3. Results and Discussion

3.1 Ets-1 cloning, expression, purification and initial crystallisation

3.1.1 Introduction

Several protocols have been described to produce and purify Ets-1 for structural and biochemical purposes, (1) Ets-1 ETS domain and Ets-1 ETS domain plus the C-terminal autoinhibitory (AI) motif were produced for NMR studies by recombinant expression in *E coli* BL21(DE3) in a pET3a vector as insoluble inclusion bodies, refolded and purified by anion and cation exchange chromatography (Petersen et al, 1995; Werner et al, 1995); (2) the minimal domain was expressed in *E coli* BL21(DE3) in the soluble fraction, followed by its purification by two successive steps of anion exchange chromatography, to prepare Ets-1 ETS DBD for crystallisation (Garvie et al, 2001); (3) diverse protocols for small-scale production in *E coli* from inclusion bodies or soluble fraction (Lionneton et al, 2001) and soluble GST-fusion protein (Kim et al, 1999) have been devised.

Nevertheless, expression of longer Ets-1 constructs (beyond Ets-1 DBD) for structural biology, a prerequisite for the structural elucidation of protein-protein complexes involving the Ets-1 N-terminal AI motif, have not been described to date and, thus, the results presented herein constitute the first such procedures.

In the next sections the molecular cloning, protein production, purification (from soluble and insoluble sources), purification, complex formation, stabilisation, characterisation and first crystallisation trials of several autoinhibited Ets-1 clones will be described.

3.1.2 Results

3.1.2.1 Molecular cloning of the Ets-1 constructs

An array of Ets-1 deletion mutants was cloned into several expression vectors for expression and stability/toxicity tests (2.3.9; 2.3.10). Out of this initial array, the constructs that were not cytotoxic to *E coli* and expressed recombinant protein at concentrations of or above 1 mg/ml were selected for further experiments (see Table 2.4 and Table 3.1). In particular, three of these constructs (E244ΔS,

E280ΔS and E301ΔS) were cloned into pETM30 (see 2.2.11) and a third construct harbouring the minimal ETS domain (E335ΔS) had been cloned into pProExHTb (2.2.11).

Construct	Vector/Tag	Ets-1[a]	MW/pI[b]
pETM30-E244ΔS	pETM30/His6-GST	Ets-1 244-441	22.67/6.81
pETM30-E280ΔS	pETM30/His6-GST	Ets-1 280-441	18.67/8.65
pETM30-E301ΔS	pETM30/His6-GST	Ets-1 301-441	16.34/9.60
pProExHTb-E335ΔS	pProExHTb/His6	Ets-1 335-441	12.76/9.26

Table 3.1 Ets-1 expression clones. Optimised expression clones for Ets-1 E244ΔS, E280ΔS, E301ΔS and E335 constructs. All constructs rendered the highest yield in soluble protein as His6-GST fusions, with the exception of E335ΔS that developed insoluble inclusion bodies unless shuffled to a His6-tagged form on pProExHTb. [a] All Ets-1 constructs were all-serine mutants, ie, every cysteine in wild-type Ets-1 was mutated into a serine to avoid oxidation (3.1.2); [b] relative molecular weight (MW) and isoelectric point (pI) for each construct, after removal of the tags by proteolysis, calcuated with the EMBOSS suite (Rice et al, 2000).

3.1.2.2 Site-directed mutagenesis of all cysteines of wild-type Ets-1 into serines

Initial attempts at producing wild-type Ets-1 in a soluble form failed because of immediate and irreversible oxidation of Ets-1 constructs as soon as *E coli* cells were broken open. Addition of up to 100 mM DTT or 20 mM βME to the lysis buffer did not improve solubility. The oxidised protein was lodged into macroscopic precipitates by intermolecular disulphide bridge formation, as demonstrated by treatment with 100 mM TCEP (2.1.1) at RT (Fig 3.1).

Site-directed mutagenesis of all cysteines present in the Ets-1 coding region into serines abolished Ets-1 oxidation and therefore contributed largely to the stabilisation of all constructs (Fig 3.2); however, it did not have any appreciable effect on expression yield or solubility. Mutagenesis was carried out as described (3.2.10) on residues C261, C350 and C416 of the full-length Ets-1 cDNA, and from there the appropriate clones were constructed by PCR amplification (2.2.9), restriction (2.2.8) and ligation (2.2.8).

3.1.2.3 Recombinant expression of Ets-1 constructs E244ΔS, E280ΔS and E301ΔS

Of all constructs assayed for expression the His6-GST fusions produced better yields of protein under standard conditions (1 mM IPTG, 37 ºC), reaching 4-35 mg per litre culture (see 2.3.10, for expression tests and small and large-scale culture setup; see 2.3.2, for protein concentration determination). E244ΔS expressed very poorly, with at most 1-2 mg per litre of culture. Furthermore, the yields were better for smaller proteins, increasing from 15-20 mg of E280ΔS to 35-55 mg of E335ΔS per litre of culture. Solubility followed the opposite trend than yield, and thus E280ΔS and E301ΔS expressed mostly as soluble proteins while E335ΔS went entirely into

inclusion bodies. After cell lysis (2.3.10), both E280ΔS and E301ΔS proteins could be isolated in the solid state as pellets by centrifugation, but they were easily resolved by slightly raising the CHAPS concentration in the lysis buffer from 0.2% (w/v) to 1.0% (w/v); conversely, E335ΔS developed inclusion bodies which resisted all attempts to solubilisation with either native or denaturing methods - even 6 M GuHCl failed to solubilise E335ΔS inclusion bodies. In the purification steps ensuing protein resolution, E280ΔS and E301ΔS were treated as soluble proteins.

Fig 3.1 TCEP incubation of wild-type Ets-1 after refolding. Protein native electrophoresis on a 8% TBE gel of 0.1-1 mg E244 samples either mock-treated (1-3) or incubated with 100 mM TCEP for 30 min (4). E244 was partially purified from inclusion bodies and refolded by the gel filtration method (2.3.8; 2.3.10; 2.3.12) in 50 mM Tris, 250 mM NaCl, 10 mM DTT, 10% (v/v) glycerol, pH 8.0. Lanes 1-3 show E244 1, 5 and 60 min after refolding. It is noticeable the shift in mobility of E244 after TCEP treatment, which runs at the approximate molecular weight of its monomer (22.67 kDa), while as soon as 1 h after refolding the aggregation and cross-linking of the protein sample is evident. The presence of a band of apparent molecular weight 6-8 times larger than that of E244 as soon as 1 min after refolding could be interpreted as an initial aggregated state, not necessarily covalently bonded (as it appears on lane 4 as well).

3.1.2.4 Recombinant expression of Ets-1 E335ΔS construct

E335ΔS was expressed as a His6-tagged protein in *E coli*, as described before (see 2.2.11 for the cloning of E335ΔS on pProExHTb; see 2.3.10 for expression tests and small and large-scale culture setup), with yields of about 5-10 mg of soluble protein per litre of culture at 37 ºC. Depending on the preparation as much as 50% (w/w) of E335ΔS could be targeted to inclusion bodies, so protein production was carried out at 30 ºC. As a side-effect of the lower temperature expression, protein yield was approximately halved (3-6 mg per litre of culture).

3.1.2.5 Protein purification

All four Ets-1 constructs E244ΔS, E280ΔS, E301ΔS and E335ΔS, were purified by an almost identical procedure. This was warranted since they share an N-terminal His6 tag for IMAC affinity chromatography (2.3.11) - the capture step - and a similar isoelectric point (pI) (Table 3.1) - for the intermediate purification step, consisting in a cation exchange chromatography (CIEX) (2.3.11). As

a final step, all proteins purified under denaturing conditions were further purified by reverse phase chromatography (2.3.11). Fig 3.3 summarises the purification strategy devised for the Ets-1 constructs for E280ΔS.

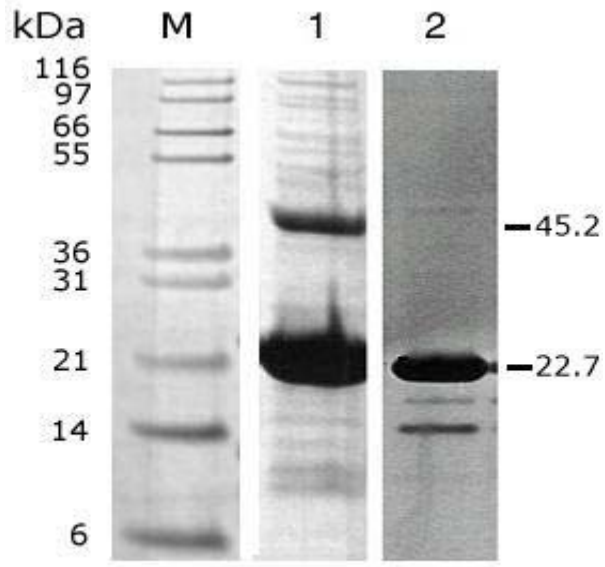


Fig 3.2 Wild-type and mutant E244. Comparison of the aggregation state of wild-type and mutant E244 expressed in pBAT4 by protein gel electrophoresis (in both cases on 10% BT gel [2.3.5]). Lane 1 shows partially purified wild-type E244 after 1 h of its refolding by the gel filtration method (2.3.12), already presenting a ladder of cross-linked products. Lane 2 shows the mutant E244ΔS (which contains mutations C261ΔS, C350ΔS, C416ΔS), which does not undergo oxidation, after 24 h. Both samples were kept at 4 ºC at all times. The lower molecular weight bands on lane 2 are contaminants, not degradation or oxidation products (they can be fully removed by additional purification, shown later).

3.1.2.6 Cell lysis and protein extraction

After expression, harvested cells were either frozen away at -80 ºC or processed immediately. In either case, the cell pellets were lysed by sonication (2.3.10) in 20 ml of Ni/NTA lysis buffer (2.1.9.2) per litre of culture. Cell debris and unbroken cells were removed by centrifugation (2.3.10), and both pellet and supernatant kept at 4 ºC for analysis by SDS PAGE electrophoresis (2.3.5).

3.1.2.7 Capture by IMAC affinity chromatography

The His6-tagged proteins were isolated, stabilised and concentrated (the capture step) by IMAC affinity chromatography (2.3.11) either in gravity flow mode (casting a suitable volume of Qiagen Ni/NTA resin on a BioRad plastic column) or by high performance liquid chromatography (HPLC) on a Pharmacia HiTrap column previously charged with nickel (2.3.11).

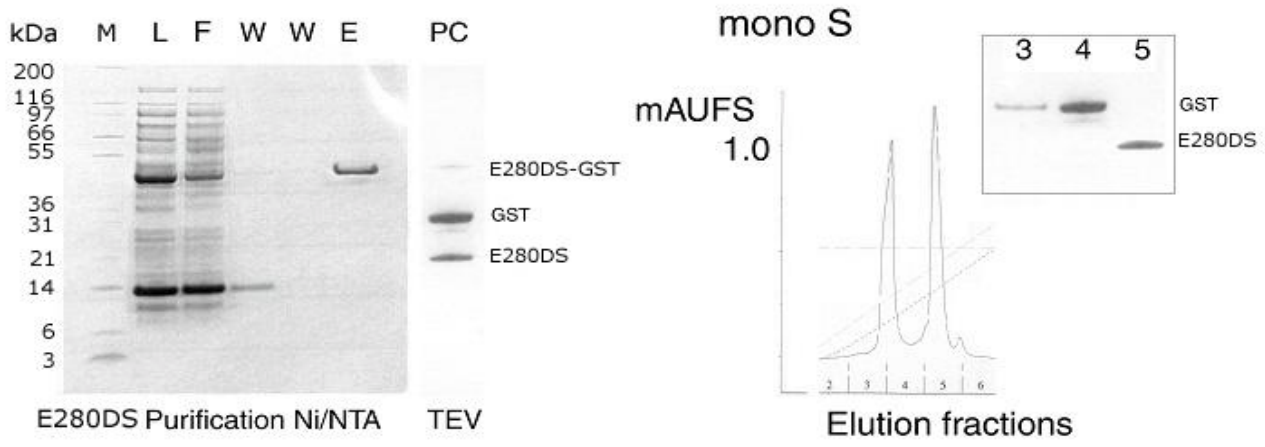


Figure 3.3 Ets-1 purification strategy. Purification protocol designed for the Ets-1 serine mutants E244ΔS, 280ΔS and E301ΔS (for clarity, results shown only for E280ΔS). The left-hand side shows a denaturing protein gel with the fractions of the Ni-NTA chromatography; 'L' is lysate, 'F' flow-through, 'W' wash and 'E' elution. The lane labelled 'PC' (for protease cleavage) shows an almost complete digestion of His6-GST-E280ΔS by TEV protease (100:1 [w/w] ratio, overnight on ice). On the right-hand side, the chromatographic profile of a cation exchange chromatography run is presented. The sample was as under 'PC,' with the fastest-eluting peak being the His6-GST moiety and the second peak pure E280ΔS. The inlet shows a SDS PAGE gel of the elution fractions, with lanes 3 and 4 with GST and lane 5 with E280ΔS.

For the His6-GST-fusion proteins, E244Δ, E280ΔS and E301ΔS, GST affinity chromatography (2.3.11) produced similar degrees of purity than IMAC affinity chromatography; however, the GST affinity resin has slower binding kinetics, requiring several applications of the sample and a relatively low flow rate (1/5 of the running flow rate) for optimum binding capacity.

The elution fraction or fractions (if several passes through a column were necessary) were collected in drop-in buffer (2.1.9.2) and kept on ice for further analysis.

3.1.2.8 Removal of affinity tag

All constructs had a TEV protease cleavage site engineered after the *c-ets-1* gene and immediately before the affinity tag for the convenient removal of the tag after purification. Proteolytic cleavage of the tag was performed as described (2.3.13), limiting the amount of protease to a weight ratio of 1/50 (protease to tagged protein) and performing all digestions overnight at 4 ºC in dialysis buffer (2.1.9.2). Protein concentration had to be reduced below 1 mg/ml to avoid precipitation during and after cleavage. Complete cleavage was confirmed by SDS PAGE electrophoresis (2.3.5) before proceeding further on.

