Fine-tuning of intrinsic N-Oct-3 POU domain allostery by regulatory DNA targets

Robert Alazard¹, Lionel Mourey¹, Christine Ebel², Peter V. Konarev³, Maxim V. Petoukhov³, Dmitri I. Svergun³ and Monique Erard¹,*

¹Institut de Pharmacologie et de Biologie Structurale, 205 Route de Narbonne, 31077 Toulouse, ²Institut de Biologie Structurale, UMR 5075 CEA-CNRS-UJF, 41 rue Jules Horowitz, 38027 Grenoble, France and ³European Molecular Biology Laboratory, Hamburg Outstation, EMBL c/o DESY, D-22603 Hamburg, Germany and Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia

Received April 13, 2007; Revised May 18, 2007; Accepted May 21, 2007

ABSTRACT

The ‘POU’ (acronym of Pit-1, Oct-1, Unc-86) family of transcription factors share a common DNA-binding domain of approximately 160 residues, comprising so-called ‘POUs’ and ‘POUh’ sub-domains connected by a flexible linker. The importance of POU proteins as developmental regulators and tumor-promoting agents is due to linker flexibility, which allows them to adapt to a considerable variety of DNA targets. However, because of this flexibility, it has not been possible to determine the Oct-1/Pit-1 linker structure in crystallographic POU/DNA complexes. We have previously shown that the neuronal POU protein N-Oct-3 linker contains a structured region. Here, we have used a combination of hydrodynamic methods, DNA footprinting experiments, molecular modeling and small angle X-ray scattering to (i) structurally interpret the N-Oct-3-binding site within the HLA DRα gene promoter and deduce from this a novel POU domain allosteric conformation and (ii) analyze the molecular mechanisms involved in conformational transitions. We conclude that there might exist a continuum running from free to ‘pre-bound’ N-Oct-3 POU conformations and that regulatory DNA regions likely select pre-existing conformers, in addition to molding the appropriate DBD structure. Finally, we suggest that a specific pair of glycine residues in the linker might act as a major conformational switch.

INTRODUCTION

The high-throughput functional identification and structural characterization of transcriptional networks are major objectives of post-genomic research (1–4). Predictive methods have an important role to play in this endeavor since the large number of protein/DNA and protein/protein interactions involved in transcriptional regulation precludes their systematic study by X-ray crystallography or NMR. Since transcription factor families are generally specified by highly conserved consensus DNA-binding domains (DBD) as well as common strategies of interaction with target DNA (5) DBD homology modeling is a particularly relevant approach (see (6) and references herein). Equally, the prepositioning of a DBD within its DNA-binding site can often be inferred by homology, a step that most docking programs cannot yet address ab initio (7). However, despite these advantages, the prediction of DBD/DNA complex 3D structures is by no means straightforward, as exemplified by complexes involving the POU DBD.

The ‘POU’ (acronym of Pit, Oct, Unc) family of transcription factors is defined on the basis of a common DBD of approximately 160 residues, first identified in the mammalian proteins Pit-1 and Oct-1 and the nematode factor Unc-86 [for a review, see (8)]. The POU DBD comprises two distinct, highly conserved sub-domains, termed ‘POUs’ and ‘POUh’, which contain respectively four and three α-helices and are connected by a flexible linker, variable in sequence and length. The crystallographic structure of the complex between the POU domain of the ubiquitous protein Oct-1 and the octamer ATGCAAAT has revealed that POUs interacts with the tetramer ATGC in a similar fashion to the phage repressors, whereas the POUh interaction with the tetramer AAAT resembles that of a homeodomain (9).

If all the POU domains can bind to the prototypic octamer ATGCAAAT, they also recognize numerous other AT-rich sequences due to the flexibility of the linker.
joining the two sub-domains (10). Remarkably, crystallographic structures of various Pit-1 or Oct-1 POU/DNA complexes have shown that the cis elements of a DNA target recognized respectively by POUs and POUh neither have to be contiguous nor even to belong to the same DNA strand (11–13). Taken together, these structures have revealed two distinct patterns of POU homodimerization, based on different relative positionings of POUs and POUh, and depending on the type of DNA target. The ‘PORE’ (Palindromic Oct-1 Responsive Elements) DNA motifs induce a POU conformation similar to that found in the initial Oct-1 POU/octamer complex. By contrast, the ‘MORE’ (More palindromic Oct-1 Responsive Element) DNA motifs elicit a POU conformation analogous to that first discovered in Pit-1 POU/DNA complexes.

N-Oct-3, the human equivalent of the mouse Brn-2 protein, is widely expressed in the developing central nervous system, and necessary to maintain neural cell differentiation (14). It is also implicated in the development of the neural-crest-derived melanocytic lineage and its over-expression in melanocytes leads to differentiation (14). It is also implicated in the development of the neural-crest-derived melanocytic lineage and its over-expression in melanocytes leads to differentiation (14). It is also implicated in the development of the neural-crest-derived melanocytic lineage and its over-expression in melanocytes leads to differentiation (14).

