Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization of redox-insensitive Oct1 POU domain with different DNA-response elements

Attila Reményi,[a] Ehmke Pohl,[a] Hans R. Schöler[b] and Matthias Wilmanns[a]*

[a]EMBL, Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, and [b]Center for Animal Transgenesis and Germ Cells Research, School of Veterinary Medicine, Department of Animal Biology, University of Pennsylvania, New Bolton Center, Kennett Square, PA 19348, USA

Correspondence e-mail: wilmanns@embl-hamburg.de

Received 17 April 2001
Accepted 2 July 2001

The POU domain of the human Oct1 transcription factor has been crystallized with two different POU-dimer-binding DNA elements. Protein–DNA cocrystals suitable for structural analysis could be obtained only with a redox-insensitive version of the POU domain. The recombinant protein expression in a prokaryotic host was adjusted for fast purification. Optimized crystals were obtained by systematically varying the length of the oligonucleotide and by modifying cryofreezing procedures. These steps are generally applicable to the preparation of protein–DNA complexes for structural studies.

1. Introduction

Many biological processes are regulated by the activity of gene products involved in specific interactions between proteins and nucleic acids. Therefore, analysis of the biological implications of these interactions has been the subject of a variety of functional and structural studies over many years. The current annotation of the human genome sequence indicates that 13.5% of the coded gene products could have a nucleic acid binding function (Venter *et al.*, 2001). In contrast, protein–nucleic acid complexes currently deposited in the Protein Data Bank represent only 4.1% (600 entries out of 14 357) of the solved macromolecular structures. These numbers indicate that complexes of proteins and nucleic acids are still underrepresented by about threefold, presumably because, on average, their structures are still more difficult to solve than those of proteins.

Here, we report a case study of the engineering and production of a redox-sensitive transcription factor of the Oct family. Members of this family play a pivotal role in transcriptional regulation of developmentally important genes in mammals (Ryan & Rosenfeld, 1997). Biochemical and structural prediction data indicated that the POU factors of the Oct family can bind to DNA in two different arrangements, depending on the type and the spacing of the specific response motifs (Botquin *et al.*, 1998). The activity of these factors is differentially regulated by conformation-specific binding to coactivators (Tomilin *et al.*, 2000). We demonstrate how the specific optimization procedures of techniques for heterologous expression, crystallization and crystal stabilization under cryoconditions were necessary for rendering initially poor crystals into crystals suitable for X-ray analysis. The key step was to eliminate the reactivity of the two solvent-exposed cysteines. The reported optimization procedures are generally applicable to the functional and structural analysis of protein–nucleotide complexes.

2. Materials and methods

2.1. Protein expression and purification

The POU domain of Oct1 was PCR amplified from an eukaryotic expression vector containing the human cDNA of Oct1 using the following primers: 5′-TTTCCATGGAGGAGCCCAGTGACCTTGAGGAG-3′ and 5′-TAATGTGCGGCCGCTCAGTTGATTCTTTTCTCCTTCTGGCG-3′. The amplified insert was cleaved by *Nco*I and *Not*I and ligated into a modified version of the pET24d vector. This vector allows the expression of inserts with an N-terminal histidine tag for affinity purification that can be removed by the TEV protease (Parks *et al.*, 1994). Protein expression was induced by 1 m*M* IPTG in Epicurian Coli BL21-Codon Plus (DE3)-RIL cells (Stratagene) at 303 K for 3 h. Cells were harvested by centrifugation and lysed in buffer *A*1 (20 m*M* HEPES pH 7.6, 150 m*M* NaCl, 5 m*M* β-mercaptoethanol, 10 m*M* imidazole). The cell lysate was applied to an Ni-NTA agarose column and washed extensively with buffer *A*2 (20 m*M* HEPES pH 7.6, 150 m*M* NaCl, 5 m*M* β-mercaptoethanol, 35 m*M* imidazole), then eluted with buffer *A*3 (20 m*M* HEPES pH 7.6, 150 m*M* NaCl, 5 m*M* β-mercaptoethanol, 300 m*M* imidazole). TEV protease (50 μg ml^{-1}) was added to the eluted protein solution and this solution was dialysed against buffer *A*4 (20 m*M* HEPES pH 7.6, 100 m*M* NaCl, 5 m*M* β-mercaptoethanol) at 277 K overnight. To remove the histidine-tagged protease and the uncleaved product, the affinity matrix was equilibrated with buffer *A*4 and the dialysed