3.1.2.9 Intermediate purification by CIEX chromatography

Removal of most bulk contaminants (termed intermediate purification) was achieved by CIEX

(2.3.11). As shown in Table 3.1, the isoelectric point (pI) of E244ΔS after tag removal is 6.81, 8.65 for E280ΔS, 9.60 for E301 and 9.25 for E335. This ensures that at pH 8.0 E280ΔS, E301ΔS and E335ΔS will bear a calculated charge of +0.78, +3.77 and +1.79 units, respectively (Rice et al, 2000). However, E244ΔS, at the same pH of 8.0, will bear a calculated charge of -1.95 units and yet it can be purified in exactly the same way. This observation reflects the dipolar character of E244ΔS, which attaches to the cation exchange resin through the positively charged ETS DBD whilst the N-terminal AI motif has excess of negative charges.

Purification was performed as described previously (2.3.11). Protein peaks were fractionated in 2-5 ml tubes and, after electrophoresis (2.3.5), pooled together according to band profile. Typically, after CIEX purity was near 95% as measured by band intensity.

Depending on the purpose and the amount of material, the purified protein could be used immediately by dialysis of the pooled fractions into a suitable buffer (2.1.9.2), or could be subjected to reverse phase chromatography (RPC) as a final step before storage by lyophilisation (2.3.11; 2.3.1). Freezing of both unprotected and cryoprotected solutions of Ets-1 constructs denatures them, as shown by large precipitates produced upon thawing.

3.1.2.10 Polishing step by RPC chromatography

The removal of DNA/RNA and/or small molecule contaminants (vg phospholipids, sugars, buffer salts) was carried out by reverse phase chromatography (RPC), which had the side effect of transferring the eluted pure protein (>99% pure) to a volatile solvent, 40-60% (v/v) acetonitrile (ACN), which is extractable by freeze-drying the sample in a lyophiliser. The detailed purification procedure has been described (2.3.11).

The lyophilised protein sample could be stored sealed at -80 ºC for long periods (over a year) without detriment or loss of activity.

3.1.2.11 Biophysical characterisation of purified protein by SDS PAGE electrophoresis, native electrophoresis, MALDI-TOF mass spectrometry and dynamic light scattering

SDS PAGE electrophoresis (2.3.5) was used throughout the purification process of each Ets-1 construct to monitor its molecular weight (MWapp) and integrity (absence of degradation or aggregation); accurate molecular weights (MWacc) were obtained by matrix assisted laser desorption ionisation (MALDI)-time of flight (TOF) mass spectrometry (MS).

The oligomeric state of Ets-1 constructs in solution was monitored by native gel electrophoresis (2.3.5). Ets-1 is expected to be a monomer, but multimers can form by covalent cross-linking in wild-type Ets-1 or as a result of non-specific aggregation (in both wild-type Ets-1 and cysteine-to-serine mutants). E335ΔS was indeed a monomer in solution at least up to 2-3 mg/ml, but E244ΔS, E280ΔS and E301ΔS form aggregates of high molecular mass (estimated as >200 kDa, >8 times larger than the corresponding monomeric Ets-1 construct) above 1 mg/ml (Fig 3.5, lane N).

Sample homogeneity was assessed by dynamic light scattering (DLS) (2.3.16) with 0.5-2 mg/ml samples. E244ΔS, E280ΔS and E301ΔS form, under a variety of conditions, aggregates of estimated size >200 kDa nm and polydispersity of 25-35%. Reducing the pH and/or the temperature reduced the polydispersity but did not change the size of the aggregates. E335ΔS was monodispersed in solution at this concentration with polydispersity of 20%.

3.1.2.12 Initial crystallisation screens of Ets-1 E280ΔS and E301ΔS

Although substantial crystallisation efforts were dedicated to E280ΔS and E301ΔS, the relatively low concentration at which both proteins can be brought to without precipitation seriously limited the experiment. Crystallisation setups were performed in sitting drop or hanging drop mode (2.3.17) and at 4 or 20 ºC, comprising a complete battery of commercial crystallisation screens that included Hampton Research Crystal Screen I and II, Crystal Cryo, Crystal Lite, PEG/Ion, Low Ionic Strength and Natrix screens. Protein samples for crystallisation were prepared systematically in the same fashion and concentrated up to 1-2 mg/ml in HN buffer (2.1.9.2). The detailed procedure has been explained previously (2.3.17).

At 4 ºC, most conditions remained clear for weeks or else showed fine precipitates. However, at 20 ºC, a substantial fraction of all conditions (>50%) showed amorphous and/or microcrystalline precipitates (as judged under the light microscope with a cross polariser attachment). Optimisation by grid screening about those conditions that rendered microcrystalline precipitates failed to produce protein crystals. Seeding of the precipitates rendered no better results.

3.1.3 Discussion

3.1.3.1 Design of wild-type Ets-1 constructs and small-scale expression tests

The design of Ets-1 constructs that would remain autoinhibited while being compact was difficult for several reasons: (1) limited proteolysis experiments had demonstrated that E280 was the minimal Ets-1 construct that still remained fully autoinhibited when compared with full-length Ets-

1 (Petersen et al, 1995); (2) in those experiments, E280 was progressively digested down to E301, which only exhibits partial autoinhibition (Petersen et al, 1995); (3) the ETS domain and the AI motifs lay within the exon VII of human *c-ets-1* gene, which spans from 244-441 and has been predicted to have its N-terminus unstructured by secondary structure prediction by PHD (Rost et al, 1994). These results implied that a fully autoinhibited construct (E244 or E280) should contain regions of high flexibility, while a more compact, smaller construct (E301 or E335) might lose biological relevance by being only partially inhibited.

A wide range of constructs were made (2.2.11) and assayed initially for expression in *E coli* BL21(DE3) cells (2.3.10). The screening consisted of 10 different Ets-1 deletion mutants (E335A, E335E, E301, E280, E244, E238, E234, E73, E51 and FL or full-length Ets-1) and 6 expression vectors (pET15b, pETM11, pETM30, pQE70, pQE70-HT and pBAT4) totalling 55 constructs (see 2.2.11 and Table 2.4 for clone nomenclature). In all cases tested Ets-1 constructs expressed poorly and always as insoluble inclusion bodies, or not at all (Table 3.2). Constructs of E234, E74, E51 and FL could not be expressed in detectable amounts (<0.1 mg per litre of culture). Among all possible combinations of insert and expression vectors that rendered yields above 0.1 mg per litre of culture, the pBAT4 clones always produced the highest yields; indeed, E238 could only be expressed as a pBAT4 clone. Interestingly, pQE70, which should also express untagged protein, did not express any of the Ets-1 clones with the exception of the two control clones E335(A2) and E335(E2). E244, E280 and E301 could be expressed either untagged (from pBAT4) or as His6-GST fusions from pETM30, but not as His6-tagged product (as from pETM11 or pQE70-HT).

Since pBAT4-E244 was the longest Ets-1 construct that could be expressed above 1 mg per litre of culture, it was chosen for further analysis.

3.1.3.2 Redox Susceptibility of wild-type Ets-1

A recurrent theme during the initial attempts to produce, isolate, stabilise and purify E244 was its marked tendency to precipitate and lodge into a white, insoluble pellet of cross-linked protein. SDS PAGE analysis (2.3.5) of the precipitates showed various degrees of cross-linking (Fig 3.4). This was interpreted as an oxidation ladder developed by intermolecular disulphide bridge formation of the unfolded protein and demonstrated by TCEP incubation (Fig 3.1). This phenomenon rendered impossible any manipulation or characterisation of the protein samples of wild-type Ets-1.

	Construct								
Vector	**E335A/E**	**E301**	**E280**	**E244**	**E238**	**E234**	**E73**	**E51**	**FL**
pBAT4	++	++	+	++	+	+	0	0	0
pQE70	+	+	0	0	0	0	0	0	0
pET15b	++	N/A	N/A	0	0	0	N/A	N/A	N/A
pETM11	++	++	0	0	0	0	0	0	0
pQE70-HT	+	+	0	0	0	0	0	0	0
pETM30	+++	+++	+	+	0	0	0	0	0

Table 3.2 Small-scale expression tests of wild-type Ets-1 constructs. A total of 55 clones were assayed for expression in small-scale cultures as previously described (2.3.10). Every cell corresponds to one such assay, and can be labelled with one or more plus signs (that construct expressed), with zero (no detectable expression or yield below 0.1 mg per litre of culture) or N/A for constructs that were not assayed. Among the constructs that did express, those labelled with a single plus sign, '+,' yielded between 0.5-1.0 mg per litre of culture, those labelled with two, '++,' yielded 1.0-5.0 mg per litre of culture and those labelled with three, '+++,' yielded more than 5.0 mg per litre of culture (the highest yield was 20 mg per litre of culture for pETM30/E335A2 and E2). All expressed protein was in the form of inclusion bodies. Protein concentration was measured by the Bradford assay (2.3.2) on 40-75% pure protein samples resuspended from insoluble pellets, and purity estimated from electrophoresis gels by densitometry (2.3.5). The two positive controls, E335A2 and E2, displayed the highest yields for all vector and construct combinations regardless of tags or fusion partners. In addition, the difference in yield between A2 and E2 for each vector was negligible, suggesting that, at least for these constructs, an alanine or a glutamic acid at position 2 did not affect expression. No E234, E74, E51 and FL clones did express. E238 could only be expressed from pBAT4. E244, E280 and E301 could be expressed either from pBAT4 as untagged proteins or as His6-GST fusions from pETM30.

3.1.3.3 Refolding of E244 and E280

Following the rationale that E244 oxidation was enhanced by the exposure of the free thiol groups of cysteines C261, C350 and C416, much effort went into obtaining a folded protein by refolding under reducing conditions. However, this strategy proved fruitless as E244 oxidised and precipitated just as fast and completely before and after refolding (Fig 3.2, lane 1, and Fig 3.4). Despite having one cysteine fewer (C261), E280 precipitated as much as E244. Table 3.3 shows the methods employed for E244 and E280 refolding (see 2.3.12 for further details).

Refolding techniques
1. Dialysis in a single step against refolding buffer
2. Dialysis in multiple steps (step-wise)
3. Dialysis against refolding buffer and additives
4. Dilution in refolding buffer (20-50 fold)
5. Refolding by gel filtration
6. On-column refolding (on a Ni-NTA column, by gravity-flow)
7. Refolding in the presence of DNA

Table 3.3 Refolding techniques tried upon with E244 and E280 (see 2.3.12 for further details). In all cases but one, E244 and E280 protein samples were produced by solubilisation of inclusion bodies in GU buffer (2.1.9.2); in one instance of the dialysis methods (1-3) and the dilution method (4), inclusion bodies were resuspended in NL buffer (2.1.9.2). The refolding buffer was based on HN or TN buffers (2.1.9.2), supplemented in method 3 with various additives (2.3.12).

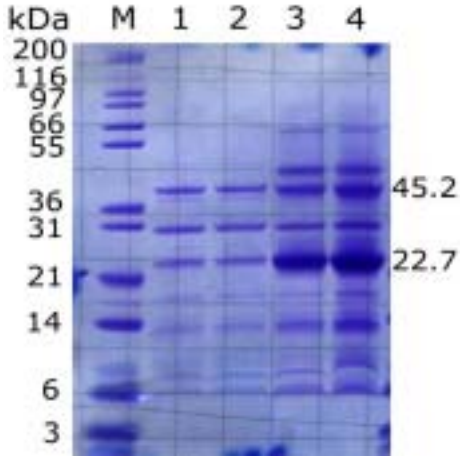


Fig 3.4 Analysis of E244 precipitates. SDS PAGE electrophoresis (2.3.5) analysis of the precipitates of wild-type E244 produced by oxidation at 4 ºC on a partially purified preparation. Lanes 1-2 show the supernatant and lanes 3-4 the precipitate (collected by centrifugation) of protein samples incubated at 4 ºC during 12 h (overnight incubation). Both supernatant and precipitate were analysed in the absence (1, 3) and presence (2, 4) of 10 mM DTT.

3.1.3.4 Site-directed mutagenesis of all cysteines in Ets-1 244-441 into serines

As the sequence alignment in Fig 1.4 shows, cysteines C350 and C416 are not strictly conserved across the ETS protein family (the same argument can be raised for C261, alignment not shown) and thus mutations C350ΔS and C416ΔS (and C261S) should in principle not interfere with protein folding or activity. However, since Ets-1 interacts with USF1 through an as yet uncharacterised interface, loss of biological relevance of the cysteine-to-serine mutants was analysed by electrophoretic mobility shift assay (Fig 3.7) (2.3.7).

Removal of C261, C350 and C416 from wild-type Ets-1 constructs lead to oxidation-resistant proteins that could be expressed, purified and characterised. This is unsurprising if the degree of solvent exposure of C350 and C416 is considered - vg by calculation of their solvent accessibility in WhatIf (Vriend, 1990) or by visual inspection of the NMR structures 2stw (Werner et al, 1995) and 1etd (Donaldson et al, 1996) - and the N-terminal end of E244, E280 and E301 is assumed partially unstructured.

3.1.3.5 Enhanced stability of mutant Ets-1 constructs

Mutation of all cysteines into serines increased Ets-1 solubility by preventing the precipitation of wild-type Ets-1 by disulphide bond formation, but it did not stop Ets-1 from developing precipitates by aggregation. Analysis of the wild-type Ets-1 precipitates showed covalent bonding between protein chains (Fig 3.4), whereas mutant Ets-1 precipitates did not reveal any cross-linking (Fig 3.2). Ets-1 precipitation seemed to arise from hydrophobic and electrostatic interactions, as illustrated by native electrophoresis (2.3.5) of E280ΔS with increasing concentrations of NaCl and

glycerol (Fig 3.5).