The ‘PORE’ (Palindromic Oct-1 Responsive Elements) DNA motifs induce a POU conformation similar to that found in the initial Oct-1 POU/octamer complex. By contrast, the ‘MORE’ (More palindromic Oct-1 Responsive Element) DNA motifs elicit a POU conformation analogous to that first discovered in Pit-1 POU/DNA complexes.

In the current study, we have used a combination of hydrodynamic methods, DNA footprinting experiments, molecular modeling and small angle X-ray scattering (SAXS) to address the following questions: (i) How should the N-Oct-3-binding site within the HLA DRz promoter be read structurally and translated into a new POU domain allosteric conformation? (ii) How do transitions between free and bound conformations occur and what are the molecular mechanisms involved? Our results lead us to conclude that there might exist a continuous spectrum of free and ‘pre-bound’ N-Oct-3 POU conformations. In addition, a specific pair of glycine residues in the linker likely acts as a major conformational switch.

**MATERIALS AND METHODS**

### DNA targets and N-Oct-3 DBD preparation

Twenty-four base-pair oligonucleotides corresponding respectively to the (−127/−104) and (−57/−34) fragments of the rat CRH gene promoter (20) and the human HLA DRz gene promoter (21), and encompassing the N-Oct-3 POU homodimer-binding sites, were prepared and purified as previously described (22). The two sequences are as follows:

(CRH) 5’GCTCCTGCATAAATAATAGGGCC3’ – (DRz) 5’ATTGATTGCATTTATAATGCTCA3’

A 100 bp fragment encompassing the DRz promoter sequence was generated by PCR using the plasmid pSVODRzlacZ (kindly provided by Dr Goding) and two flanking primers. DNase I footprinting assays were performed as described (19).

The N-Oct-3 His-tag DBD was purified as before with the exception of the final gel filtration on a Superdex 75 HR 16/60 column instead of the heparin sepharose chromatography (22). Protein samples were concentrated and buffer exchanged with 25 mM Tris pH 7.5, 500 mM NaCl, 1% glycerol, 2 mM DTT, by ultrafiltration using Microcon centrifugal filter devices, then stored at −70°C and thawed prior to the experiments. The concentration was calculated from absorption measurements at 280 nm using an estimated molar extinction coefficient of 12900 M⁻¹ cm⁻¹. The dispersity of each protein preparation was assessed by dynamic light scattering (DLS) measurements using a DynaPro molecular sizing instrument. The N-Oct-3 DBD folding was checked by circular dichroism using a Jobin-Yvon Mark VI dichrograph.

**FPLC size-exclusion chromatography**

Analytical size-exclusion chromatography was performed at 5°C on a Superdex 75 16/60 column (Pharmacia) equilibrated with 50 mM Tris pH 7.5, 0.1 M NaCl, 2% glycerol, 2 mM DTT. The column was calibrated using the Pharmacia low molecular weight calibrating kit containing bovine serum albumin (M = 67 kDa, Rs = 35.5 Å), ovalbumin (M = 43 kDa, Rs = 30.5 Å), chymotrypsinogen (M = 25 kDa, Rs = 20.9 Å) and ribonuclease A (M = 13.7 kDa, Rs = 16.4 Å). Hydrodynamic or Stokes radii (Rs) were calculated from the plot of (−log Kav)¹/² versus Rs.

**Analytical ultracentrifugation**

Sedimentation velocity analysis was performed using a Beckman XL-A analytical ultracentrifuge and an AN-60 Ti rotor (Beckman Instruments). Experiments were carried out at 12°C in 50 mM Tris pH 7.5, 0.5 M NaCl, 2% glycerol, 0.3 mM TCPH at protein concentrations of 1 and 2 mg/ml. Samples of 400 μl were loaded into 12-mm path-length double-sector cells and centrifuged at 42000 r.p.m. Their absorbance was recorded at 280 nm. The solvent density, ρ, and viscosity, η, were measured at 20°C as 1.027 g/ml and η/ηH2O = 1.134 using a density-meter DMA 5000 and viscosity-meter AMVn (Anton PAAR). The values at 12°C were determined to be 1.028 g/ml and η = 1.398 cp. The partial specific volume of the protein, v, was estimated from the amino acid composition at 0.731 ml/g using the SEDNTERP program (V1.01; developed by Haynes, Laue, and Philo; available at http://www.bbri.org/RASMB/rasmb.html).