protein solution was allowed to flow through the column. In the cleaved polypeptide chain, a Gly-Ala-Met peptide remains N-terminal to the first glutamate residue of the POU domain. The correct molecular weight of this protein, 18 620 Da, was confirmed by mass-spectrometry analysis (data not shown). As a second purification step, a Hi-Trap heparin-Sepharose column was equilibrated with buffer *B*1 (20 m*M* HEPES pH 7.6, 100 m*M* NaCl, 1 m*M* EDTA), the sample was loaded and then eluted with buffer *B*2 (20 m*M* HEPES pH 7.6, 1 *M* NaCl, 1 m*M* EDTA). The protein eluted at around 300 m*M* NaCl. In the final step, the sample was diluted 2.5-fold by the addition of double-distilled H_2O in order to decrease the salt concentration and concentrated with a 10K Centricon membrane to a protein concentration of 8 mg ml^{-1}. The oligonucleotides were purchased from Metabion, Munich and purified by reverse-phased chromatography. The single-stranded DNA was annealed in buffer *C* (10 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 10 m*M* $MgCl_2$) first by heating the solution to 368 K and then cooling it down slowly to room temperature in a water bath. The double-stranded DNA was desalted on a C18 cartridge and lyophilized in aliquots. The protein–DNA complexes were formed by adding protein solution to lyophilized aliquots of the oligonucleotide in a molar ratio of 1:1.5.

2.2. Dynamic light-scattering measurement

The data was measured with a DynaPro 99 DLS-instrument and analysed using the *DYNAMICS* Version 5.24.02 from Protein Solutions software. Polydispersity was defined as (standard deviation of the particle size)/(average particle size), assuming a monomodal size distribution of the system. The protein concentration was about 1 mg ml^{-1} in a solution of 150 m*M* NaCl, 20 m*M* HEPES pH 7.0.

2.3. Crystallization

The initial sparse-matrix screens for crystallization contained polyethylene glycol (PEG) solutions as precipitating agents with different molecular weights (PEG 400 to PEG 8000) at three different pH values (5.5, 7.0 and 8.0). Organic solvents (ethanol, 2-propanol, dioxane, butandiol and hexanediol), divalent cations (Ca^{2+} and Co^{2+}) and polyamines (spermine, spermidine) were employed as additives. The crystallization trials were carried out in hanging drops by vapour diffusion with a drop size of 2 µl at 293 K.

2.4. Crystal stabilization for cryofreezing

The crystals were cross-linked by a gentle vapour-diffusion glutaraldehyde treatment according to the procedure described by Lusty (1999). The cover slips containing the crystallization droplets were transferred over a sitting-drop well containing the cystallization solution and 5 µl of glutaraldehyde solution (from Sigma). Diffusion of glutaraldehyde into the crystallization drop was allowed for 30 min. The crystals were transferred by a loop into a solution identical in composition to the crystallization solution but containing 20% glycerol as a cryoprotectant and were finally flash-frozen in a nitrogen stream at 100 K.

The POU–PORE crystals were transferred from the crystallization droplet into 50 µl of a mixture of two polyfluorinated oils (Kottke & Stalke, 1993). After the removal of the crystallization solution from the surface, the crystals were directly flash-frozen.

2.5. X-ray diffraction experiments

The diffraction experiments were carried out at the synchrotron beamlines at EMBL/DESY, Hamburg using either a MAR 345 mm imaging-plate scanner or a MAR 165 mm CCD detector at cryogenic temperature (100 K). Test images were indexed and integrated with the *HKL*2000 program suite (Otwinowski & Minor, 1997). The estimated mosaicity values were taken from the output of the scaling program *SCALEPACK*.