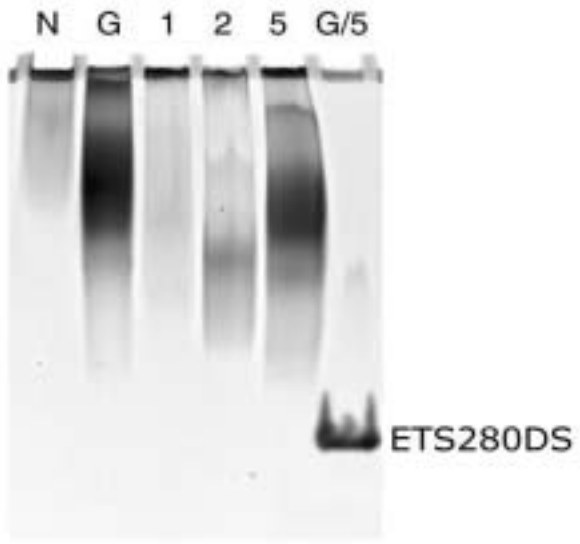


Fig 3.5 Effect of ionic strength and glycerol on the stability/solubility of E280ΔS. Native gel electrophoresis on a 8% TBE gel of mutant E280ΔS with varying concentrations of NaCl and glycerol. 'N' is a control lane without additives, 'G' is 10% (v/v) glycerol; '1,' '2' and '5' are lanes with 100, 200 and 500 mM NaCl, respectively; and 'G/5' has both 10% (v/v) glycerol and 500 mM NaCl. Only when both 10% (v/v) glycerol and 500 mM NaCl does E280ΔS run as an isolated species. At lower salt concentrations, or with 10% (v/v) glycerol, E280ΔS behaves as a particle of >200 kDa which does not even enter the gel. The fact that the two additives must be present for E280ΔS to migrate as a single, isolated protein is indicative that both hydrophobic and ionic interactions are responsible for the aggregation observed.

The most substantial loss of material by precipitation occurred during the proteolytic removal of the His6-GST tag of E244ΔS, E280ΔS and E301ΔS and of the His6 tag of E335ΔS cleavage, with losses of 75-80% (w/w) protein. Precipitation was reduced by filtering the protease solution through a 0.22 mm membrane or by using TEV protease produced under native conditions. Taken together, these data suggest a role for the TEV protease in the precipitation of Ets-1.

Despite the protective role of glycerol against non-specific aggregation of mutant Ets-1 (Fig 3.5), glycerol was avoided during protein preparation for crystallisation of isolated Ets-1 constructs and Ets-1 complexes.

3.1.3.6 Purification of Ets-1 mutant constructs

The purification of E244ΔS, E280ΔS, E301ΔS and E335ΔS, comprised three main steps: A high-capacity capture step by Ni-NTA affinity chromatography, an intermediate step by CIEX chromatography and, in certain cases, a final step by RPC chromatography (2.3.11 and 3.1.2). Several techniques were tested as capture step, such as GST affinity chromatography, unspecific

DNA affinity chromatography and heparin-sepharose affinity chromatography (2.3.11), but the Ni-NTA technology proved the best in protein yield, enrichment and manipulation speed.

Interestingly, E244ΔS, with the slightly acidic pI of 6.81, could be purified by CIEX in the same way as the other constructs, whose pI's are above 8.5. A likely explanation is that the binding of E244ΔS to the CIEX resin is mediated by the positively charged ETS domain, whilst the N-terminus, which adds 2 net negative charges to the construct, remains free.

This protocol has been successfully applied to other transcription factors (USF1, see above), and can be automated since the elution buffers are all compatible with the loading buffers of the following steps and there is no sample dilution during the entire process, which would make necessary manual concentration. The possibility to perform on-column digestion of the His6-GST and His6 tags was verified experimentally (2.3.13), and could possibly lead to an increased automation of the process.

3.1.3.7 Crystallisation of E280ΔS and E301ΔS

Both purified E280ΔS and E301ΔS could be concentrated by ultrafiltration (2.3.3) up to 1-2 mg/ml in HN buffer (2.1.9.2), concentration that is generally too low for successful crystallisation (typically, 5-20 mg/ml [Ducruix & Giege, 1999]). Although no precipitation could be evidenced at this concentration, DLS experiments (3.1.2) revealed extensive aggregation beyond 1-2 mg/ml with polydispersity of 25-35%. These two facts, the low concentration of the protein samples and the development of aggregates at slightly higher concentrations, provide an explanation for the failure to obtain suitable crystallisation conditions for E280ΔS and E301ΔS despite quite an extensive initial screening (3.1.2). Most precipitates that presented a microcrystalline appearance under the polarimeter were (a) assumed as starting conditions for further optimisation and (b) utilised as crystallisation seeds, but neither approach has so far rendered protein crystals.

An open question about the crystallisation of autoinhibited Ets-1 constructs was how the N-terminal AI motif, which is partially unstructured according to secondary structure predictions, would affect nucleation and crystal growth. Mutation of a number of residues within this motif, whose phosphorylation reinforces autoinhibition, into glutamic and/or aspartic acid residues might aid crystallisation by making the Ets-1 constructs more compact (Cowley & Graves, 2000).

3.1.4 Summary

Four Ets-1 constructs, pETM30-E244ΔS, pETM30-E280ΔS, pETM30-E301ΔS and pETM11-

E335ΔS were expressed in *E coli* BL21(DE3) cells (2.3.10), purified by IMAC affinity chromatography and cation exchange chromatography, and polished by reverse phase chromatography (2.3.11) for storage as lyophilate (2.3.1). Their stability was significantly improved with respect to that of the corresponding wild-type Ets-1 constructs, which underwent prompt and irreversible oxidation by intermolecular disulphide bridge formation. Nevertheless, all proteins exhibited a strong tendency to form non-specific aggregates at concentrations above 1-2 mg/ml and thus were kept at lower concentrations (typically, 1 mg/ml). At that concentration, all proteins form monodisperse solutions when analysed by dynamic light scattering.

Although crystallisation setups were performed with E280ΔS and E301ΔS protein solutions, no crystals of enough size (at least 0.5 mm along every direction) have been grown to date. In the absence of a clear lead after an initial crystallisation screening that encompassed several commercial screens, microcrystalline precipitates were used for optimisation and streak seeding in hopes of producing small protein crystals. The likeliest explanation for the lack of protein crystals is the rather low highest concentration attainable with E280ΔS and E301ΔS, roughly 1 mg/ml, which severely limits the sampling of the space of crystallisation conditions.

3.2 USF1 cloning, expression and purification

3.2.1 Introduction

Throughout this study we have focused on the expression, purification and crystallisation (as a ternary complex with Ets-1 and DNA) of the bHLH/ZIP (LU) and bHLH (SU) domains of USF1 (residues 196 through 310 or 260, respectively). Protocols existed for the expression, refolding from inclusion bodies, purification and crystallisation of USF1 (Ferre-D'Amare, 1994; Sieweke et al, 1998), modifications of which were used to express different USF1 constructs for complex formation with Ets-1.

3.2.2 Results

3.2.2.1 Molecular cloning of the USF1 constructs

Two USF1 constructs spanning residues 194-310 (LU) and 194-260 (SU) were cloned into several expression vectors as described (2.2.11) (see Table 2.4 for details on nomenclature and vectors) and tested for stability/toxicity (2.3.9) and recombinant protein expression (2.3.10). Of the resulting combinations, the pETM11-USF1 clones produced the highest yields, 5-10 mg of pure protein per litre of culture. The corresponding clones carrying cysteine-to-serine mutations were cloned by

identical procedures into pETM11 and pProExHTb (2.2.11). Table 3.4 lists the clones used for expression.

Construct	Vector/Tag	USF1[a]	MW/pI[b]
pETM11-LUΔS	pETM11/His6	USF1 194-310	13.8/9.31
pETM11-SUΔS	pETM11/His6	USF1 194-260	8.0/10.56
pProExHTb-LUΔS	pProExHTb/His6	USF1 194-310	13.8/9.31
pProExHTb-SUΔS	pProExHTb/His6	USF1 194-260	8.0/10.56

Table 3.4 USF1 expression clones. Best expression clones for the USF1 LUΔS and SUΔS constructs. All constructs rendered the highest yield in soluble protein as His6-tagged proteins. [a] All USF1 constructs were all-serine mutants, ie, every cysteine in wild-type USF1 was mutated into a serine to avoid oxidation; [b] relative molecular mass, or molecular weight, and isoelectric point, pI, for every protein, after removal of the tags by proteolysis.

3.2.2.2 Site-directed mutagenesis

Production of wild-type USF1 was achieved in the range of 5-10 mg per litre of culture in soluble form, but after overnight incubation at 4 °C USF1 started to precipitate in an irreversible fashion. SDS PAGE electrophoresis of the precipitate revealed cross-linking and extensive oxidation, revertible only by treatment with 100 mM TCEP (2.1.1) at RT. Therefore, LU and SU were modified by site-directed mutagenesis by changing their two cysteines C229 and C248 into serines (2.2.10). The mutant proteins (SUΔS and SUΔS) exhibited excellent behaviour in terms of solubility (no precipitation after several days incubation at 4 ºC, no cross-linked precipitates), and were retained henceforth for structural biology purposes. The preferred clones for production of USF1 were pProExHTb-LUΔS and pETM11-SUΔS.

3.2.2.3 Expression of SUΔL and SUΔS constructs

LUΔS and SUΔS were expressed in *E coli* BL21(DE3) as N-terminally His6-tagged recombinant proteins from a pProExHTb and a pETM11 vector, respectively, with typical yields about 5-10 mg protein per litre of culture each. Both proteins expressed mostly in the soluble fraction.

3.2.2.4 Protein purification

LUΔS and SUΔS were purified following identical procedures, since they share an N-terminal His6-tag for IMAC affinity chromatography (2.3.11) and similar isoelectric point (pI) of 9.31 and 10.56, respectively. Furthermore, their purification followed closely the purification of the mutant Ets-1 constructs, with Ni-NTA affinity purification as capture step, CIEX as intermediate purification step, and usually a final polishing step by RPC. Fig 3.6 shows the purification strategy

devised for the USF1 constructs.

3.2.2.5 Protein extraction and cell lysis

After protein expression and centrifugation, harvested cells could be frozen away at -80 ºC or processed immediately (2.3.10). In either case, the cell pellets were lysed with 20 ml of Ni-NTA lysis buffer (2.1.9.2) per litre of culture by sonication (2.3.10), supplemented with 5-10 mM βME. Cell debris and unbroken cells were removed by centrifugation, and both pellet and supernatant were kept at 4 ºC for analysis by SDS PAGE electrophoresis (2.3.5).

3.2.2.6 Capture by IMAC chromatography

The His6-tagged proteins were isolated, stabilised and concentrated (the capture step) by IMAC affinity chromatography (2.3.11) either in gravity-flow mode (casting a suitable volume of Qiagen Ni-NTA resin on a BioRad plastic column) or by high performance liquid chromatography (HPLC) on a Pharmacia HiTrap column previously charged with nickel (2.3.11). The exact procedure has been detailed (2.3.11).

The elution fraction or fractions (if several passes through a column were necessary) were collected in drop-in buffer (2.1.9.2) and kept on ice for further analysis.

3.2.2.7 Removal of the affinity tag

All constructs had a TEV protease cleavage site engineered between the *usf1* gene and the tag for its convenient removal. Proteolytic cleavage of the tag was performed as described before (2.3.13), limiting the amount of protease to a weight ratio of 1/50 (protease to tagged protein) and performing all digestions at 4 ºC overnight. LUΔS concentration had to be reduced below 1 mg/ml to avoid precipitation, whilst SUΔS would stay in solution up to greater than 5 mg/ml. Complete cleavage was confirmed by SDS PAGE electrophoresis (2.3.5) before proceeding further on.

In contrast with wild-type and mutant Ets-1 clones, USF1 clones did not undergo enhanced precipitation during proteolytic treatment, probably because of their greater solubility.

3.2.2.8 Intermediate purification

Removal of most bulk contaminants (intermediate purification) was achieved by CIEX (2.3.11). As shown in Table 3.4, the isoelectric point (pI) of SUΔS after the removal of the tag by protease cleavage is 9.31 and that of LUΔS is 10.56, what ensures that at pH 8.0 all of them will bear a (calculated) net positive charge of +2.42 and +5.36 units, respectively (Rice et al, 2000).

Purification was performed as described (2.3.11). Protein peaks were fractionated in 2-5 ml tubes and, after SDS PAGE electrophoresis (2.3.5), pooled together according to the band profile. Typically, after CIEX the purity was near 95% as measured by densitometry of the protein bands on a SDS PAGE gel.

Purified USF1 could be used immediately by dialysing the pooled fractions against a suitable buffer (vg HN or TN, 2.1.9.2), or could be subjected to RPC as a final purification step before storage by lyophilisation (2.3.1). Freezing of both unprotected and cryoprotected solutions of USF1 (with 50% [v/v] glycerol) constructs denatured them, as shown by the appearance of precipitates upon thawing.

3.2.2.9 Polishing by RPC chromatography

The removal of DNA/RNA and/or small molecule contaminants (vg phospholipids, sugar, buffer salts) was carried out by RPC. As elution of the bound sample was performed with increasing amounts of acetonitrile (ACN), which is volatile, the eluate, consisting of mostly pure protein (>99%), could be freeze-dried by lyophilisation (2.3.1). The lyophilised protein sample could be stored sealed at -80 ºC for long periods (over a year) without detriment or loss of activity.

3.2.2.10 Refolding of lyophilised USF1

The refolding of lyophilised USF1 constructs was done by dilution (2.3.12) in refolding buffer in the presence of one of the cognate DNA recognition sequences LTR33 and LTR32 (2.1.5). First, 10 mg of lyophilate was dissolved in 0.5 ml GU buffer (2.1.9.2, 2.3.8); then, dissolved protein sample was added gently and stepwise to 20 ml of refolding buffer with 20-30 mM LTR33 or LTR32. The refolding buffer was based on the dialysis buffer HN (2.1.9.2), but with 20 mM HEPES instead of 10 mM HEPES (20 mM HEPES was the minimum buffer concentration necessary to achieve refolding). The final ratio of protein to DNA is 1:1.1-1.2.