Data processing was carried out using the SEDFIT program (http://www.analyticalultracentrifugation.com/). Continuous distributions were obtained considering 200 particles of frictional ratio 1.5 with sedimentation
coefficients between 0.1 and 5.0 S, and using a regularization procedure (F ratio 0.7) (23). The non-interacting single-component model analysis was used to determine independently the sedimentation coefficient ($s$) and molecular mass ($M$) from the sedimentation velocity profiles. The two analyses take advantage of a systematic noise evaluation procedure (24,25). The corrected sedimentation coefficients, $s_{20\text{,}w}$, were derived from the experimental ones ($s$) using the following equation:

$$ s_{20\text{,}w} = s((1-\nu p)20\text{,}w)/(1-\nu p)(\eta/\eta 20\text{,}w) $$

The Svedberg equation was used to relate $s$, $M$ and the hydrodynamic radius $R_H$ as follows:

$$ s = M((1 - \nu p)/\left( N_A 6\pi\eta R_H \right) $$

**Molecular modeling**

Models were generated using the *Accelrys* modules InsightII, Biopolymer, Discover, Docking, Homology and Decipher (version 2005), run on a Silicon Graphics Fuel workstation, following the main outlines as previously described (19). Models of the 24 bp DNA fragments from the CRH and DR$\alpha$ gene promoters were built based on respective local homology with the NORE motif (19) and the MORE motif [PDB accession number: 1E3O (12)] after assignment of the POUs and POUh tetrameric binding sites. The four inter base-pair structural parameters (rise, twist, tilt and roll) were inferred from the homologous templates. The N- and C-terminal regions of the N-Oct-3 DBD were modeled in an extended conformation. The two-step docking was performed as before (19).

An automated conformational search procedure based on torsion driving was applied to the CRH-induced form of the N-Oct-3 DBD. The Gly 98 $\Phi$ and Gly 110 $\psi$ dihedral angles were selected as rotors, and systematically modified by 18$^\circ$ increments in the –180$^\circ$ to 180$^\circ$ range. The 441 resulting conformers were first filtered out using an energy threshold ($\leq$2.10$^4$ kcal/mol), and then divided into structural families. Each cluster was defined by conformations with similar relative orientations of the POUs and POUh sub-domains and overall backbone configurations superimposable within 4–5 Å.

**Scattering experiments and data analysis**

The synchrotron radiation X-ray scattering data were collected on the X33 camera (26,27) of the European Molecular Biology Laboratory (EMBL) at the storage ring DORIS III (Deutsches Elektronen Synchrotron) using a linear gas detector (28). The scattering patterns from the free N-Oct-3 DBD and from the 24-bp CRH and DR$\alpha$ promoter fragments, either free or in complex with the DBD, were measured at several solute concentrations between 2.5 and 8 mg/ml and in 50 mM Tris pH 7.5, 0.4 M NaCl, 2% glycerol, 2 mM DTT. The data were collected at 12$^\circ$C at a sample-detector distance of 2.3 m covering the momentum transfer range 0.15 < $s$ < 3.5 nm$^{-1}$ ($s = 4\pi \sin \theta /\lambda$), where $2\theta$ is the scattering angle and $\lambda = 0.15$ nm the X-ray wavelength. The data collected in 15 successive 1-minute frames to check the radiation damage were normalized and processed using the program PRIMUS (29). The difference curves after buffer subtraction were extrapolated to infinite dilution following standard procedures (30).

The maximum particle dimensions $D_{\text{max}}$ were estimated using the orthogonal expansion program ORTOGNOM (31). The forward scattering values $I(0)$ and the radii of gyration $R_g$ were evaluated using the Guinier approximation (32) and by using the indirect transform package GNOM (33), which also provides the distance distribution functions $p(r)$ of the particles. The molecular masses ($M$) of the solutes were evaluated by comparison of the forward scattering with that from a reference solution of bovine serum albumin ($M = 66$ kDa).

The scattering patterns from the predicted models of the free N-Oct-3 DBD, the CRH and DR$\alpha$ DNA fragments, and their respective complexes, were computed using the program CRYSOL (34). Given the atomic coordinates, the program fits the experimental scattering curve by adjusting the excluded volume of the particle and the contrast of the hydration layer surrounding the particle in solution to minimize the discrepancy estimated as follows:

$$ \chi^2 = \frac{1}{N-1} \sum_j \left[ \frac{I_{\text{exp}}(s_j) - I_{\text{calc}}(s_j)}{\sigma(s_j)} \right]^2, $$

where $N$ is the number of experimental points, $c$ is a scaling factor, $I_{\text{exp}}(s_j)$, $I_{\text{calc}}(s_j)$ and $\sigma(s_j)$ are the experimental and calculated intensity, and the experimental error at the momentum transfer $s_j$, respectively.