3. Results and discussion

3.1. Choice of the protein construct for structural studies

The aim of our study was to provide a structural basis for the differential dimeric arrangement of the POU domains of the Oct1 transcription factor bound to two different DNA-response elements, termed PORE (ATTTGAAAGGCAAAT) and MORE (ATGCATATGCAT) (Tomilin *et al.*, 2000). The Oct1 constructs were chosen to cover the DNA-binding POU-factor domains (Fig. 1) but not the flanking N-terminal and C-terminal regions of Oct1.

3.2. Optimization of recombinant protein expression

Since the usage of synonymous codons varies between different species, the expression of eukaryotic genes in a prokaryotic host (for example in *E. coli*) is frequently affected by the presence of codons rare to the host genome. Particularly in the case where they occur in consecutive sequence, they may cause the ribosome to terminate translation or trigger a +1 frame shift (Rosenberg *et al.*, 1993).

In the POU domain of the mammalian Oct1 transcription factor, residues Arg100 and Arg101 (numbered from the N-terminus of the construct used) are coded by two consecutive rare arginine codons (AGG-AGG). Termination of translation at this sequence would produce an intact POU-specific N-terminal domain, which has been reported to retain DNA-binding activity (Klemm & Pabo, 1996) without the C-terminal POU-homeo domain. Indeed, the SDS–PAGE from the cell lysate after IPTG induction (Fig. 1, lanes 1 and 2) showed a truncated product corresponding to about half the size of the intact POU

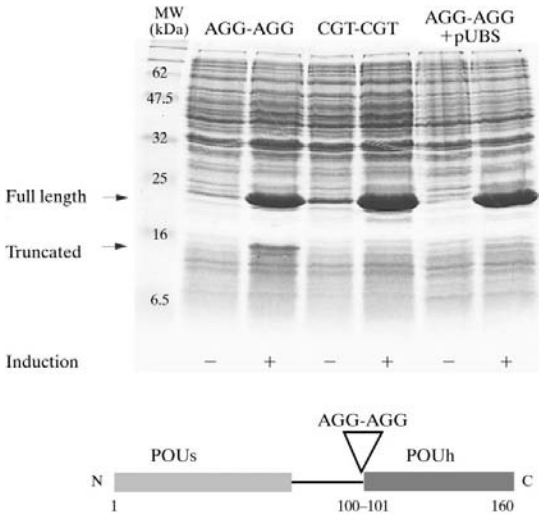

Figure 1
Effect of consecutive rare arginine codons on the expression of the POU domain of human Oct1 in *E. coli*. Bacterial cell lysate before and after induction with IPTG was loaded on SDS–PAGE. In lanes 1 and 2, the wild-type POU domain with the AGG-AGG codons was expressed. In lanes 3 and 4, the rare codons were mutated to the frequently used CGT arginine codons. In lanes 5 and 6, the expression plasmid containing the wild-type POU domain was cotransformed with another plasmid (pUBS) containing the rare $tRNA_{arg4}$ gene responsible for recognition of the AGG arginine codon. Bottom: schematic presentation of the sequence of the POU factor domain of mammalian Oct1. The POU domain is comprised of the POU-specific domain (POUs, 75 amino acids), POU homeo domain (POUh, 60 amino acids) and an unstructured linker sequence (24 amino acids). The position of the rare arginine tandem codons (AGG-AGG) is indicated.

domain (Fig. 1, lanes 1 and 2). However, it was not possible to purify the full-length construct from this mixture to homogeneity because of the different molecular-weight bands. To test the hypothesis of rare codon-induced termination (*a*) the two AGG codons were mutated to CGT, which is frequently found in *E. coli*, and (*b*) a plasmid (pUBS) containing an IPTG-inducible rare tRNA gene was cotransformed into the expression host (Schenk *et al.*, 1995). The truncated product was no longer present in either of the cases (Fig. 1, lanes 4 and 6), confirming our assumption. The second option involved the use of an *E. coli* strain with the rare tRNA encoded in its genome, which was applied for scaling up protein expression. A two-step purification scheme for the histidine-tagged POU domain, using a nickel-affinity matrix and an ion-exchange chromatography step, was sufficient to purify the protein to at least 99% homogeneity (Fig. 2, lane 1).

