of amino acid residues separated by ~0.38 nm between adjacent C_α atoms in the primary sequence. At a resolution of 0.5 nm, a protein structure can be considered as an assembly of dummy residues (DR) centered at the C_α positions. A three-dimensional model of the protein may therefore be constructed from solution scattering data by finding a chain-compatible spatial arrangement of the DRs that fits the experimental scattering pattern. That such a model adequately describes scattering patterns of proteins was verified by simulations.

5.3 Protein crystallization

The properly purified and prepared protein samples are the first prerequisite for obtaining crystals suitable for X-ray analysis. The sample has to be pure. The usual concentration of the sample for handling is from 10 to 30 mg/ml. Initial protein crystallization is mainly a trial-and-error procedure in which the protein is slowly precipitated from its solution. The classical procedure for inducing proteins to separate from solution and produce a solid phase is to gradually increase the level of saturation by addition of a precipitant. Very often the protein separates as a precipitate, but with appropriate care protein crystals can be grown. The precipitation point or solubility minima are usually dependent on the pH, temperature, the chemical composition of the precipitant, and the properties of both the protein and the solvent (McPherson, 1990).

A common precipitation method involves increasing the effective concentration of the protein, usually by adding a salt (salting-out) or polyethilenglycol (PEG), or by taking away a salt (salting-in). The salt ions and macromolecules compete for the solvent molecules (water) because both, salt ions and protein molecules, require hydration layers to maintain their solubility. When competition between them becomes intense, the protein molecules begin to self associate (Fig. 61).

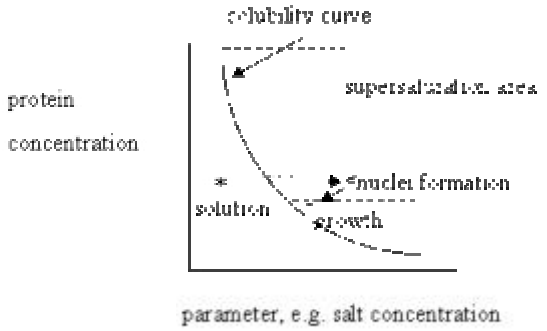


Fig. 61 **A solubility curve for a protein, as a function of the salt concentration or other parameter**

The other common precipitation methods involve changes in pH and temperature at constant ionic strength. Proteins may exhibit a number of different solubility minima as a function of the variables, and each of this minima may afford the opportunity for crystal formation. In practice it is possible to vary pH and

temperature and keep the constant concentration of precipitating solution, or to increase concentration of precipitating agent at constant pH and temperature. The precipitating agent may be a salt such as ammonium sulphate, an organic solvent such as ethanol or methylpentanediol, or a highly soluble synthetic polymer such as polyethylene glycol (PEG).

The most commonly used methods for initial crystal trials are: the hanging drop and sitting drop vapour-diffusion methods, dialysis and batch method. The hanging and sitting drop methods relies on the transport of either water or some volatile agent between a micro-drop of mother liquor and much larger reservoir solution.

In hanging drop method drops are prepared on a siliconized microscope glass cover slip by mixing 1~3 µL of protein solution with the same volume of precipitant solution (Drenth, p. 5). The cover slip is placed over a small well containing 1 ml of the precipitating solution (Fig. 62).

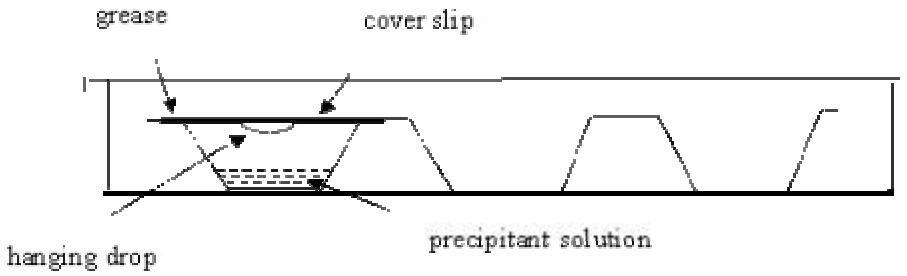


Fig. 62 **A schematic drawing of the hanging drop method for protein crystallisation.**

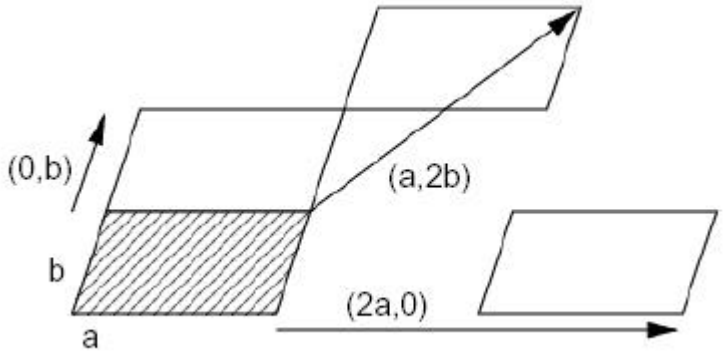
In the sitting drop method the samples are placed in transparent containers, which hold in additions 1 ml of the precipitating solution.