**RESULTS AND DISCUSSION**

Hydrodynamic properties show that free N-Oct-3 POU is monomeric

The N-Oct-3 DNA-binding domain (DBD) purifies as a single species of 20 kDa molecular mass as judged by SDS-PAGE (Figure 1A). In order to investigate the oligomerization state and hydrodynamic radius of this POU domain, we first carried out dynamic light scattering (DLS) and analytical gel filtration experiments. DLS measurements recorded at 20°C and at a maximal concentration of 4 mg/ml indicated a low polydispersity and a narrow particle size distribution diagram corresponding to a hydrodynamic radius of 29.3 Å (Figure 1B). The purified N-Oct-3 POU domain eluted from a FPLC-size exclusion chromatography column between the 43 and 25 kDa calibration proteins and the elution volume served to calculate its Stokes radius (Figure 1C). The resulting Rs value of 27.6 Å was very similar to that calculated by DLS, but significantly higher than those of globular proteins of an equivalent molecular weight. This indicates the presence of either a dimer or an elongated monomer in solution.

The N-Oct-3 DBD was then submitted to sedimentation velocity analysis, and the data were processed as described in the Materials and Methods section. A selection of
sedimentation profiles performed in the same conditions, along with their best-fits using a single component, are shown in Figure 2A, the corresponding residuals being displayed in Figure 2B. Identical sedimentation coefficients were obtained (1.84 S) at the two concentrations used (1 and 2 mg/ml), and the deduced molecular mass (21 kDa) indicates, when compared with the theoretical mass (19.9 kDa), that the N-Oct3 DBD is a monomer. In addition, the analysis of the sedimentation profiles in terms of a continuous distribution of elongated particles showed narrow single peaks at both concentrations (Figure 2C). This clearly demonstrates the homogeneity of the solution and the lack of any association–dissociation processes, thereby confirming the monomeric status of the free N-Oct-3 DBD. Thus we can conclude that the N-Oct-3 POU, either induced by the NORE motif of the CRH gene promoter (19) or by an element of the HLA DRα gene promoter. In the latter case, it was first necessary to characterize the interaction between the N-Oct-3 POU and its DNA target.

**Structural reading of the N-Oct-3-binding site within the HLA DRα gene promoter and POU domain allosteric**

We have previously shown (19) that the N-Oct-3 POU domain can adopt three different conformations and corresponding homodimerization patterns in response to the particular distribution of potential POUs and POUh tetrameric binding sites which characterize the respective PORE, MORE and NORE motifs evoked earlier. In the same report, we defined a structural framework suitable for the analysis of any interaction between the N-Oct-3 POU domain and a DNA target. Most importantly, the POUs and POUh tetrameric binding sites for each monomer are non-contiguous and on opposite strands in the MORE mode, whereas they are contiguous and on the same strand in the PORE mode. This results in a different relative positioning of the POUs and POUh sub-domains within each monomer between the two modes. Finally, the NORE motif designates the 14-bp sequence element TNRRTAAATAATRN (N: any nucleotide; R: purine residues) which is common to a set of neuronal promoters, including the CRH gene promoter, and which is capable

---

**Figure 1.** Characterization of the N-Oct-3 POU domain. (A) Detection of a single band with the expected N-Oct-3 DBD molecular mass by Coomassie-blue staining in 13% SDS-PAGE (see the molecular mass markers on the right). (B) Dynamic light scattering of the N-Oct-3 DBD (see text). (C) Calibration curve obtained by FPLC size-exclusion chromatography of globular proteins of known Stokes radii ‘Rs’ (see the Materials and Methods section). The arrow indicates the elution position of the N-Oct-3 POU domain.
of eliciting a novel homodimerization mode exclusive to the N-Oct-3 DBD. Both the NORE and PORE motifs elicit a ‘POUh-dominant’ mode of N-Oct-3 DBD homodimerization with a strong anchoring into the DNA minor groove. However, in the case of the NORE mode, the two POUh-binding sites are overlapping, which explains the non-cooperative character of the homodimerization.

DNAse I footprinting is a particularly valuable tool to determine which homodimerization mode is elicited by a given DNA regulatory element. Bearing in mind the strong correlation between N-Oct-3 over-expression in melanomas and the up-regulation of HLA-DR\(\alpha\) gene expression (15,18), we used this approach, coupled to molecular modeling, to analyze N-Oct-3 binding to the HLA-DR\(\alpha\) gene promoter. Electrophoretic mobility shift assays (EMSA) showed that the N-Oct-3 POU domain binds as a non-cooperative homodimer to the DR\(\alpha\) DNA, a 24-bp DNA fragment of the HLA-DR\(\alpha\) gene promoter (Figure 3), with an effective dissociation constant \(K_d\) of \(5 \times 10^{-10}\) M for the first monomer (see Figure 3A legend) and an apparent dissociation constant \(K_d\) \(\leq 2.6 \times 10^{-8}\) M for the second monomer [see Figure 3B legend; (35)]. DNAse I footprinting of the first N-Oct-3 DBD binding to a promoter fragment encompassing this high-affinity binding site reveals a total protection of both DNA strands (lanes 1 in Figure 4A and B). We therefore deduce that the relative positioning of the POUs and POUh sub-domains within this first bound monomer must be elicited by a MORE-type motif, the only one with POUs and POUh-binding sites on both strands of the DNA.