3.3. Monodisperse protein solution after addition of DNA

The purified protein sample, loaded on a native gel under low-salt conditions, appeared to be aggregated in the absence of bound oligonucleotides. This aggregation was confirmed by gel filtration and by dynamic light-scattering (DLS) experiments. Addition of stoichiometric amounts of double-stranded DNA to the protein solution eliminated the protein aggregation and transformed the solution into a system with a low polydispersity value of 0.1 (calculated with the DynaPro software). Values under ~0.3 are positively correlated with a high chance of crystallization (Ferre-D'Amare & Burley, 1997).

3.4. Engineering an oxidation-insensitive protein–DNA complex

In the initial crystallization trials of POU–DNA complexes extensive amorphous transparent protein precipitation, sometimes described as 'skin', was observed at the solvent–air interface. SDS–PAGE and native gels with samples taken from a number of crystallization drops indicated that the DNA remains soluble, whereas most of the protein was found in the amorphous transparent precipitate. Higher molecular-weight bands observed in non-reducing SDS gels suggested non-native disulfide bond formation by gradual oxidation of the cysteines (Fig. 2).

From these observations, we concluded that the protein–DNA complex did not remain stable over the long period of the crystallization experiments (2–3 d for the POU–MORE and 2–3 weeks for the POU–PORE complex). Inspection of a previous structure of the Oct1 POU domain complexed with the DNA octamer motif as a monomer (Klemm *et al.*, 1994) reveals that both cysteine residues, Cys61 and Cys150, are highly solvent-exposed in the absence of DNA (Fig. 2*b*). Furthermore, Cys150 is surrounded on the surface by two positively charged residues (Arg143 and Arg156). This local positively charged environment could stabilize the thiolate form of these cysteines, which could lead to an increased rate of disulfide bridge formation (Snyder *et al.*, 1981). Cys150 is conserved in all POU domains known so far (Herr & Cleary, 1995), but it has been shown to be dispensable for DNA binding when it was exchanged with a serine residue (Rigoni *et al.*, 1993). Therefore, these two cysteines were replaced by serines using site-directed mutagenesis and the mutant protein was expressed and purified as described for the wild-type sequence. We confirmed the functional integrity of the double mutant, which was subsequently used for crystallographic studies, by measuring its binding capacity to the DNA motifs in a gel-retardation assay. The binding remained unchanged with respect to the wild-type protein that was used as reference (data not shown).

This mutant stayed soluble in the crystallization drops and remained bound to DNA even after more than a month; no covalently linked multimer formation was observed in SDS–PAGE. Protein–DNA cocrystals with the MORE and PORE motifs suitable for diffraction analysis were obtained only with this modified protein. Similar observations were previously reported for the DNA-binding Runt domain, where a cysteine-to-serine mutation rendered the otherwise *in vitro* sensitive