After the addition of USF1, the refolding mixture was incubated at 4 ºC for 2-3 h and, immediately afterwards, concentrated by ultrafiltration (2.3.3) and buffer exchanged by dilution and re-concentration. This last step helped to lower GuHCl concentration to 0.3 M. The maximum protein concentration attainable was of 10-15 mg/ml.

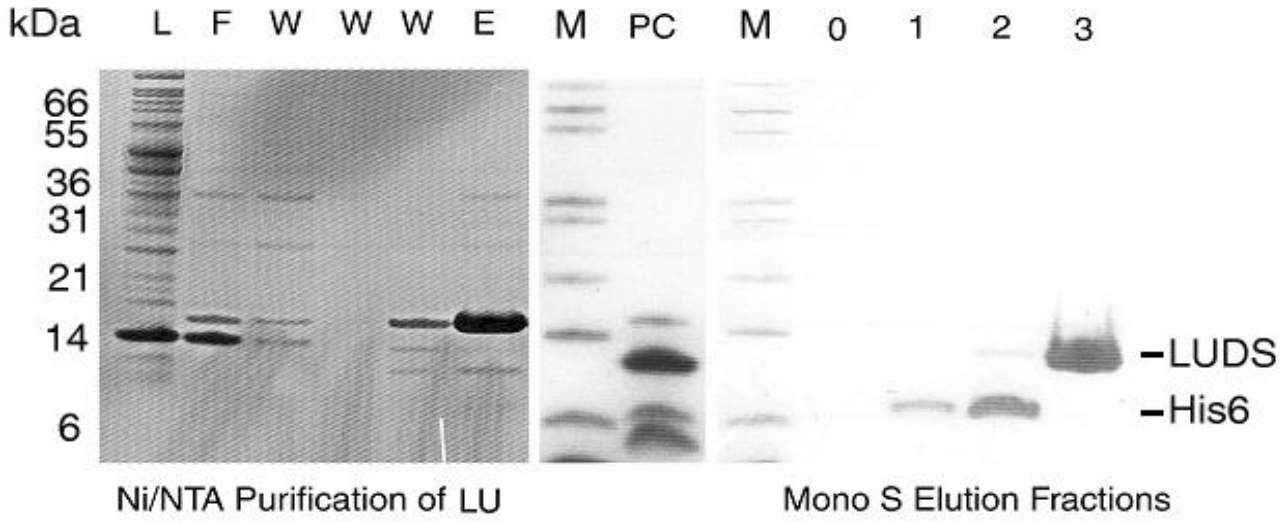


Fig 3.6 Purification of mutant USF1 constructs. Overview of the purification of USF1 constructs (shown only for LUΔS for clarity). The left-hand side shows a SDS PAGE gel with the fractions of the Ni-NTA chromatography (2.3.11); 'L' is lysate, 'F' flow-through, 'W' wash and 'E' elution. The lane labelled 'PC' (for protease cleavage) shows an almost complete digestion of His6-LUΔS by TEV protease (100:1 [w/w] ratio, overnight on ice). The left-hand panel shows a SDS PAGE (2.3.5) gel of the elution fractions of the CIEX step, with lanes 1-2 with His6 tag and lane 3 with pure LUΔS.

3.2.2.10 Biophysical characterisation of purified LUΔS and SUΔS by SDS PAGE electrophoresis, native electrophoresis, MALDI-TOF mass spectrometry and dynamic light scattering

SDS PAGE electrophoresis (2.3.5) and MALDI-TOF MS were used throughout the purification process of both USF1 clones to monitor correct approximate molecular weight and integrity (absence of degradation or aggregation).

The oligomeric state of USF1 in solution was monitored by native gel electrophoresis (2.3.5). LUΔS is expected to be a homotetramer, since it homodimerises through the HLH motif and forms dimers of homodimers via the LZ interface (Ferre-D'Amare et al, 1994; Ferre-D'Amare and Burley, 1994), which was confirmed in all cases.

Sample homogeneity was assessed by DLS with 1 mg/ml samples (2.3.16), which showed that they form monodisperse solutions of particles of the expected size.

3.2.3 Discussion

3.2.3.1 Construct design of USF1 clones and expression tests

USF1 194-310 was first expressed as a pET15b-LU, with an N-terminal His6 tag and a thrombin

cleavage site. Although expression yields were high (above 5-10 mg per litre of culture), thrombin treatment to cleave off the tag lead to a thrombin product within the USF1 sequence, which accumulated with time. To circumvent this problem, USF1 was re-cloned into the vectors pETM11, pETM30 and pProExHTb (2.2.11 and Table 2.4) which introduced a TEV protease cleavage site, and assayed for expression (2.3.10) and proteolytic cleavage, as described (2.3.13). Indeed, TEV protease treatment did not produced unwanted proteolytic products.

Although wild-type USF1 clones expressed in the soluble fraction, the recombinant protein proved unstable within hours and precipitated. As discussed before for Ets-1 (3.2.2, 3.2.3), USF1 underwent oxidation by disulphide bridge formation, even in the presence of 100 mM DTT in degassed, nitrogen-purged buffers. Site directed mutagenesis (2.2.10) of cysteines C229 and C248 into serines removed all major oxidable groups and stopped USF1 precipitation. The preferred clones for expression of the cysteine-to-serine mutants were pProExHTb-LUΔS and pETM11-SUΔS.

3.2.3.2 Redox Susceptibility of wild-type USF1 clones

As with Ets-1, USF1 possesses a number of cysteines, C229 and C248, situated on the loop connecting the two Á-helices of the HLH motif. C229 and C248 side chains protrude towards the bulk solvent and are thus prone to oxidation. Removal of both cysteines abolished the oxidation problem and augmented the long-term solubility of USF1 constructs (from hours to days at 4 ºC). Since C229 and C248 have such a dramatic effect on USF1 redox susceptibility, they could be involved in the redox regulation proposed as a physiological regulatory mechanism for USF1 (Pognonec et al, 1992).

3.2.3.3 Stability of the mutated USF1 constructs

As commented above, oxidation of USF1 cysteines caused most of the solubility problems for both LU and SU, therefore their substitution by serines eliminated the problem. However, USF1 long-term solubility was still compromised by denaturation and aggregation, which was more severe for LUΔS than for SUΔS, possibly because of interactions mediated by the LZ interface of LUΔS.

3.2.3.4 Purification of LUΔS and SUΔS constructs

The purification strategy followed with the USF1 constructs LUΔS and SUΔS reflects the protocols established for the mutated Ets-1 clones, with a capture step consisting of an IMAC chromatography on a Ni-NTA column, proteolytic removal of the His6 tag by TEV protease treatment, application onto a CIEX column and, once the right fractions were pooled together:

either (1) the pooled fractions were dialysed against a suitable buffer (vg HN or TN) (2.1.9.2) or (2) it was applied on a RPC for polishing and storage (2.3.1).

3.2.3.5 Refolding of LUΔS and SUΔS in the presence of the recognition site

The realisation that LUΔS and SUΔS could be refolded in the presence of their binding sites with highest efficiency came from a series of comparative experiments with the gel filtration refolding technique (Ferre-D'Amare et al, 1994) (2.3.12). The presence of the DNA target sequence in the refolding buffer could favour proper folding by (1) actively promoting folding or, alternatively, (2) by sequestering properly folded protein into a more stable protein-DNA complex and, thus, displacing the equilibrium of the refolding reaction towards folding.

The loss of material occurred during concentration of the refolding reaction proceeds from unfolded USF1 and, beyond 3-5 mg/ml, to excess (folded or unfolded) USF1. After the developing precipitates are removed by filtration, the sample can be further concentrated to a maximum of 10-15 mg/ml.

3.2.4 Summary

The two cysteine-to-serine USF1 constructs SUΔS (USF1 194-260) and LUΔS (USF1 194-310) were cloned into pETM11 and pProExHTb, respectively, for expression of USF1 as a His6-tagged protein in *E coli* BL21(DE3) (2.3.10). Typical yields reached 5-10 mg per litre of culture. Purification of both proteins started off with an IMAC affinity chromatography step with Qiagen Ni/NTA resin, followed by proteolytic removal of the tag with TEV protease, CIEX and a final polishing by RPC (2.3.11) for storage as lyophilate (2.3.1). The purified proteins were >95% pure as judged by densitometry of SDS PAGE gels. The stability of both mutant proteins was significantly improved with respect to that of the corresponding wild-type USF1 constructs, which underwent irreversible oxidation by intermolecular disulphide bridge formation. Nevertheless, LUΔS exhibited a tendency to form non-specific aggregates at concentrations above 2-3 mg/ml, whilst SUΔS would remain soluble up to 5 mg/ml at least.

Oligomeric state, integrity and aggregation state were followed by denaturing and native electrophoresis (2.3.5). Polydispersity was assessed by DLS (2.3.16).

Refolding of lyophilised LUΔS and SUΔS needed the presence of their DNA binding sites in the refolding buffer for maximum efficiency. This observation forms the basis of the protein-DNA complex formation of USF1/DNA and of the ternary complexes Ets-1/USF1/DNA, and it

constitutes, together with the substitution of cysteines by serines, a key technical aspect of this work.

3.3 Complex formation, purification and initial crystallisation of several Ets-1/USF1/DNA ternary complexes

3.3.1 Introduction

As mentioned before (3.1.1), Ets-1 cannot bind to its cognate DNA target sequences on its own since it is constitutively autoinhibited (Lim et al, 1992; Jonsen et al, 1996). Therefore, it needs to establish stable interactions with a partner protein that itself can bind to DNA and which recruits Ets-1 onto DNA. One such partner protein is USF1 (Sieweke et al, 1998). The interaction of DNA-bound USF1 with free Ets-1 brings the latter to an adjacent EBS by a synergistic mechanism yet to be uncovered (see Fig 1.5 for a model).

Although there is structural information on several ETS family proteins and the complexes they form (Table 1.2 and references thereof), including the ternary complex of the minimal ETS domain of Ets-1 with Pax and their combined site (Garvie et al, 2001), little is known about how the release of autoinhibition necessary for Ets-1 productive binding to DNA operates - all known structures are from non-inhibited forms of Ets-1. Since the switch between Ets-1 free and bound states is an important regulatory mechanism of Ets-1 activity, a complex between Ets-1 in an autoinhibited form and a protein partner, with DNA, would provide solid grounds for the test of the proposed models.

3.3.2 Results

3.3.2.1 Ternary complex formation

Ets-1/USF1 complexes on DNA could be formed and purified from the wild-type proteins, but even as a complex their stability was so seriously compromised by aggregation and cross-linking that the complex was quickly trapped into a pellet of denatured protein. However, the sole fact that both proteins displayed such a (comparatively) long half-life as a ternary complex clearly indicated that they were protected or stabilised when bound to DNA. For maximal stability and for the purpose of crystallisation, Ets-1/USF1/DNA complexes were formed from cysteine-to-serine mutant Ets-1 and USF1. More particularly, the complexes studied are listed in Table 3.5.

To reconstitute each ternary complex, lyophilised Ets-1 was renatured by dilution (2.3.12) in the presence of 1:1.2 molar ratio of USF1/DNA (LTR33 or LTR32, 2.1.5) complex in refolding buffer (20 mM HEPES, 100 mM NaCl, 2 mM DTT, 2 mM EDTA, pH 7.5). In more detail, first, 10 mg of Ets-1 lyophilate was dissolved in 0.5 ml GU solubilisation buffer (2.1.9.2, 2.3.8) to make a 20 mg/ml (denatured) protein solution. Then, this protein solution was added slowly (in 50-100 ml aliquots to avoid too high a local concentration of GuHCl) onto a 20 ml 15-25 μM USF1/DNA complex solution, so that the final concentration of the ternary complex have an upper bound of 15-20 μM. Excess Ets-1 will precipitate immediately or upon concentration. The refolding reaction is incubated at 4 ºC for 2-3 h. The stepwise addition of Ets-1 is a key step in the entire process, since the efficiency of the refolding reaction is poor and can worsen with inefficient mixing. Losses of protein up to 25-40% (w/w) were expected during routine complex production.

Ets-1	USF1	DNA
E244ΔS	LUΔS, SUΔS	LTR33, LTR32
E280ΔS	LUΔS, SUΔS	LTR33, LTR32
E301ΔS	LUΔS, SUΔS	LTR33, LTR32
E335ΔS	LUΔS, SUΔS	LTR33, LTR32

Table 3.5 Ternary Ets-1/USF/DNA complexes for crystallisation. The complete list of complexes reconstituted and used in crystallisation setups can be generated by the combination of every Ets-1 construct with any of the two USF1 constructs and any of the two DNA elements, yielding a total of 4x2x2 or 16 distinct complexes (see Table 2.4 for construct nomenclature and 2.1.5 for LTR33 and LTR32 sequences).

USF1/DNA complexes were pre-formed by a similar process (see 3.2.2 for a detailed description), whereby USF1 was mixed at 4 ºC with a DNA solution by stepwise addition of the protein at a final molar ratio of 1:1-1.2 and an expected final concentration of 20-30 μM (approximately 1 mg/ml of USF1/DNA).

Once the Ets-1 ternary complex was formed, it was concentrated by ultrafiltration (2.3.3) up to a maximum concentration of 10-12 mg/ml (15 mg/ml for the E335ΔS complexes due to the much better solubility of the minimal domain). In all cases, concentration triggered considerable precipitation (sometimes up to 40% [w/w]) of protein, most likely excess and/or denatured protein. The concentrated complex retained 0.6 M GuHCl in its buffer, which should be removed by dialysis or by using a desalting column (2.3.11).