A MORE motif is characterized by two strong POUs anchoring sites on opposite DNA strands and on either side of the pseudo-dyad axis. The sequence of these binding sites is most often ATG(/A)C, but an ATNN motif is sufficient to establish the highly specific set of interactions with the conserved Gln and Thr residues of the POUs recognition helix. Based on the DNAse I footprint, the A12T13T14T15 tetramer on the upper strand and the overlapping A12BT13BG14BC15B tetramer on the lower strand of the HLA-DR\(\alpha\) gene promoter possess the appropriate structural requirements for the two POUs-binding sites in the MORE configuration (Figure 4C). In line with this, the non-cooperativity of the homodimerization observed by EMSA (Figure 3B) is consistent with the overlap of the two POUs-binding sites. Furthermore, the mutagenesis of the A12T13T14 triplet is sufficient to abolish the binding of both monomers (data not shown).

Following the assignment of the two POUs-binding sites as A12T13T14T15 on the upper strand and A12BT13BG14BC15B on the lower strand, the two corresponding POUh-binding sites can now be predicted as G14BC15BA16BA17B and T14T15T16A17 respectively, based on the known MORE motif organization.
In this mode, each POUh-binding site overlaps the POU-binding site of the other monomer on the same strand (see Figure 4C and its legend). The extent of the DNAse I footprint on the lower strand as a consequence of the first monomer binding designates G14BC15BA16BA17B as the first POUh-binding site (lane 1 in Figure 4B), and hence the A12T13T14T15 as the first POU-binding site. This implies that the A12BT13BG14BC15B tetramer on the lower strand is the second POUs-binding site and the T14T15T16A17 tetramer on the upper strand is the second POUh-binding site.

It is important to underline that, as for the so-called canonical sequence of the human immunoglobulin heavy chain gene promoters IgG V_H (19,36), the prototypic octamer sequence ATGCAAAT on the lower strand is not ‘read’ as a single continuous POU-binding site but, instead, as the second POUs-binding site (ATGC) overlapping the first POUh-binding site (GCAA). As a consequence, the terminal AT is still cleaved by DNAse I since it does not take an active part in the interaction (see green-colored marking in Figure 4B and C).
Now that the POUs and POUh-binding sites have been assigned, the bound structure of the HLA DR\(\alpha\) promoter DNA fragment can be built and docked with the corresponding sub-domains. The resulting model is displayed in Figure 5A and B. It is known that the generic MORE mode can accommodate variable spacings between the two POUs insertion sites. For example the ‘MORE+2’ mode, corresponds to a 2 bp spacing (37). Following this nomenclature, the DR\(\alpha\)/DBD complex represents a new MORE subtype, which can be designated by ‘MORE-2’. In this mode, the two POUs DNA recognition helices are inserted into overlapping sites in the major groove (see the red-colored star in Figure 5B).

If regulatory conformations of the N-Oct-3 POU domain require molding by the respective DNA structure, we need to ask what is the molecular mechanism responsible for this remarkable adaptation to the promoter structure.

A pair of Gly residues in the N-Oct-3 POU linker as potential actors in the conformational switch: a combined molecular mechanics and SAXS approach

A comparative analysis of the N-Oct-3 POU conformation induced by the DR\(\alpha\) DNA (Figure 5C) with that induced by the CRH DNA (Figure 5D) taking the position of the POUs as a fixed reference, reveals that
the two POUh sub-domain orientations can be superimposed by an ~180° rotation around the linker taken as a virtual axis.

Before dealing with the structural determinants of N-Oct-3 linker flexibility, we first need to recall its distinctive features. Using circular dichroism, we previously observed an increase in the a-helical content of the N-Oct-3 DBD when binding to its DNA targets, in contrast to the Oct-1 DBD (38). Since the only significant difference between these two highly conserved DBDs is their respective linker sequences, we engineered chimeric proteins where the N-Oct-3 and the Oct-1 linkers were interchanged. This showed that the replacement of the N-Oct-3 DBD linker by that of Oct-1 abolished the increase in a-helical structure, whereas the replacement of the Oct-1 linker by that of N-Oct-3 resulted in the typical increase in the a-helical content following protein/DNA complex formation. Since a number of reliable secondary structure prediction methods indicated that the heptapeptide motif IDKIAAQ specific to the N-Oct-3 linker could adopt an a-helical structure, we built another set of chimeric proteins where this heptapeptide was removed from the N-Oct-3 linker and embedded within the Oct-1 linker. As the results were similar to those for the entire linker interchange experiments, we concluded that the ability of the N-Oct-3 linker to adopt an a-helical structure when binding to a DNA target could be ascribed to the IDKIAAQ motif (see its location in the DBD sequence in Figure 6A). We now show that the potential secondary structure of this heptapeptide motif can also be stabilized independently of DNA binding, when free DBD concentrations are greater than 0.7 mg/ml (see Figure S1 and its legend), which are the conditions of the hydrodynamic and SAXS experiments reported here. Note that the link between protein folding and molecular concentration has been revealed in a number of recent works (see for example (39,40)). Thus the N-Oct-3 linker has the characteristics of a ‘helical linker’ as defined by George and Heringa based on an extensive compilation of inter-domain linkers (41). Interestingly, the helical heptapeptide IDKIAAQ is preceded by the 4-residue motif SPTS (Figure 6A), shown to form a b-turn in a number of proteins and polypeptides, the structures of which were solved by crystallography or NMR (42–44).