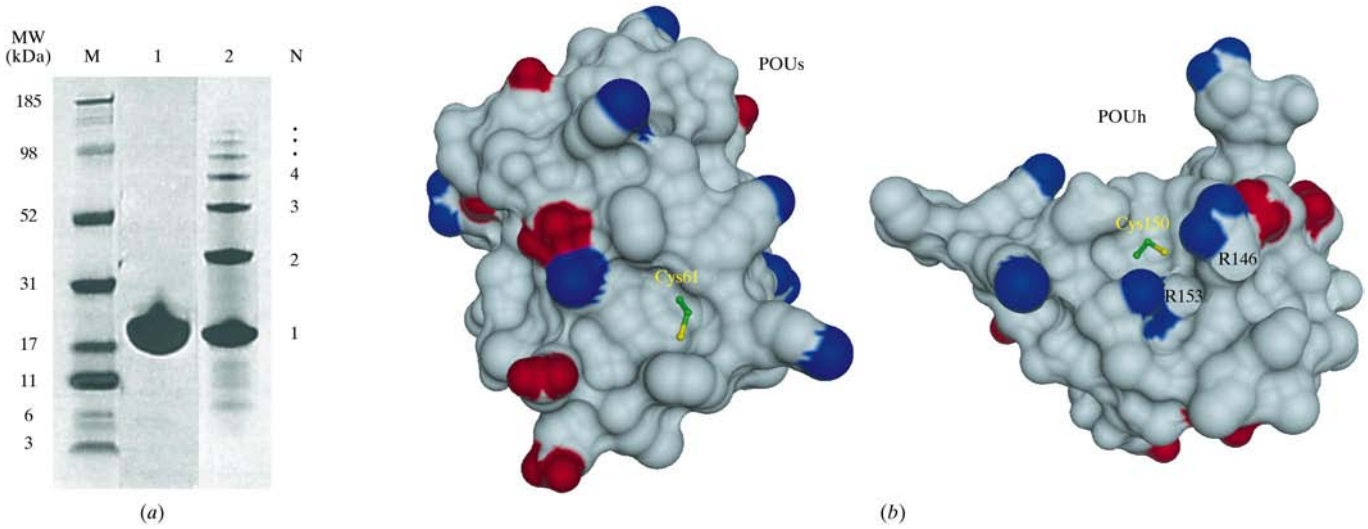

Figure 2
Cysteine oxidation in the POU domain causes covalent linkage of POU-domain monomers. (*a*) SDS–PAGE analysis: lane 1, freshly purified wild-type POU domain prior to crystallization trial; lane 2, crystallization drop containing stoichiometric amounts of wild-type Oct1 POU factor and the PORE oligonucleotide after 3 d (in PEG 3350 pH 5.0 and 10 m*M* β-mercaptoethanol). M, marker lane; N, estimated oligomerization state. (*b*) Positions of Cys61 and Cys150 in the POU-specific and POU-homeo domains of Oct1. The images were produced with *DINO* (Phillipsen, 2001). The molecular surfaces were calculated for each domain by the program *MSMS* with a probe radius of 1.4 Å, using the coordinates of a POU monomer of Oct1 (PDB code 1oct), neglecting Cys61 and Cys150 from the calculation. Positively (Lys NZ, Arg NH1 and NH2) and negatively (Glu OE1 and OE2, Asp OD1 and OD2) charged atoms of basic and acidic side chains are coloured blue and red, respectively. POUs, POU-specific domain; POUh, POU-homeo domain.

Table 1
Summary of the final crystallization conditions and crystal parameters.

	POU–MORE	POU–PORE
Crystallization conditions	22–24% PEG 3350, 50 m*M* HEPES pH 7.0, 1.8 m*M* spermine	20–23% PEG 1000, 100 m*M* Na citrate pH 5.3, 5%(*v*/*v*) glycerol
Length of DNA	21 bp, TA overhang	22 bp, blunt end
Crystal size (mm)	0.3 × 0.3 × 0.3	0.7 × 0.5 × 0.4
Space group	*C*2	$P6_222$
Unit-cell parameters (Å, °)	*a* = 93.3, *b* = 52.4, *c* = 69.0, *β* = 127.6	*a* = 131.2, *c* = 116.8
Max. resolution (Å)	1.9 (3.2†)	2.7 (3.2†)
Mosaicity (°)	0.5 (∼2†)	0.7 (∼1.5†)
Method for cryofreezing	Glutaraldehyde cross-linking + 20% glycerol	Polyfluorinated polyether oil

† The best value obtained with conventional soaking into different cryosolutions; the mosaicity values were estimated from the analysis of diffraction images after scaling.

protein oxidation insensitive (Nagata *et al.*, 1999).

3.5. Optimized crystal growth by varying DNA length

It is well known that the length of the oligonucleotide greatly influences protein–DNA cocrystal formation (Jordan *et al.*, 1985; Schultz *et al.*, 1990). Since there is no generally applicable rationale for optimum length and terminal structure of the oligonucleotide, we have chosen a systematic approach by setting up crystallization trials in parallel with oligonucleotides differing in length and terminus (blunt-end or overhang).