5.4 Principles of X-ray crystallography

A crystal arranges huge numbers of molecules in the same orientation, so that scattered waves can add up in phase and raise the signal to a measurable level. A crystal consists of an infinite number of copies of one object. The objects must form a regular pattern, the “crystal lattice”.

Unit cell and asymmetric unit

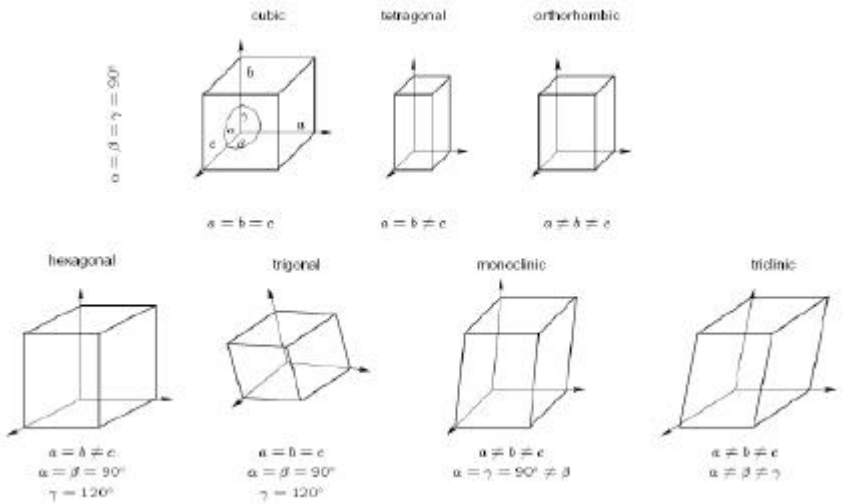
In every crystal there is always a smallest box, defined by its edges a, b, and c, and the angles they enclose, a, ß, and ? , from which one can create the whole crystal solely by integer translations along its sides. This is called the unit cell.



A unit cell consists itself of a smallest unit from which one can create the unit cell by applying all symmetry operators that belong to the crystals space group. This is called the asymmetric unit.

Seven Lattice Types

The lattice is an infinite repetition of one “box”, the unit cell. It is defined by the lengths of three edges a, b, c, and the angles a, ß, and ?. There are 7 different lattice types that allow to fill an infinite space.



Space Groups

In a three dimensional lattice, symmetry operations cannot be combined arbitrarily. Together with translations, there are 230 allowed combinations, the 230 space groups. Of those, only 65 are chiral, i.e. suitable for macromolecules like proteins, RNA, or DNA.

Matthews coefficient

The Matthews coefficient allows to estimate this number. The Matthews coefficient (Vm) is easily calculated as

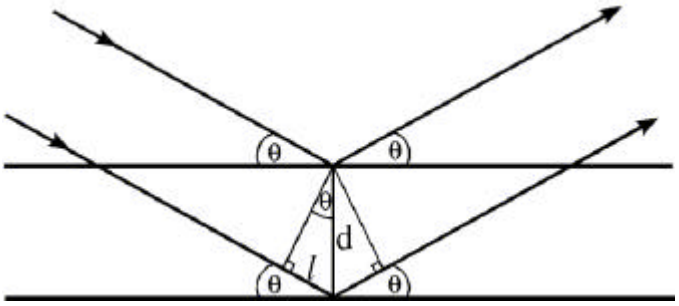
$$\frac{\text{volume of your unit cell}}{\text{the molecular weight of your macromolecule} * Z * X}$$

Where Z is the number of asymmetric units in the unit cell (i.e. the number of symmetry operators in the space group). The unknown variable, X, is the number of molecules in the asymmetric unit. A series of Matthews coefficients with X= 0.5, 1.0, 2.0, etc are calculated. The most probable values of X are those, which give Matthews coefficients within the empirically observed range (Matthews, 1968).