A crucial feature of hinge residues is that they have very few packing constraints in their main chain atoms (45,46). As such, the Gly residues are well suited to promote hinge motion (47,48). The two Gly residues present in the N-Oct-3 DBD linker (Figure 6A) could therefore act as major molecular pivots in the conformational transitions. To examine this further, we performed automated conformational searches by systematically sampling the φ and ψ dihedral angles of Gly 98 and Gly 110, using the CRH-bound conformation as a starting structure. We found the combination of Gly 98φ and Gly 110 ψ dihedral angles to be the most efficient to explore the N-Oct-3 DBD conformational space (see the Materials and Methods section and Figure S3A and B). After filtering using an energy threshold, the resulting conformers could be clustered within a discrete number of conformational families, based on overall R.M.S. values of 4–5 Å and corresponding to different relative orientations of the POUs and POUh sub-domains such as those displayed in Figure 6B–D. In order to identify potential free forms amongst these structures, we first compared their calculated radius of gyration (Rg) to the free N-Oct-3 DBD hydrodynamic radius. To select the most likely candidates, we then combined molecular mechanics with SAXS methodology following the main outlines of a recent study (49).

Processed X-ray scattering patterns corresponding to the free N-Oct-3 DBD are presented in Figure 7A and B (data groups 1), alongside those from the free DNA fragments (data groups 2) and from the equimolecular N-Oct-3 DBD/DNA complexes (data groups 3). The structural parameters computed from the experimental data, including the radius of gyration (Rg) and maximum particle dimension (Dmax), are displayed in Table 1. The estimated effective mass (Meff) of the free N-Oct-3 DBD agrees within experimental error with the value expected from the sequence (Mseq), confirming that the protein is monomeric in solution. The distance distribution functions computed from the experimental data (Figure 8) emphasize the elongated shape of the free form(s), and the similarities between the gyration radii of the free N-Oct-3 DBD and of its complexes with each promoter DNA fragment. Note the good agreement between the free N-Oct-3 DBD gyration and hydrodynamic radii.

In all cases, the theoretical scattering patterns of the predicted structures were computed using the program CRYSOL and then compared to the experimental data. The accuracy of the fit was assessed by the discrepancy value χ as explained in the Material and Methods section, where typical values between 0.8 and 1.1 indicate good agreement. Thus, the computed scattering curves corresponding to the models of both the CRH DNA fragment and the N-Oct-3 DBD/CRH complex agree well with the respective experimental curves, with discrepancy values of 1.05 and 1.09, respectively (data groups 2 and 3 in Figure 7A and Table 1; Figure S2A). The same observations can be made for the models of the DRz DNA fragment and the N-Oct-3 DBD/DRz complex (data groups 2 and 3 in Figure 7B and respective discrepancy values of 0.82 and 1.09 in Table 1; Figure S2B). Fitting the computed scattering curves of the N-Oct-3 DBD in the predicted CRH- or DRz-bound conformations with the experimental data for the free N-Oct-3 DBD yields slightly higher discrepancy values (see respective χ values of 1.18 and 1.23 in Table 1 and data groups 1 in Figure 7A and B). In order to accurately interpret this in terms of similarities versus differences between free and bound conformations, we must first build a referential of free-form models. For this, we systematically computed the theoretical scattering curves of the molecular mechanics-derived structures and fitted them to the free DBD experimental data.

According to their χ values in the 1.06–1.09 range, a number of conformers appear as good candidates to represent free N-Oct-3 DBD conformations. These can be divided into two distinct clusters which are themselves part of larger conformational families, ‘F1’ and ‘F1’,
defined by respective overall R.M.S. values of 4.9 Å (Figure 6B) and 4.4 Å (Figure 6C). Importantly, the χ value dispersion observed in both cases, 1.06–1.19 and 1.06–1.27 respectively, is compatible with the conservation of a given overall POU domain conformation within each family.

A more detailed analysis indicates that each conformational family contains structural sub-classes characterized by a particular distance between the POUs and POUh recognition helices (‘RHdist’) within the 18–35 Å range. Interestingly, the conformers with the lowest RHdist (Figure S3B) tend to be less energetically stable.