Blunt-ended oligonucleotides containing even numbers of base pairs (20–28) were initially used in crystallization trials for both complexes. It was noticed that oligonucleotides shorter in length displayed a higher propensity for microcrystal formation under the crystallization conditions we used. The oligonucleotide length was further refined by using DNA molecules of 20–23 base pairs with or without a TA overhang. In both cases, the complex with optimized oligonucleotide length yielded crystals from several conditions of the sparse-matrix screen, albeit with different morphologies (Table 1). The complex crystals used for X-ray data collection are shown in Fig. 3.

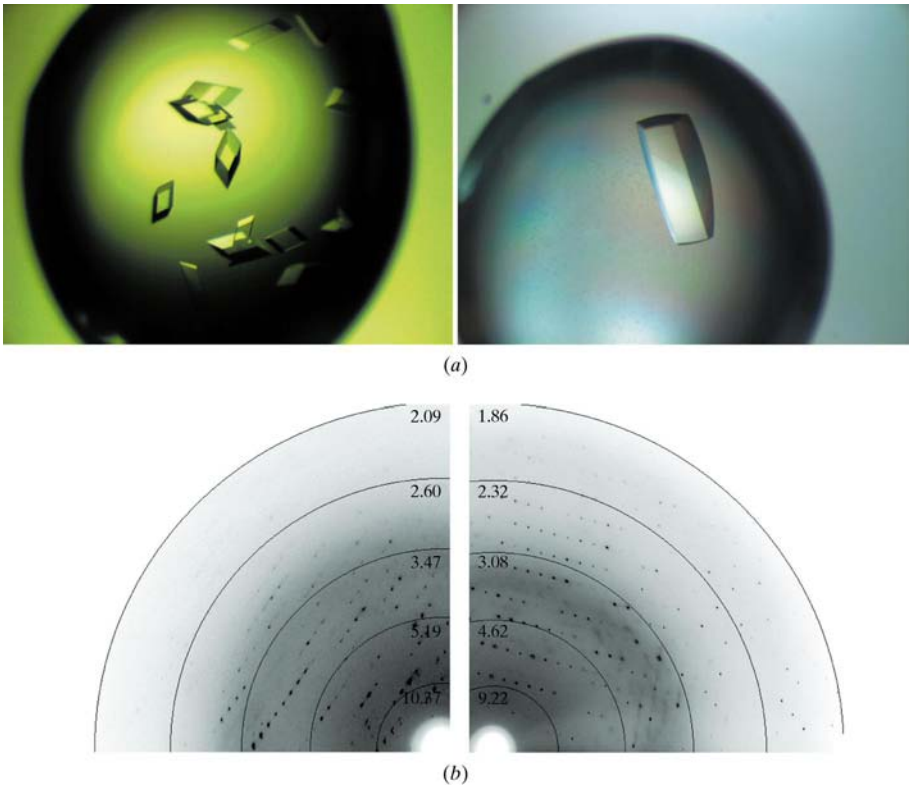

Figure 3
X-ray suitable crystals of the Oct1 POU factor (C61S, C150S) mutant in complex with the MORE and PORE motifs. (*a*) Morphology and size of the crystals containing the POU–MORE complex (right) and POU–PORE complex (left). (*b*) Improvement of the diffraction properties of the POU–MORE crystals by glutaraldehyde cross-linking: left, before and right, after glutaraldehyde treatment. Concentric circles indicate the resolution of the diffraction data. Further data describing these crystals are summarized in Table 1.

3.6. Optimized X-ray diffraction under cryofreezing

X-ray data collection carried out at cryogenic temperature reduces the radiation damage of sensitive macromolecular crystals (Hope, 1988). However, the addition of cryoprotectants frequently reduces or even abolishes their diffraction properties (Garman & Schneider, 1997). Upon soaking into a broad variety of cryosolutions with different cryoprotection protocols, the POU–DNA cocrystals consistently developed cracks and deteriorated. In a few cases, the crystals diffracted to about 3–4 Å resolution, displaying large and distorted diffraction spots with an estimated mosaicity of more than 1°. X-ray exposure of fresh crystals at room temperature permitted diffraction to higher resolution (<3 Å) and less distorted diffraction spots. This indicated that the use of these conventional cryofreezing methods caused the diffraction quality of the crystals to decline, possibly owing to osmotic or mechanical stress.