3.3.2.2 Biological activity

To confirm that the ternary complexes Ets-1/USF1/DNA could form and hold stably,

electrophoretic mobility shift assays (EMSA) were conducted with biotin-labelled LTR33 (as described in 2.3.7) (Fig 3.7). Interestingly, LUΔS does not homotetramerise either in the absence or in the presence of Ets-1, observation that is corroborated by gel filtration of the complexes (2.3.11, Fig 3.8). DLS of the complexes, though not quantitative, is also consistent with the absence of LUΔS homotetramers in the presence of Ets-1.

3.3.2.3 Gel exclusion chromatography of the ternary complexes

All formed ternary complexes were further purified by GF (2.3.11) for two distinct purposes: (1) To exchange the refolding buffer by GF buffer (2.1.9.2; 10 mM HEPES, 100 mM NaCl, 2 mM DTT, 2 mM EDTA, pH 7.5); and (2) the separation of the active complex (containing Ets-1, USF1 and DNA) from the remaining free DNA and/or unbound protein(s). In most, the amount of protein and/or DNA in excess was very low (as seen in the chromatograms), perhaps because of the precipitation of superfluous material during concentration of the complex.

Ternary complexes eluted at about the expected sizes (gel filtration standards were run routinely before any preparative run) of 46-75 kDa (see Fig 3.8 for a gel filtration chromatogram of the complex E301ΔS/SUΔS/LTR33).

Elution fractions were confirmed by SDS PAGE electrophoresis (2.3.5) to contain both proteins, and the ratio of A_{260}/A_{280} (about 1.8-2.0) identified the complex as binding DNA; those fractions were then pooled together and re-concentrated to the highest concentration possible.

3.3.2.4 Stability of the ternary complexes

There was extensive precipitation as complex formation proceeded by refolding of denatured Ets-1 onto a pre-formed USF1/DNA binary complex. Presumably, excess unfolded Ets-1 formed the precipitate whilst refolded Ets-1 stayed stable as a ternary complex. It is also conceivable that part of the pre-formed USF1/DNA complex may have been disrupted by locally high GuHCl concentration or by interaction with unfolded Ets-1. As the refolding reaction, initially 20 ml, was concentrated to prepare it for gel filtration, precipitates developed that were removed by filtration. At certain point, precipitation ceased and henceforth concentration could continue, what was suggestive of a complete elimination of excessive free protein. Precipitation could account for up to 40% (w/w) loss in protein, most of it Ets-1 (by analysis of the precipitate by SDS PAGE electrophoresis [2.3.5]). The ternary complexes could be concentrated up to 10-12 mg/ml (although E335 complexes were exceptionally soluble and could be concentrated up to 15 mg/ml). The end point for the concentration was the onset of new precipitates, this time from the complex, so it

corresponded to the solubility of the complex in GF buffer (2.1.9.2) at 4 ºC.

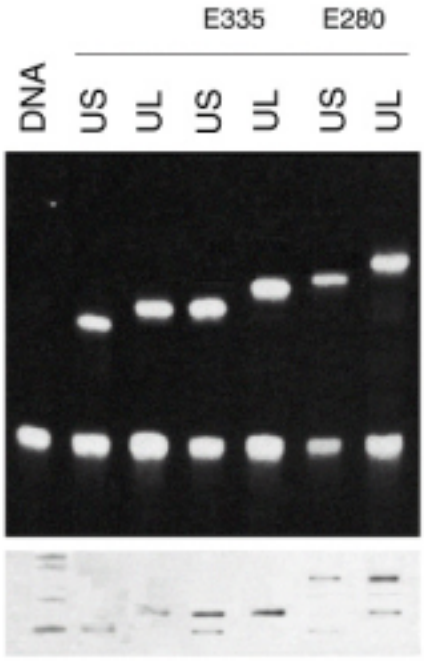


Fig 3.7 Native electrophoresis of the ternary complexes E280/LTR33 and E335/LTR33 with LUΔS and SUΔS. The top panel shows several Ets-1/USF1 complexes with biotinylated LTR33, on a 8% TBE native gel, developed by chemiluminescence (2.3.7). The bottom panel shows a Coomassie-stained SDS PAGE (2.3.5) gel with identical samples.

3.3.2.5 Complex homogeneity/polydispersity

The polydispersity of the Ets-1 ternary complexes was analysed by DLS at 20 ºC (2.3.16). Typically, 2 µl of purified, concentrated complex was diluted to 20-24 µl of GF buffer (2.1.9.2) to make up a 1 mg/ml complex solution. DLS analysis showed that most complexes were remarkably monodisperse (within 15-20%), while E335 complexes tended towards higher polydispersity indexes (25-30%) (see Table 3.6 and Fig 3.9). This phenomenon may have resulted from the absence of interactions between the N-terminal AI motif and USF1 and/or DNA, since it is absent in E335ΔS but not in all other complexes. The generally low polydispersity (always lower than 30%) ensures that the ternary complexes are the only species in solution, as suggested by gel filtration, and therefore rules out any dissociation of the complex at 20 ºC after gel filtration (at least in the time scale of the experiment setup, 10-20 min).

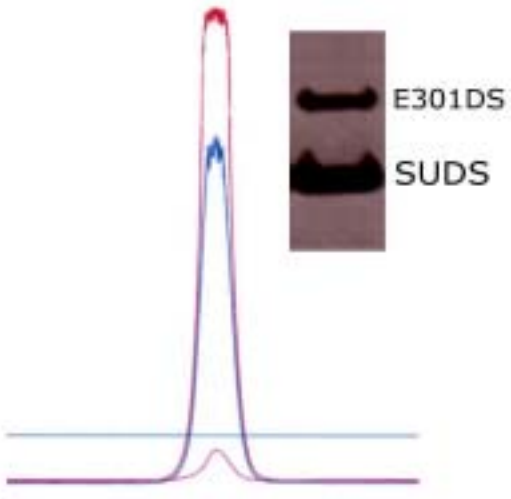


Fig 3.8 Gel filtration chromatography of E301ΔS/SUΔS/LTR33. The polishing step after any complex formation was performed by gel filtration chromatography (2.3.11). The ratio of the UV absorbance at 260 (red trace) and 280 nm (blue trace) confirmed the presence of a DNA complex under the peak, and SDS PAGE electrophoresis (2.3.5) established the presence of both E301ΔS and SUΔS (inlet). During concentration of the ternary complex for the gel filtration run, the precipitation and/or the escape through the concentrator's membrane of isolated components enriches the protein sample in ternary complex, and this is reflected upon in the absence of bands of excess components in the chromatograms (although this varies greatly depending on complex and preparation).

3.3.2.6 Crystallisation of Ets-1/USF1/DNA (Initial screenings)

The crystallisation of the ternary complexes (see Table 3.5 for the complete list of complexes) suffered from a shortage of protein supply, since complex production was laborious and yields extremely low. This has been partly the reason why complete optimisation of first crystallisation leads has not yet been completed (work in progress). Additionally, and to avoid a waste of material, high ionic strength conditions were ruled out from the beginning to avoid complex dissociation (which occurs at >350-400 mM NaCl).

Initial screening entailed 4x2x2 (16) combinations of Ets-1 (E244ΔS, E280ΔS, E301ΔS and E335ΔS), USF1 (LUΔS and SUΔS) and DNA (LTR33 and LTR32), at 10-12 mg/ml, in the sitting-drop vapour diffusion setup (2.3.17). Depending on sample availability, 6-8 commercial screens were setup per complex (1 μl/drop, a total of 200-300 drops/complex), choosing those screens tailored for conditions of low to moderate ionic strength. The selected screens comprised the Hampton Cryo Screen, Hampton Lite screen, the Wizard Cryo screens I and II, Hampton Low Ionic Strength screen, Hampton PEG/Ion screen and, when sample permitted it, the more traditional, higher ionic strength screens Hampton Crystal Screens I and II.

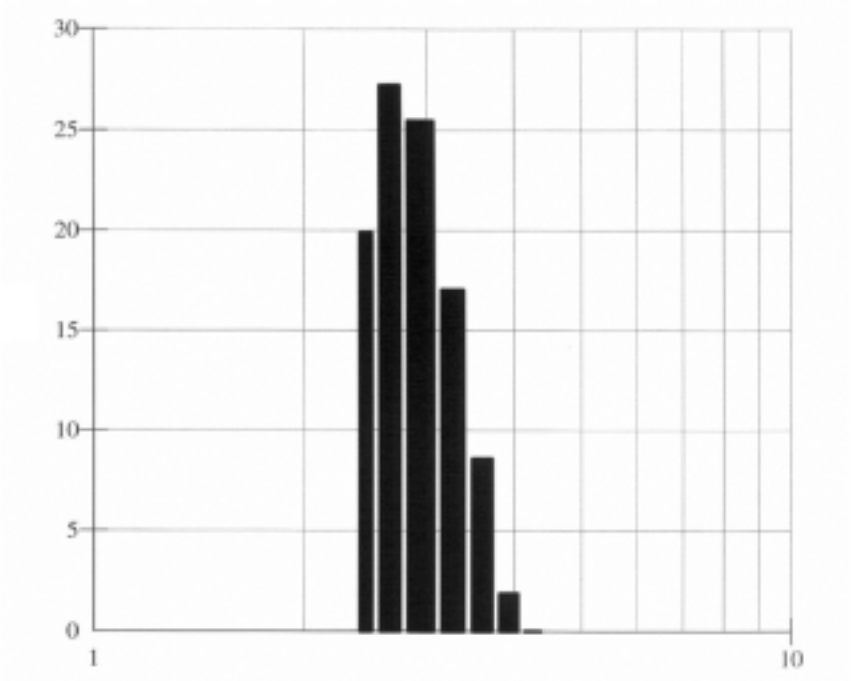


Fig 3.9 Dynamic light scattering (DLS) measurements of the complex E280ΔS/LUΔS/LTR33. Protein sample and measurements were done according to the procedures explained above (2.3.16). This complex is well behaved in comparison with most other related complexes, which have polydispersity indexes about 15-25%. Its hydrodynamic radius of 3.02 nm and calculated molecular weight, 59.4 kDa, agrees with the predicted value of 66.3 kDa.

Ets1/USF1/DNA	Polydispersity (%)
E244ΔS/LUΔS/LTR33, E244ΔS/LUΔS/LTR33	20.5/26.3
E244ΔS/LUΔS/LTR32, E244ΔS/LUΔS/LTR32	25.3/27.1
E280ΔS/LUΔS/LTR33, E280ΔS/LUΔS/LTR33	20.1/22.5
E280ΔS/LUΔS/LTR32, E280ΔS/LUΔS/LTR32	24.8/26.9
E301ΔS/LUΔS/LTR33, E301ΔS/LUΔS/LTR33	15.9/16.4
E280ΔS/LUΔS/LTR32, E280ΔS/LUΔS/LTR32	17.5/16.9
E335ΔS/LUΔS/LTR33, E335ΔS/LUΔS/LTR33	28.9/25.4
E335ΔS/LUΔS/LTR32, E335ΔS/LUΔS/LTR32	32.8/35.1

Table 3.6 Dynamic light scattering (DLS) polydispersity index for Ets-1/USF1/DNA ternary complexes. All DLS measurements were performed as described before (2.3.16), with 1 mg/ml complex solutions as obtained from the final gel filtration chromatography (2.3.11) after concentration. Polydispersity indices lay within the range 15-35%, with the highest values for E335ΔS complexes and the lowest for E301ΔS complexes. Although there seems to be a trend which suggests that LUΔS-containing complexes have lower polydispersity than the equivalent SUΔS-containing complexes, that tendency is reversed for E280ΔS/LUΔS/LTR32 and E335ΔS/LUΔS/LTR33.

Of the eight ternary complexes formed with LTR33, most drops yielded neither crystals nor birefringent, microcrystalline precipitates, staying clear at least during 2-3 months. Since in the majority of conditions the precipitating agent was based on low to high molecular weight polyethyleneglycols (PEG), the equilibration times for thorough mixing were fairly long. Later on, amorphous precipitates appeared in >50% of the drops evidencing protein denaturation. After 4-5 months, from E301ΔS/LUΔS/LTR33 drops there appeared rosettes made up by the stacking of thin plates in a star-like arrangement, under the conditions #40 (40% [v/v] ethanol, 0.1 M phosphate

citrate, pH 4.2, 5% [w/v] PEG3000) and #44 (40% [v/v] ethyleneglycol, 0.1 M HEPES, pH 7.5) of the Emerald Wizard Cryo I screen (Fig 3.10). On the other hand, from E335ΔS/LUΔS/LTR33 drops grew a fairly isolated crystal with thin plate morphology, surrounded by tiny precipitates, under condition #4 of the Emerald Wizard Cryo I screen (40% [w/v] PEG3000, 0.1 M HEPES, pH 7.5) (Fig 3.10). No other construct produced anything similar, either the drops stayed clear or protein developed into an amorphous precipitate.

The crystallisation setups of the remaining eight ternary complexes with LTR32 showed many more conditions with amorphous precipitates and fewer conditions stayed clear. Interestingly, the same constructs, E301ΔS/LUΔS/LTR32 and E335ΔS/LUΔS/LTR32, produced very similar crystals (in number, size and morphology) under exactly the same conditions that the corresponding LTR33 complexes. However, as with the LTR33 complexes, the number and size of the crystals was too small to attempt to characterise them.

Under condition #11 of the Hampton Crystal Screen Lite (0.2 M tri-Sodium citrate dihydrate, 0.1 M sodium cacodylate, 15% [v/v] 2-propanol, pH 6.5) and only for complex E301ΔS/SUΔS/LTR33, there appeared a shower of mycrocrystals (Fig 3.10). They were too small to permit their analysis by X-ray crystallography (~0.05 mm each side), but were produced in sufficient number for them to be analysed by SDS PAGE electrophoresis (2.3.5) to confirm their being protein. Fig 3.11 shows a non-denaturing PhastGel gel (Pharmacia) of the collected crystals (60-80 crystals across several drops) that provides evidence that those tiny crystals contain the ternary complex and in the expected stoichiometry.