Figure 6. Conformational search by torsion driving. (A) Location of the linker (brown-coded) within the sequence of the N-Oct-3 DBD: the Gly 98 and Gly 110 residues (highlighted) flank the SPTSIDKIAAQ undecapeptide (underlined). Other critical features are the Gln 63 and Asn 162 residues (red-coded) in the respective POUs and POUh DNA recognition helices (purple-coded). Display code for the remaining elements as follows: gray for the POUh N-terminal arm, blue for helices 1, 2, 4, 5, 6, green for the regions between secondary structure elements, black for exogenous regions resulting from the DBD cloning. (B–D) Clustering of molecular mechanics-derived structures in families of potential free forms (B, C) and extended conformers (D). The conformers Cα traces are structurally aligned within a 4–5 Å R.M.S. range in each cluster. (E–G) The conformers Cf 183 (E), Cf 194 (F) and Cf 221 (G) are the best representatives of each family, respectively FI (B), FII (C) and NF (D). In all cases, Gly 98 and Gly 110 are coded in brown, Gln 63 and Asn 162 in red, the POUs and POUh recognition helices in purple. RHdist is monitored in Å.

MGSSHAAAAHSSGLVPRGSHMTPTSDDLEQFAKQFKQRRIK
LGFQTQADVGLALGTYVNGVSQTTICREALQLSFKNMCKKLK
PLLKNKWEAADSSGSPTSIDKIAAQGRKRRKRTSIEVSVKG
ALESHEFLKCPKPSAQEITSLADLQLEKEVVRVWFCNRRQKE
KRMTPPGGTLP
(Figure S3A), but are closer to the respective CRH- and DRz-bound conformations for which RHdist is comprised within the 15–20 Å range (Figure 5C and D). Taken together, these results imply that the two populations of putative free forms, F1 and FII, most likely coexist, and also that there could be a structural continuum running from free to less stable ‘pre-bound’ conformations. In line with this, the fitted scattering curve of the CRH-bound modeled structure is very close to that of ‘Cf 183’ (see the respective red- and turquoise-colored curves of data group 1 in Figure 7A, and the corresponding χ values of 1.18 and 1.09 in Table 1), Cf 183 being the best F1 representative (Figure 6E). Similarly, the fitted scattering curve of the DRz-bound modeled structure is very close to that of ‘Cf 194’ (see the respective blue- and magenta-colored curves of data group 1 in Figure 7B, and the corresponding χ values of 1.23 and 1.08 in Table 1), Cf 194 being the best FII representative (Figure 6F). By contrast, the fitted scattering curve of ‘Cf 221’ significantly deviates from the free N-Oct-3 DBD experimental data with a χ value of 1.90 (see the dashed green-colored curve in data group 1 in Figure 7A and B, and Table 1). Indeed, this conformer (Figure 6G), with its higher Rg (32 Å) and RHdist (50 Å) values, cannot represent the free form and belongs to a large conformational family of extended structures, characterized by RHdist values within the 40–50 Å range (Figure 6D).

Model fitting against experimental SAXS data is a useful means to interpret scattering information in terms of higher-resolution structures (50). Fitting of multiple models generated by molecular mechanics or dynamics has also been applied to analyze conformer ensembles in solution, especially in relation to protein unfolding (51). Along these lines, a recent report [see (52) and references therein] has explored how multiple well-defined protein conformations in a sample influence the scattering data. Test cases were established, based on simulation of SAXS data from reconstituted ensembles of protein structures, such as ensembles comprising various weighted proportions of the extended and collapsed states of calmodulin, a protein comprising two globular domains connected by a flexible helical linker. One of the main conclusions of this study is that the ability of ab initio modeling to differentiate static structures from dynamic structures depends strongly on the extent of the variability of the ensemble. Hence, an ab initio low-resolution model of the free N-Oct-3 DBD can be expected to reflect distinct properties from respective members of the FI and FII conformational families, but probably not from members of the same family. Indeed, a molecular envelope of the N-Oct-3 DBD generated using the GASBOR program (53) can accommodate the DRz and the CRH-bound conformations at different sites (see Figure S4 and its legend). As these conformations bear similarities with the respective overall structures of the FI or FII families’ members, this lends support to the likely coexistence of these two conformational families, inasmuch as they are energetically equiprobable (see Figure S3 and its legend).
CONCLUDING REMARKS

Initially structural studies performed on the POU-type DNA-binding domain showed that individual POUs and POUh sub-domains could be considered as rigid bodies when interacting with DNA (9,54,55). The adaptability of several POU proteins to a variety of DNA targets was then ascribed to the flexibility of the linker joining the POU sub-domains (56,57). However, despite the critical importance of the linker with regards to the molding of specific regulatory POU conformations to the target DNA, no detailed molecular mechanism for this flexibility has so far been proposed. One of the main reasons for this of course is that neither Oct-1 nor Pit-1 POU linker structures can be resolved in the available crystallographic data derived from POU/DNA complexes.