Theory of X-ray diffraction by crystal

The electrons in a crystal are emitters of waves. When the emitted waves emit add up, they interfere with one another. The total path does not depend on the direction of the incoming and outgoing waves and the positions of the electrons relative to each other. A crystal amplifies the diffraction pattern in certain directions, where the various unit cells diffract in phase, and eliminates it in other directions. The relationship between scattering angle and the interplanar spacing is given by Bragg's law:

$$2d \sin\theta = n\lambda$$



Ewald sphere with radius $1/\lambda$ is a geometrical construction to help visualize which Bragg planes are in the correct orientation to diffract (fig. 63). The vectors of the incoming ray (labelled 1) and the diffracted ray (labelled 2) are both at an angle θ from a set of Bragg planes in the crystal. The diffracted ray has its base at the center of the sphere (the origin of the crystal). The vector of the difference (shown in red) between the direct beam passing undeflected through the crystal (labelled 3) and the diffracted ray is perpendicular to the Bragg planes. In the small internal triangles, each side corresponding to the halve of the difference vector of $\sin\theta/\lambda$ length is equal to 1/2d. Such a construction could be made for any set of planes for incoming ray, and the corresponding reciprocal space vector would be seen to go from the position of the undeflected direct beam to the tip of the vector representing the diffracted ray. All of these reciprocal space vectors start from the same point, which is defined as the origin of reciprocal space.

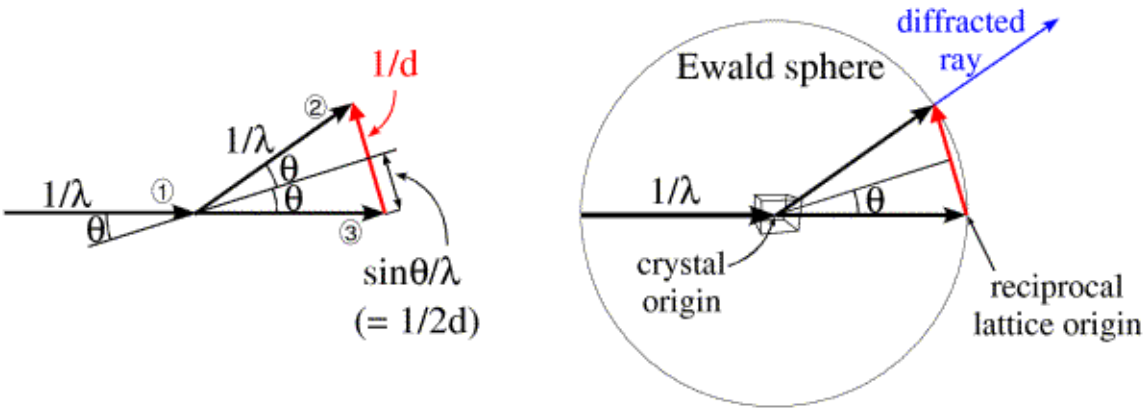


Fig. 63 **Ewald sphere construction**

If a set of planes is in the diffracting condition, the corresponding reciprocal space vector has to end on the surface of the Ewald sphere. Conversely, if the direct beam does not strike the planes with the correct angle θ, the reciprocal space vector will not be on the surface of the Ewald sphere.

The intensities of the reflections (h, k, l) measured by an X-ray diffraction experiment are proportional to the square modulus of the structure factors (F(h, k, l)).

The structure factors are related to the electron density distribution within the unit cell by an expression called Fourier transformation:

$$\rho(x,y,z) = \frac{1}{V_{unitcell}} \sum_{h,k,l=-\infty}^{h,k,l=\infty} F(h,k,l) \cdot e^{-2\pi i(hx+ky+lz)}$$

and its inversion

$$F(h,k,l) = \int_{V_{unitcell}} d^3x \, \rho(x,y,z) e^{2\pi i(hx+ky+lz)}$$

If all structure factors are known, one could calculate the electron density in the whole unit cell. A major effort of crystallography lies in the determination of as many and as accurate structure factors as possible.

Phase problem

The phase problem arises because we need to know both the amplitude and the phase of the diffracted waves to compute the inverse Fourier transform. During the experiment the intensities of each reflection is measured, but there is no practical way of measuring the relative phase angles for the different reflections, and the phase has been lost.

There are different technics to solve the phase problem in protein X-ray crystallography: the isomorphous replacement method, the multiple wavelength anomalous diffraction method, the molecular replacement method and the direct methods (Drenth 1999).

The multiple isomorphous replacement method (MIR) requires the crystal of the protein containing one or more heavy atom derivatives of the protein. This method uses the differences observed in the diffraction intensities, upon incorporating heavy atoms into the crystals. The first step of this method requires the determination of the coordinates of these heavy atoms in the unit cell. The positions and occupancies of heavy atoms are the starting point for the determination of the protein phase angles (Drenth, 1999).

The multiple wavelength anomalous diffraction method (MAD) is based on the idea of using the anomalous scattering of protein crystal containing heavy atom derivatives at different wavelengths. The idea of separating out the normal scattering of the atoms from any anomalous scattering is used.