3.3.3 Discussion

3.3.3.1 Complex formation

The stabilisation of the individual Ets-1 and USF1 proteins was a prerequisite for complex formation, which was achieved by mutation of all cysteines on Ets-1 (C261, C350 and C416) and USF1 (C229 and C248) into serines. Attempts to co-express Ets-1 and USF1 in *E coli* cells with a two-plasmid system (pACYC/pET system, 2.3.11) or in a cell-free translation system (2.3.10), intended to obtain a binary complex that were more stable than the individual proteins, failed to produce yields and/or a balanced expression of the two proteins.

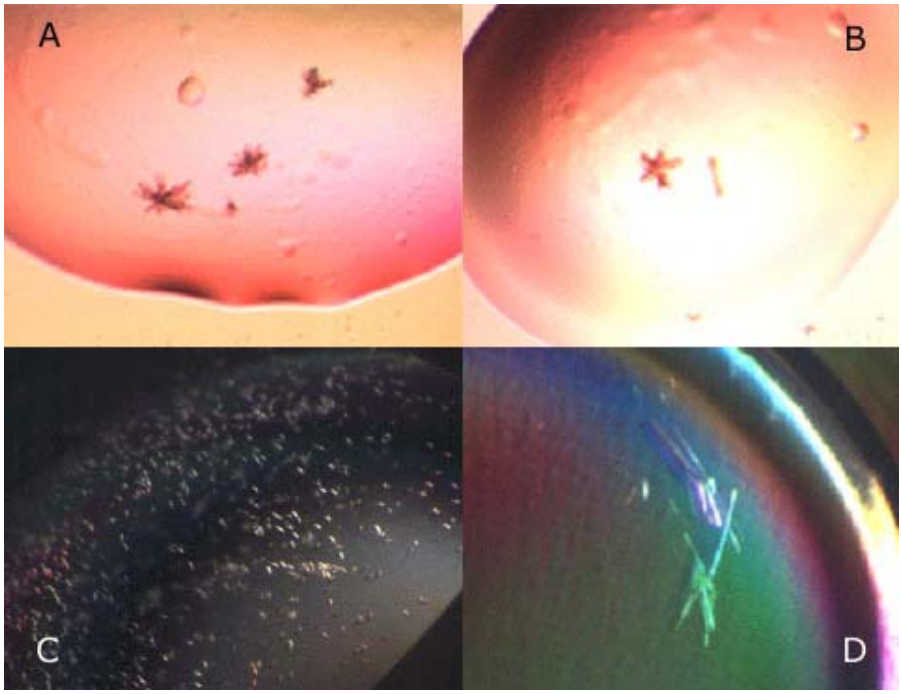


Fig 3.10 Preliminary crystal conditions for optimisation. Conditions #44 (A) and #40 (B) of the Wizard Cryo Screen I (Emerald) produced a few crystals of star- or rosette-like morphology for the E301ΔS/LUΔS/LTR33; in a drop containing E335ΔS/LUΔS/LTR33, condition #4 of the same kit generated thin plates that grew in close contact with one another (D). The complex E301ΔS/SUΔS/LTR33 developed a shower of microcrystals under condition #11 of the Crystal Screen Lite (Hampton), which were of insufficient size for crystallographic characterisation (C).

Although Ets-1 could be refolded by several methods (vg gel filtration, on-column refolding) (2.3.12), prompt precipitation by non-specific aggregation and tremendous loss of material discouraged those procedures. The method of choice was rather to achieve refolding of Ets-1 on a diluted solution of a pre-formed USF1/DNA complex by slow stepwise addition of an Ets-1 concentrated solution in 6 M GuHCl (2.3.12). This procedure had the advantage of being relatively efficient (greater than 50% [w/w] Ets-1 could be successfully refolded per assay) and producing the desired ternary complex. The success of this strategy suggests that the supply of the right biological context (here in the form of a complex able to recruit folded Ets-1) may be crucial for refolding proteins with multiple motifs and/or domains. It also resembles refolding of enzymes in the presence of their substrates, and could provide a useful generalisation in protein preparation from inclusion bodies.

After complex formation, concentration of the complex proved a very wasteful process since >25% (w/w) protein precipitated and needed to be removed to stop further precipitation. As most of the precipitates contained Ets-1, it seems that it is unfolded protein or excess folded protein which precipitates; this might be beneficial by increasing the purity of the ternary complex.

3.3.3.2 Biological activity

In order to proceed with the crystallisation of the complex, it was essential to prove that the correct complex could form and display the expected biological activity. EMSA (2.3.7, Fig 3.10) of several binary and ternary complexes showed convincingly that the complexes hold with the expected stoichiometry. LUΔS, which homotetramerises through its LZ domain in the absence of DNA (Ferre-D'Amare et al, 1994; Ferre-D'Amare et al, 1994), fails to do so in the presence of the LTR site and of Ets-1, suggesting that a conformational change may happen on USF1 in complex with Ets-1 and/or in complex with the LTR sequence.

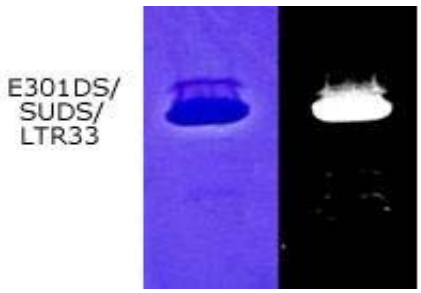


Fig 3.11 Non-denaturing PhastGel gel of crystals of E335ΔS/SUΔS/LTR33. More than 60 small crystals were collected from several crystallisation drops, washed and dissolved in a small amount of native electrophoresis loading buffer. Run was done according to a suitable program provided by the manufacturer. To locate protein and DNA, a double staining of Coomassie blue for protein and ethidium bromide or SYBR Green for nucleic acids was performed sequentially, which shows how protein and DNA are present in the same complex.

3.3.3.3 Homogeneity and polydispersity

Most of all ternary complexes (with the exception of the E335ΔS complexes) had low polydispersity (8-20%) and therefore did not aggregate or precipitate at 1 mg/ml. E335ΔS complexes, however, had slightly higher polydispersity of about 25-30%, which suggests that a broader distribution of particle sizes in solution could occur. Although the relationship between monodispersity (polydispersity below 30%) and crystallisability is not yet well established, there is partial evidence supporting that monodisperse solutions tend to favour crystallisation over polydisperse solutions, especially when precipitates or aggregates are present (Ferre-D'Amare & Burley, 1994).

3.3.3.4 Crystallisation of Ets-1/USF1/DNA complexes

The initial crystallisation screenings for all the sixteen combinations of Ets-1/USF1/DNA complexes were chosen to have low ionic strength to avoid dissociation of the complex, which limited the commercial screenings to the Hampton Cryo Screen, Hampton Screen Lite, Emerald Wizard Cryo Screens I and II, Hampton Low Ionic Strength Screen and Natrix Screen. In one case,

with E301ΔS/SUΔS/LTR33, a simple sparse matrix especially tailored for protein-nucleic acid complexes was also tried (see Fig 2.2). Crystals of E301ΔS/LUΔS/LTR33 and E335ΔS/LUΔS/LTR33 (and those of the LTR32 complexes) were found only after 4-5 months and under a very limited set of conditions, with either ethanol, ethyleneglycol, PEG3000 or 2-propanol as precipitating agents.

To analyse the condition where a shower of microcrystals appeared, 50-70 small crystals were collected from the drop, washed in fresh mother liquor, dissolved in non-denaturing sample buffer (3.1.9.2) and run on a Pharmacia PhastSystem to confirm the presence of DNA and protein in the complex. Fig 3.11 shows that the crystals contained a complex of molecular weight equivalent to that of the ternary complex (Ets-1, USF1 and DNA) and also DNA. Measurement of protein concentration in the mother liquor from the drops and careful manipulation of the crystals during pickup helped avoid cross-contamination of protein from the drop over to the crystals.

Although small crystals of similar size and quality were obtained under the same conditions from E301ΔS/LUΔS and E335ΔS/LUΔS complexes, the equivalent SUΔS complexes, E301ΔS/SUΔS and E335ΔS/SUΔS, developed thin, plate-like morphologies. None of the other Ets-1 proteins produced crystals.

Taken together, these data suggest a very narrow optimum for crystallisation of the Ets-1/USF1/DNA complex, where the protein construct (for both Ets-1 and USF1) and the DNA (since LTR33 rendered better crystals) bear importance, and where the range of compatible crystallisation conditions is narrow (organic solvents, low molecular weight PEG, neutral to acidic pH). Interestingly, most protein-DNA complexes tend to crystallise about neutrality (pH 6-8) and PEG is usually found as precipitant. On the other hand, organic solvents as ethanol, 2-propanol or ethyleneglycol are not frequently encountered in macromolecular crystallisation.

3.3.4 Summary

The 4x2x2 (16) ternary complexes Ets-1/USF1/DNA were reconstituted from the isolated proteins and DNA in a two-step procedure, whereby USF1 and DNA are mixed together to render a binary complex which is, after 2-3 h at 4 ºC, concentrated and buffer exchanged; subsequently, denatured Ets-1 is refolded by dilution in 20 ml of refolding buffer with 25-30 mM of USF1/DNA (2.3.12), concentrated and buffer exchanged into GF buffer (2.3.11). This last process is extremely wasteful in terms of protein material, since up to 40% (w/w) may be lost in precipitates. The resulting ternary complexes are stable under native electrophoresis (2.3.5) and gel filtration chromatography (2.3.11),

and forms a monodisperse solution by DLS (2.3.16). Furthermore, they can be concentrated up to 10-12 mg/ml without further precipitations.

Crystallisation of a series of Ets-1/USF1/DNA complexes (see Table 3.5) has been attempted with partial success. From the E301ΔS/LUΔS/LTR33 and E335ΔS/LUΔS/LTR33 complexes, crystals of reduced size have been grown under conditions dominated by organic solvents, low molecular PEG and neutral to acidic pH. Their morphologies are less than ideal (plates, rosettes). For the complex E301ΔS/SUΔS/LTR33, a shower of microcrystals of reduced size grew under 15% (v/v) 2-propanol as precipitant at pH 6.5, the analysis of which by native electrophoresis (2.3.5) permitted to establish that they contain protein and DNA (Fig 3.11).

Now the next step lies ahead where optimisation of the few initial, 'promising' conditions must be pursued to narrow down those conditions from where crystals of larger size and better diffraction properties can grow (work in progress).

3.4 References

Cowley, DO & Graves, B (2000) Phosphorylation represses Ets-1 DNA binding by reinforcing autoinhibition. *Genes Dev* **14**(3): 366-76.

Donaldson LW, Petersen JM, Graves BJ, and McIntosh LP (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J* **15**(1): 125-34.

Ducruix, A and Giege, R (1999) Crystallization of nucleic acids and proteins: A practical approach. 2nd ed, Oxford University Press.

Ferre-D'Amare AR and Burley SK (1994) Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. *Structure* **2**(5): 357-9.

Ferre-D'Amare AR, Pognonec P, Roeder RG, and Burley SK (1994) Structure and function of the b/HLH/Z domain of USF. *EMBO J* **13**(1): 180-9.

Garvie CW, Hagman J, and Wolberger C (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* **8**(6): 1267-76.

Jonsen MD, Petersen JM, Xu QP, and Graves BJ (1996) Characterization of the cooperative function of inhibitory sequences in Ets-1. *Mol Cell Biol* **16**(5): 2065-73.

Kim, YJ, et al (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**(4): 599-608.

Lim F, Kraut N, Framptom J, and Graf T (1992) DNA binding by c-Ets-1, but not v-Ets, is

repressed by an intramolecular mechanism. *EMBO J* **11**(2): 643-52.

Lionneton, F, Drobecq H, and Soncin F (2001) Expression and purification of recombinant mouse Ets-1 transcription factor. *Protein Expr Purif* **21**(3): 492-9.

Petersen JM, Skalicky JJ, Donaldson LW, McIntosh LP, Alber T, and Graves BJ (1995) Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. *Science* **269**(5232): 1866-9.

Pognonec P, Kato H, and Roeder RG (1992) The helix-loop-helix/leucine repeat transcription factor USF can be functionally regulated in a redox-dependent manner. *J Biol Chem* **267**(34): 24563-7.

Rice, P, Longden, I, and Bleasby, A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**(6): 276-277

Rost B, Sander C, and Schneider R (1994) PHD - an automatic mail server for protein secondary structure prediction. *Comput Appl Biosci* **10**(1): 53-60.

Vriend, G (1990) WHAT IF: a molecular modelling and drug design program. *J Mol Graph* **8**: 52-56.

Sieweke, MH, et al (1998) Cooperative interaction of ets-1 with USF-1 required for HIV-1 enhancer activity in T cells. *EMBO J* **17**(6): 1728-39.

Werner MH, Clore M, Fisher CL, Fisher RJ, Trinh L, Shiloach J, and Gronenborn AM (1995) The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation. *Cell* **83**(5): 761-71.