The N-Oct-3 DBD linker has dual structural properties. On the one hand, it contains a helical peptide motif, in common with approximately half of the known inter-domain linkers (41), which might constrain the relative orientation of the two POU sub-domains. On the other hand, this linker also functions as a hinge region, as best exemplified in the transition between the CRH- and the DRα-bound conformations. A number of studies dealing with hinge motion (45–48) designate the pair of Gly residues present in the linker as potential key-players in the N-Oct-3 DBD conformational transitions. Based on these working hypotheses, we have combined various hydrodynamic and SAXS data with the results of a conformational search through torsion driving. We have shown that the linker flexibility resulting from rotations around this pair of Gly residues is sufficient to generate the transitions between the free and bound conformations, whilst at the same time respecting the local structuring of the linker. We have identified two families of putative free N-Oct-3 POU conformations, which can be interconverted by rotation around a virtual Gly–Gly hinge axis. As specified earlier in the text, the distances between the DNA recognition helices (‘RH dist’) in these conformers lie within the 18–35 Å range, which favors the concerted DNA-binding activity of the two POU sub-domains.

There might exist an equilibrium between these two families of putative free conformers and, for each family, between best free form representatives and less stable ‘pre-bound’ conformers. We propose that NORE- or MORE-2-type DNA motifs select conformers closer to the final CRH- or DRα-bound conformations, respectively. Note that the importance of the Gly residues does not exclude the contribution of other residues to the overall flexibility of the linker, especially in the final adjustments required upon DNA binding.

In conclusion, our results indicate that regulatory DNA regions most likely select pre-existing N-Oct-3 DBD conformations, in addition to molding the appropriate DBD structure. More generally, our study emphasizes the necessity not only to employ a structural reading of nucleic regulatory sequences but also to integrate information about protein flexibility when predicting functional

<table>
<thead>
<tr>
<th>Sample</th>
<th>R_g (nm)</th>
<th>D_max (nm)</th>
<th>M_eff (kDa)</th>
<th>M_seq (kDa)</th>
<th>χ</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Oct-3 DBD</td>
<td>2.93 ± 0.05</td>
<td>10.0 ± 0.5</td>
<td>17 ± 3</td>
<td>20.0</td>
<td>1.18 (CRH-bound)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.23 (DRα-bound)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.09 (Cf 183)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.08 (Cf 194)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.90 (Cf 221)</td>
</tr>
<tr>
<td>CRH DNA</td>
<td>2.37 ± 0.04</td>
<td>8.5 ± 0.5</td>
<td>14 ± 2</td>
<td>15.0</td>
<td>1.05</td>
</tr>
<tr>
<td>DRα DNA</td>
<td>2.89 ± 0.03</td>
<td>11.0 ± 0.5</td>
<td>36 ± 4</td>
<td>35.0</td>
<td>1.09</td>
</tr>
<tr>
<td>N-Oct-3/CRH</td>
<td>2.85 ± 0.03</td>
<td>11.0 ± 0.5</td>
<td>34 ± 4</td>
<td>35.0</td>
<td>1.09</td>
</tr>
<tr>
<td>N-Oct-3/DRα</td>
<td>2.85 ± 0.03</td>
<td>11.0 ± 0.5</td>
<td>34 ± 4</td>
<td>35.0</td>
<td>1.09</td>
</tr>
</tbody>
</table>

R_g, D_max and M_eff designate, respectively, the radius of gyration, maximum size and effective molecular mass, calculated from the scattering data. For DNA-containing samples, the fact that the DNA contrast is higher than that of the protein was taken into account when estimating the M_eff value. M_seq is the molecular mass of the solutes predicted from the appropriate sequence. χ denotes the discrepancy between the experimental data and the scattering curves computed from the models. In the case of the N-Oct-3 DBD, the χ values have been calculated for the bound conformations, induced by the CRH or DRα DNA, and for conformations derived from molecular mechanics.

Figure 8. Distance distribution functions of the free N-Oct-3 DBD (green), the free CRH DNA (magenta), the N-Oct-3/CRH (red) and the N-Oct-3/DRα (blue) complexes.

Table 1. Summary of the structural parameters computed from the scattering data

![Image of distance distribution functions](image-url)
structure. Indeed a number of recent studies address the critical issue of the indirect readout of promoter DNA sequences (for example see (58–61), whilst new concepts and methods are emerging to explore protein flexibility and allostery (62,63). Along these lines, combining an ensemble optimization method with SAXS is a highly promising approach as perfectly illustrated in our recently published study (64).

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank David Barker for critical reading of the manuscript. R.A. and L.M. are grateful for financial support from the European Community to access the X33 EMBL beamline at DESY (Contract N° RI13/CT/2004/5060008 of the FP6 Program ‘Research Infrastructure Action’). This work was supported by research grants from the Région Midi-Pyrénées (AO N° 03001137) and from the ‘Action Concertée Incitative Biologie Cellulaire, Moléculaire et Structurale’ (AO N° 050031). Funding to pay the Open Access publication charges for this article was provided by Centre National de la Recherche Scientifique.

Conflict of interest statement. None declared.

REFERENCES