The molecular replacement method (MR) requires the existence of the homologous protein structure. The first identification of the suitable model (known structure) can be based on sequence homology with the protein for which the structure must be determined. To find a correct orientation of the model molecule in the target unit cell two steps (rotation and translation) are used. In the rotation step the spatial

orientation of the known molecule in the unknown unit cell is determined, while in the translation step the correctly oriented molecule is positioned with respect to the symmetry elements in the unknown cell.

Electron density maps, model bilding

To build the model and fit into the map an interactive computer graphics program, such as the program named 'O' or Coot are used.

Refinement

The success of an atomic model is often judged through the standard crystallographic R-factor, which is simply the average fractional error in the calculated amplitude compared to the observed amplitude of the structure factors (Drenth, 1999). Another important parameter is free R-factor.

R-factor:

"residual-factor" or agreement factor:

$$R-factor=\frac{\sum_{(h,k,l)} \| F_{obs}(h,k,l)| - |F_{calc}(h,k,l)\|}{\sum_{(h,k,l)} |F_{obs}(h,k,l)|}$$

Free R-factor:

an R-factor calculated on a partial data set (5%) that is not used in the refinement of a structure.

Refinement is the process of adjusting the model to find a closer agreement between the calculated and observed structure factors by least-squares methods or molecular dynamics. The method of least squares is an iterative process in which the parameters to be refined change. When the changes in the parameters become small enough the refinement has then converged to the final parameter set. The radius of convergence is the maximum distance, which can be achieved by the atoms to move to their ideal positions and is limited by the resolution of data employed.

There are two types of crystallographic refinement: rigid body and restrained refinement. In **rigid body refinement**, big parts of the protein, such as subunits, are

refined as rigid bodies. In the simplest case, the entire protein is treated as one rigid body. Rigid body refinement is useful in the early stages of structure determination and it is usually done with low resolution data (15-3Å). In **restrained refinement**, the stereochemical parameters are allowed to vary around a standard value, controlled by an energy term. The atomic coordinates are the variables and the restraints are on the bond length, bond angles, torsion angles and van der Waals contacts. This allows an easy movement of small parts of the structure, but it is difficult to move large parts, for instance, an entire molecule or domain (Drenth, 1999).

List of references (for Appendixes)

1. Svergun D. I., Feifin L. A. (1986) X-ray and neutron small angle scattering (in Russian)
2. McPherson (1990) Crystallization of biological macromolecules. Gold Spring Harbor Press
3. Drenth J. (1999) Principles of protein X-ray crystallography (2nd edition). Heidelberg: Springer-Verlag
4. McRee D. E. (1999) Practical protein crystallography (2nd edition)
5. E. Golemis (2002) Protein-protein interaction. A molecular cloning manual
6. Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* 28, 768-773.
7. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80, 2946-53.
8. Petoukhov, M. V. & Svergun, D. I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89, 1237-50.
9. Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). MASSHA - a graphic system for rigid body modelling of macromolecular complexes against solution scattering data. *J. Appl. Crystallogr.* 34, 527-532.
10. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36, 1277-1282.
11. Matthew B. W. (1968). J. Mol. Biol. 33, 491-97

Acknowlengements

I would like to thank my supervisor Matthias Wilmanns for giving me the opportunity to work on these projects.

I would like to say many, many thanks to Galina Kachalova (Max-Plank Institute, Hamburg) for giving me wonderful advices, for helpful scientific discussions and for being very supportive in all the moments I needed.

I would like to thank Dimitry Svergun, Peter Konarev and Maxim Petoukhov for the help with SAXS experiments.

I would like to thank Michel Koch for the support in preparation of the thesis and for very interesting and helpful discussions.

I would like to thank Gleb Bourenkov (Max-Plank Institute, Hamburg) and Sasha Popov for collecting datasets for me in Grenoble.

I would like to thank Matthew Groves for giving me good advices and for the help in preparation of the thesis.

I would like to thank Stiv for the assistance in preparation of the thesis.

I would like to thank Andreas Hoenger and Dilem Hizlan (EMBL, Heidelberg) who helped me to perform EM experiments.

I would like to thank Philip Bastiaens (EMBL, Heidelberg) and Rosie Jordanova who helped me with FRET experiment.

I would like to thank Joerg Mueller (EMBL, Heidelberg), a member of my thesis advisory committee, for giving me good advices.

I would like to thank Fransisco Fernandez for providing me expression constructs for USF1 and Ets1.

I would like to thank Ulrich Mayer and Michael Sieweke for providing me the vector containing HIV1 LTR region and for intereseting discussions.

I would like to thank Robert Janowski and Ela Nowak for being very kind and helpful colleagues. Thank you for your help.

I would like to thank Yong-Hwa Song and Peijian Zou for help in the lab.

And, finally, I would like to thank all the members of EMBL for great time, which I had in Hamburg.