4. Appendix. Vector Maps

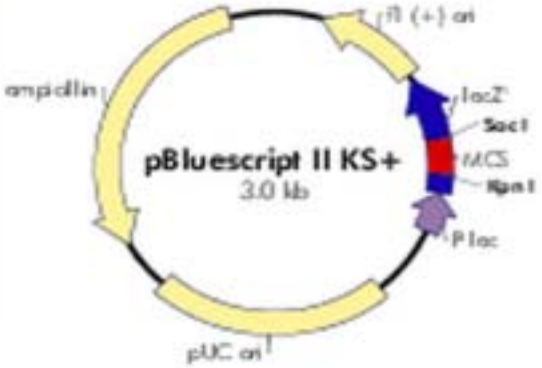
4.1 pCR-Blunt

Comments for pCR®-Blunt
3513 nucleotides

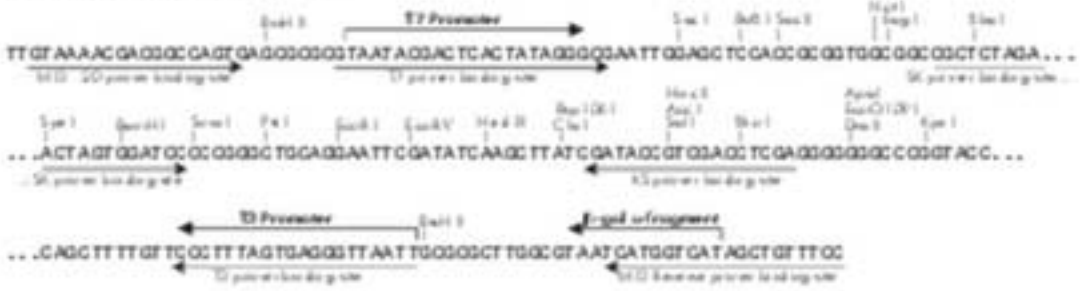
Lac promoter/operator region: bases 95-216
M13 Reverse priming site: bases 205-221
LacZ-alpha ORF: bases 217-570
Multiple Cloning Site: bases 248-393
T7 promoter priming site: bases 400-419
M13 Forward (-20) priming site: bases 427-442
Fusion joint: bases 571-579
ccdB lethal gene ORF: bases 580-882
Kanamycin resistance ORF: bases 1231-2025
Zeocin resistance ORF: bases 2232-2606
pUC origin: bases 2674-3387

4.2 pBluescript II KS+

f1 (+) origin 135–441
β-galactosidase α-fragment 460–816
multiple cloning site 653–760
***lac* promoter** 817–938
pUC origin 1158–1825
ampicillin resistance (*bla*) ORF 1976–2833

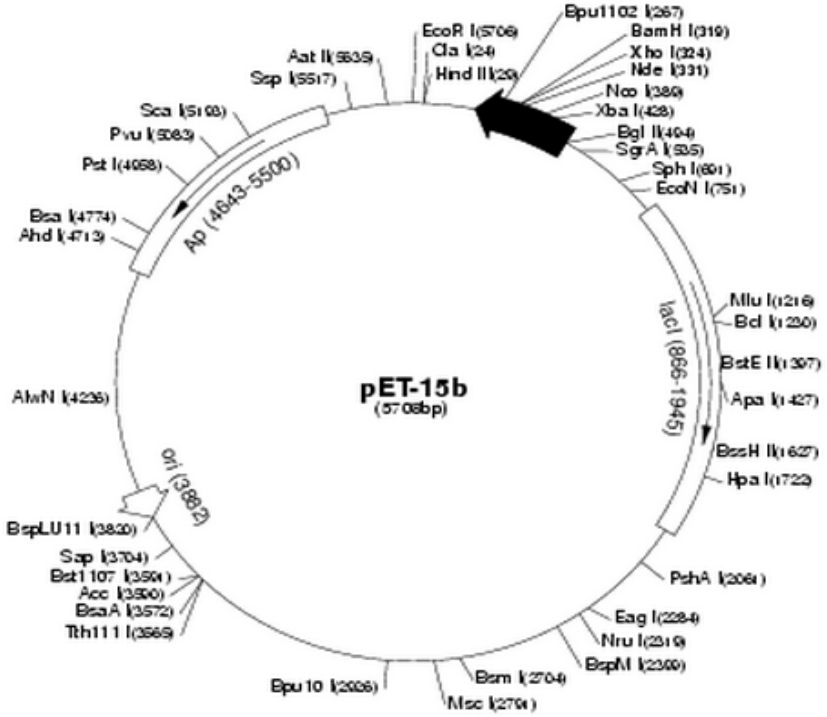


pBluescript II KS (+/–) Multiple Cloning Site Region
(sequence shown 598–826)



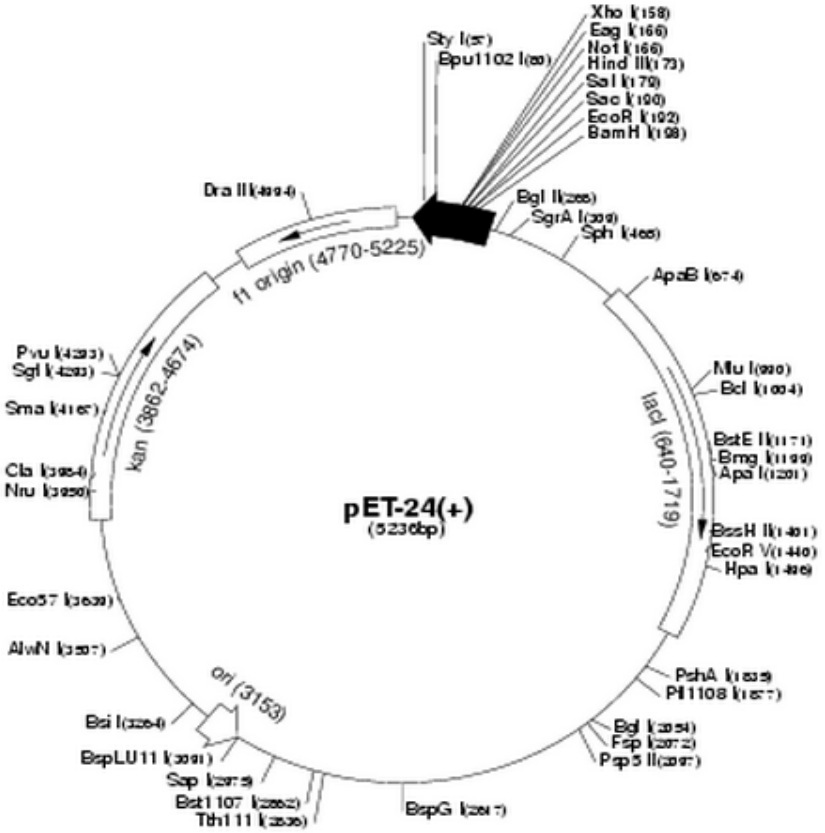
4.3 pET15b

pET-15b sequence landmarks	
T7 promoter	463-479
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites	
(*Nde* I - *Bam* H I)	339-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
bla coding sequence	4643-5500

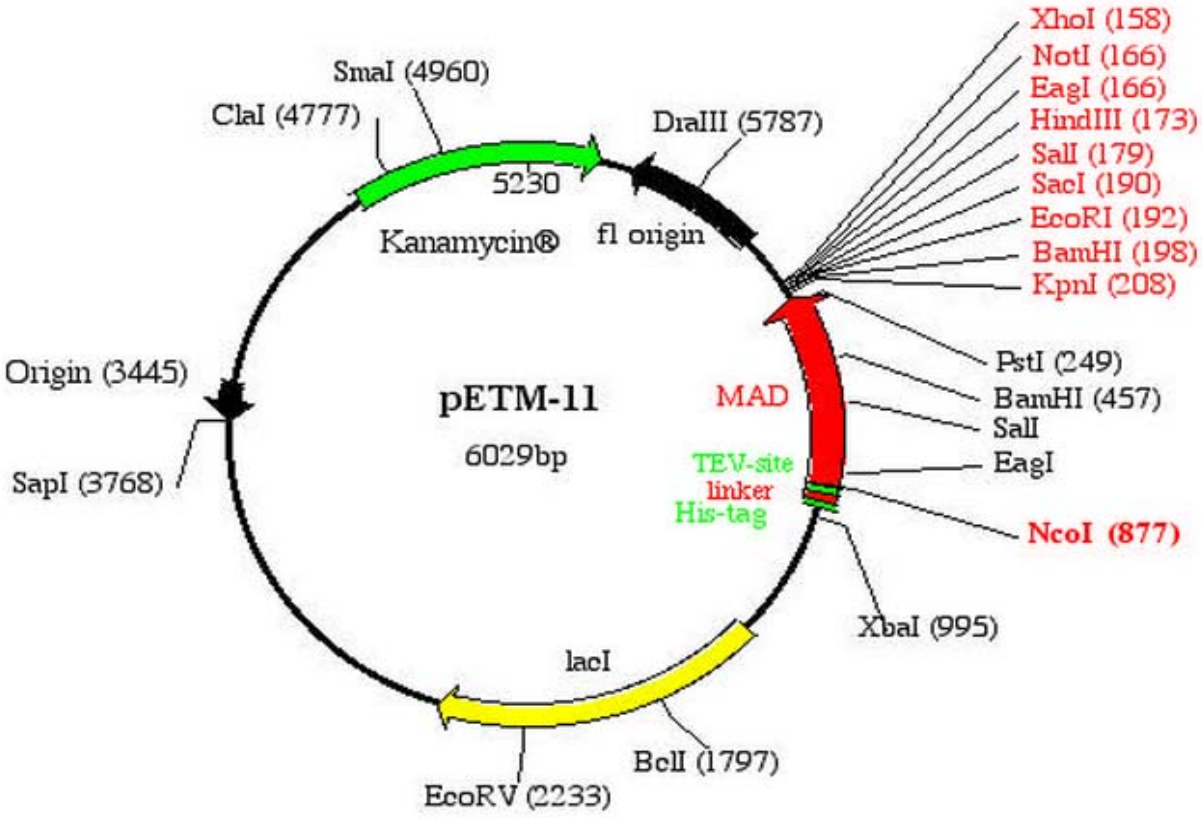


4.4 pET24d

pET-24(+) sequence landmarks	
T7 promoter	237-253
T7 transcription start	236
Multiple cloning sites	
(*Bam* H 1- *Xho* 1)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	640-1719
pBR322 origin	3153
Kan coding sequence	3862-4674
f1 origin	4770-5225

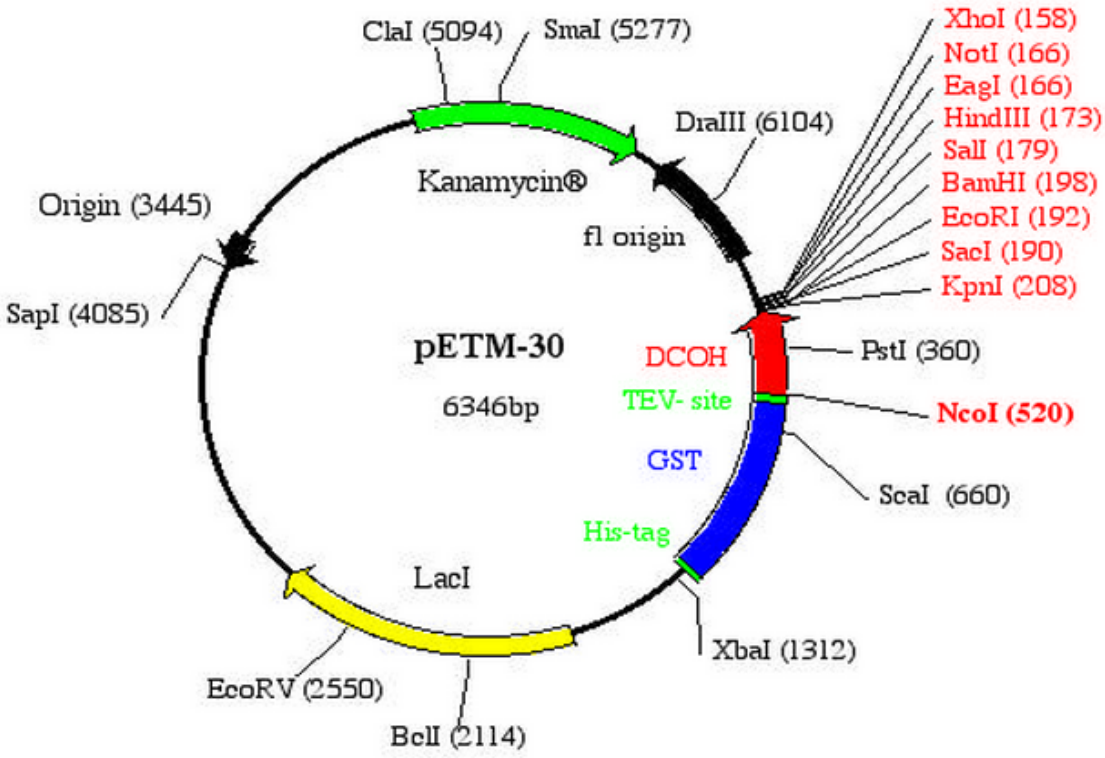


4.5 pETM11




```
    T7 promoter  -->                lac operator    XbaI
GAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAAT
CTTTAATTATGCTGAGTGATATCCCCTTAACACTCGCCTATTGTTAAGGGGAGATCTTTA

                        rbs                        His.Tag
AATTTTGATTTAACTTTAAGAAGGAGATATACCATGAAACATCACCATCACCATCACCCC
TTAAAACTAAATTGAAATTCTTCCTCTATATGGTACTTTGTAGTGGTAGTGGTAGTGGGG
                                  METLysHisHisHisHisHisHisPro

                                  TEV site      NcoI
ATGAGCGATTACGACATCCCCACTACTGAGAATCTTTATTTTCAG GGCGCCATGGCGGCG
TACTCGCTAATGCTGTAGGGGTGATGACTCTTAGAAATAAAAGTC CCGCGGTACCGCCGC
MetSerAspTyrAspIleProThrThrGluAsnLeuTyrPheGln|GlyAlaMETAlaAla

GCGGTTCGGATGAAC..612bp..GACAGTCACAAGGCGTGTCTTGGTCTCTAACTAGTG
CGCCAAGCCTACTTG...MAD...CTGTCAGTGTTCCGCACAGAACCAGAGATTGATCAC
AlaValArgMetAsn..204aa..AspSerHisLysAlaCysLeuGlyLeu***

                                        NotI
      BamHI       SacI                  EagI
KpnI        EcoRI        SalI HindIII        XhoI   His-tag
GTACCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCAC
CATGGCCTAGGCTTAAGCTCGAGGCAGCTGTTCGAACGCCGGCGTGAGCTCGTGGTGGTG
                                              HisHisHis
```