Therefore, other methods to further improve their mechanical robustness were investigated. This could be achieved for the POU–MORE complex but not for the POU–PORE crystals by cross-linking them with glutaraldehyde using a gentle vapour-diffusion technique (Lusty, 1999). The crystals not only retained their sharp edges without any visible fractures in the cryosolution containing 20%(*v*/*v*) glycerol in addition to the mother liquor, but also showed increased diffraction to about 1.9 Å resolution (Fig. 3). Their mosaicity value was reduced to about 0.5°. A similar method of glutaraldehyde cross-linking was successfully applied to the protein–DNA complex of Pit-1 (Jacobson *et al.*, 1996).

The diffraction properties of the POU–PORE crystals during cryomounting could be maintained by transferring them into polyfluorinated polyether oil mixtures (Kottke & Stalke, 1993). The mosaicity of these crystals treated this way was about 0.7°; they diffracted to 2.7 Å resolution.

We are very grateful to Gunter Stier for providing us with bacterial expression vectors and the pUBS plasmid as well as for his valuable advice in recombinant protein-expression techniques.

References

Botquin, V., Hess, H., Fuhrmann, G., Anastassiadis, C., Gross, M. K., Vriend, G. & Scholer, H. R. (1998). *Genes Dev.* **12**, 2073–2090.

Ferre-D'Amare, A. R. & Burley, S. K. (1997). *Methods Enzymol.* **276**, 157–166.

Garman, E. F. & Schneider, T. R. (1997). *J. Appl. Cryst.* **30**, 211–237.

Herr, W. & Cleary, M. A. (1995). *Genes Dev.* **9**, 1679–1693.

Hope, H. (1988). *Acta Cryst.* B**44**, 22–26.

Jacobson, E. M., Li, P., Rosenfeld, M. G. & Aggarwal, A. K. (1996). *Proteins*, **24**, 263–265.

Jordan, S. R., Whitcombe, T. V., Berg, J. M. & Pabo, C. O. (1985). *Science*, **230**, 1383–5.

Klemm, J. D. & Pabo, C. O. (1996). *Genes Dev.* **10**, 27–36.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W. & Pabo, C. O. (1994). *Cell*, **77**, 21–32.

Kottke, T. & Stalke, D. (1993). *J. Appl. Cryst.* **26**, 615–619.

Lusty, C. J. (1999). *J. Appl. Cryst.* **32**, 106–112.

Nagata, T., Gupta, V., Sorce, D., Kim, W. Y., Sali, A., Chait, B. T., Shigesada, K., Ito, Y. & Werner, M. H. (1999). *Nature Struct. Biol.* **6**, 615–619.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–325.

Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. (1994). *Anal. Biochem.* **216**, 413–417.

Phillipsen, A. (2001). *DINO: Visualizing Structural Biology*, http://www.biozentrum.unibas.ch/~xray/dino.

Rigoni, P., Xu, L., Harshman, K., Schaffner, W. & Arnosti, D. N. (1993). *Biochim. Biophys. Acta*, **1173**, 141–146.

Rosenberg, A. H., Goldman, E., Dunn, J. J., Studier, F. W. & Zubay, G. (1993). *J. Bacteriol.* **175**, 716–722.

Ryan, A. K., & Rosenfeld, M. G. (1997). *Genes Dev.* **11**, 1207–1225.

Schenk, P. M., Baumann, S., Mattes, R. & Steinbiss, H. H. (1995). *Biotechniques*, **19**, 196–200.

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

Snyder, G. H., Cennerazzo, M. J., Karalis, A. J. & Field, D. (1981). *Biochemistry*, **20**, 6509–6519.

Tomilin, A., Reményi, A., Lins, K., Bak, H., Leidel, S., Vriend, G., Wilmanns, M. & Schöler, H. R. (2000). *Cell*, **103**, 853–864.

Venter, J. C. *et al.* (2001). *Science*, **291**, 1304–1350.