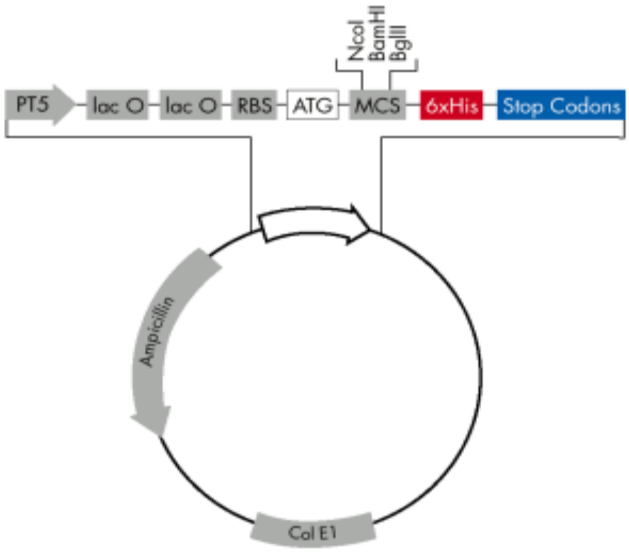
4.6 pETM30



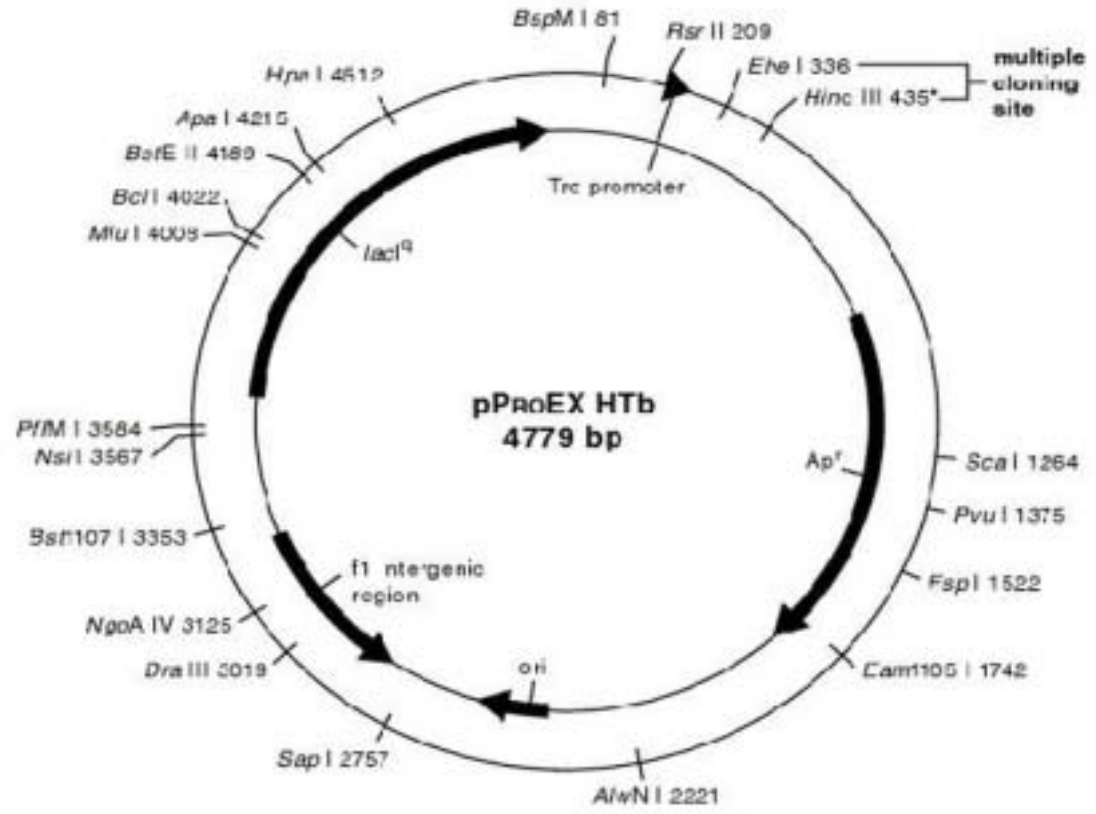

```
     T7 promoter  -->            Lac operator     XbaI
CGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTT
GCTTTAATTATGCTGAGTGATATCCCCTTAACACTCGCCTATTGTTAAGGGGAGATCTTTATTAAA

                rbs                      His-tag
TGTTTAACTTTAAGAAGGAGATATACCATGAAACATCACCATCACCATCACAACACTAGTAGCAAT
ACAAATTGAAATTCTTCCTCTATATGGTACTTTGTAGTGGTAGTGGTAGTGTTGTGATCATCGTTA
                          METLysHisHisHisHisHisHisAsnThrSerSerAsn


TTCATGTCC..633bp..GACCATCCTCCAACTAGTGGATCTGGTGGTGGTGGCGGATCCATGAGC
AAGTACAGG...GST...CTGGTAGGAGGTTGATCACCTAGACCACCACCACGCCTAGGTACTCG
PheMetSer..211aa..AspHisProProThrSerGlySerGlyGlyGlyGlySerMetSer

          TEV-site      NcoI
GAGAATCTTTATTTTCAG GGCGCCATGGCTGGCAAGCACACA..279bp..GCCGTGTCTATGACA
CTCTTAGAAATAAAAGTC CCGCGGTACCGACCGTTCGTGTGT...DCOH..CGGCACAGATACTGT
GluAsnLeuTyrPheGln|GlyAlaMETAlaGlyLysHisThr...93aa..AlaValSerMetThr

                                        NotI
                                        EagI
         BamHI        SacI
   KpnI        EcoRI          SalI HindIII          XhoI   C-His-tag
TAGGTACCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCAC
ATCCATGGCCTAGGCTTAAGCTCGAGGCAGCTGTTCGAACGCCGGCGTGAGCTCGTGGTGGTGGTG
***
```

4.7 pQE70

Positions of elements in bases	
Vector size (bp)	3431
Start of numbering at *Xho*I (CTCGAG)	1–6
T5 promoter/lac operator element	7–87
T5 transcription start	61
6xHis-tag coding sequence	133–150
Multiple cloning site	113–132
Lambda t_0 transcriptional termination region	173–267
rrnB T1 transcriptional termination region	1033–1131
ColE1 origin of replication	1608
β-lactamase coding sequence	3226–2366



4.8 pProExHTb



RBS HIS-TAG

```
ACAGGAAACAGACCATGTCGTACTACCATCACCATCACCATCACGATTACGATATCCCA
TGTCCTTTGTCTGGTACCACATGATGGTAGTGGTAGTGGTAGTGCTAATGCTATAGGGT
          MetSerTyrTyrHisHisHisHisHisHisAspTyrAspIlePro
```

EheI BamHI

TEV-site NcoI EcoRI StuI SalI

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

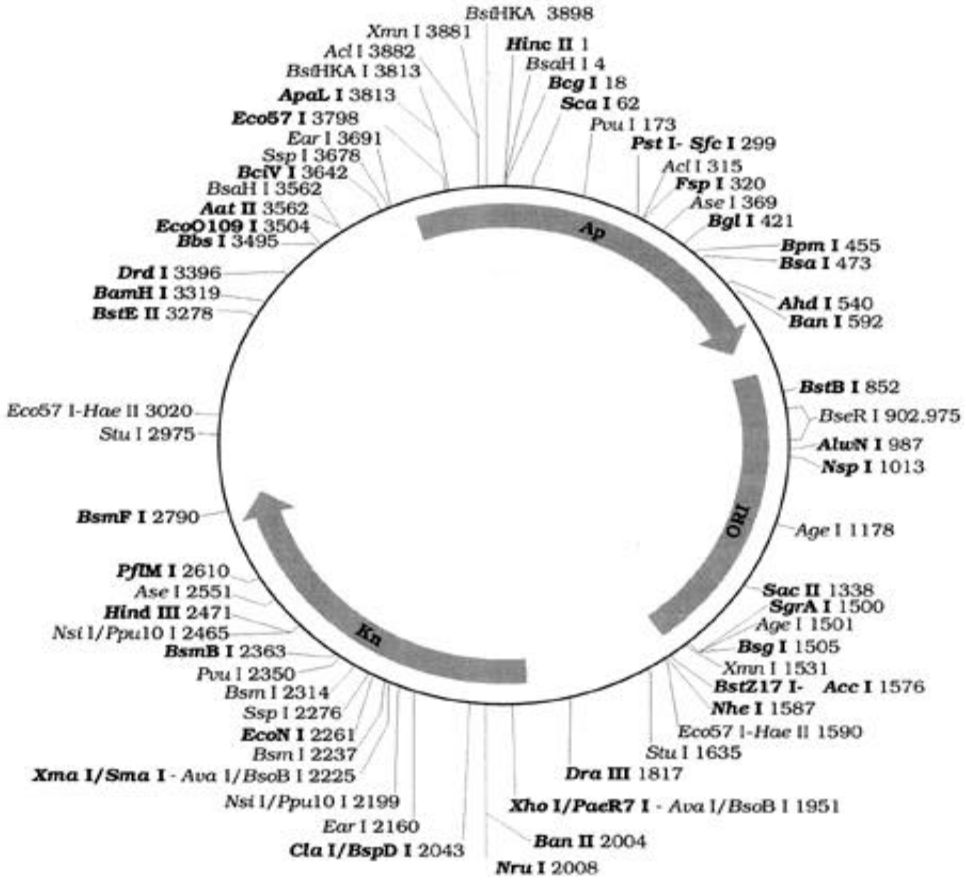
SstI SpeI NotI HspV XbaI PstI XhoI KpnI

```
ACGAGCTCACTAGTCGCGGCCGCTTTCGAATCTAGAGCCTGCAGTCTCGAGGCATGCGGT
TGCTCGAGTGATCAGCGCCGGCGAAAGCTTAGATCTCGGACGTCAGAGCTCCGTACGCCA
ThrSerSerLeuValAlaAlaAlaPheGluSerArgAlaCysSerLeuGluAlaCysGly
```

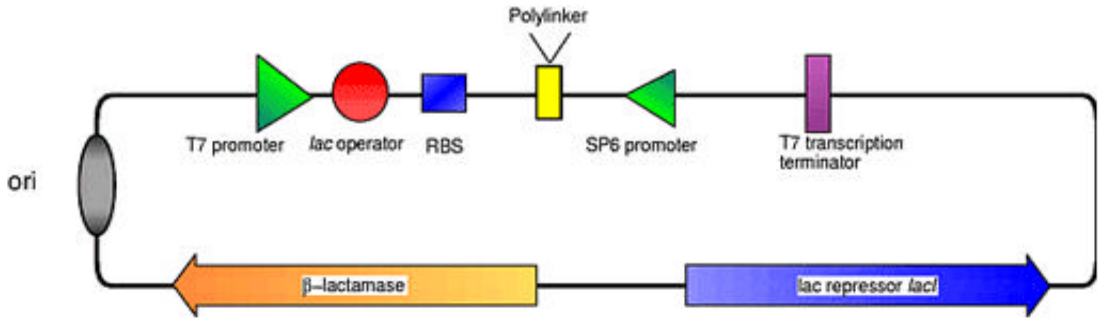
HindIII

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

4.9 pACYC9d


BsaHKA 3898
Xmn I 3881
Hinc II 1
Acl I 3882
BsaH I 4
BsaHKA I 3813
Bcg I 18
ApaL I 3813
Sca I 62
Eco57 I 3798
Pvu I 173
Ear I 3691
Pst I- Sfc I 299
Ssp I 3678
Acl I 315
BciV I 3642
Fsp I 320
BsaH I 3562
Ase I 369
Aat II 3562
Bgl I 421
EcoO109 I 3504
Bbs I 3495
Bpm I 455
Bsa I 473
Drd I 3396
BamH I 3319
Ahd I 540
BstE II 3278
Ban I 592
Ap
BstB I 852
Eco57 I-Hae II 3020
BseR I 902.975
Stu I 2975
AluN I 987
Nsp I 1013
ORI
BsmF I 2790
Age I 1178
PflM I 2610
Sac II 1338
Ase I 2551
SgrA I 1500
Hind III 2471
Age I 1501
Nsi I/Ppu10 I 2465
Bsg I 1505
Kn
BsmB I 2363
Xmn I 1531
Pvu I 2350
BstZ17 I- Acc I 1576
Bsm I 2314
Nhe I 1587
Ssp I 2276
Eco57 I-Hae II 1590
EcoN I 2261
Stu I 1635
Bsm I 2237
Dra III 1817
Xma I/Sma I - Ava I/BsoB I 2225
Nsi I/Ppu10 I 2199
Xho I/PaeR7 I - Ava I/BsoB I 1951
Ear I 2160
Ban II 2004
Cla I/BspD I 2043
Nru I 2008

4.10 pBAT4



Features	position
T7 promoter	1-17
Lac operator	21-42
T7 gene 10 ribosome binding site	50-80
Polylinker	83-152
SP6 promoter	157-175
T7 transcription terminator	254-295
LacI gene	705-1786 (reversed)
β-Lactamase gene	2556-3406
Origin of replication	3481-4176

```
  T7 promoter  -->          Lac operator      XbaI
ATTAATACGACTCACTATAGGGAATTGTGAGCGGATAACAATTCCTCTAGAAATAATTTT
TAATTATGCTGAGTGATATCCCTTAACACTCGCCTATTGTTAAGGAGATCTTTATTAAAA

                rbs          NcoI        EcoRI SalI  XhoI  BamHI
GTTTAACTTTAAGAAGGAGATATATCCATGGATATCGAATTCGTCGACCTCGAGGGATCC
CAAATTGAAATTCTTCCTCTATATAGGTACCTATAGCTTAAGCAGCTGGAGCTCCCTAGG
                           METAspIleGluPheValAspLeuGluGlySer

       XbaI    NotI   NsiI  HindIII   <--    SP6 promoter
GGGCCCTCTAGATGCGGCCGCATGCATAAGCTTGAGTATTCTATAGTGTCACCTAAATCC
CCCGGGAGATCTACGCCGGCGTACGTATTCGAACTCATAAGATATCACAGTGGATTTAGG
GlyProSerArgCysGlyArgMetHisLysLeuGluTyrSerIleValSerProLysSer
```