

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

Robert Alazard<sup>1</sup>, Lionel Mourey<sup>1</sup>, Christine Ebel<sup>2</sup>, Peter V. Konarev<sup>3</sup>, Maxim V. Petoukhov<sup>3</sup>, Dmitri I. Svergun<sup>3</sup> and Monique Erard<sup>1,\*</sup>

<sup>1</sup>Institut de Pharmacologie et de Biologie Structurale, 205 Route de Narbonne, 31077 Toulouse, <sup>2</sup>Institut de Biologie Structurale, UMR 5075 CEA-CNRS-UJF, 41 rue Jules Horowitz, 38027 Grenoble, France and <sup>3</sup>European 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 DR $\alpha$  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  $\alpha$ -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 POUUs interact 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

\*To whom correspondence should be addressed. Tel: +33 (0) 562175496; Fax: +33 (0) 562175994; Email: Monique.Erard@ipbs.fr

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 POU 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 POU 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 tumorigenesis *via* the dysregulation of a number of genes (15–18). The fact that N-Oct-3 can interact with such a variety of targets is due to the structural plasticity of its POU domain. In a previous report (19), we have shown that the N-Oct-3 DBD, in addition to forming the classical homodimers in association with PORE and MORE sequences, can also adopt a novel mode of homodimerization when bound to a set of neuronal promoters, including the CRH (corticotropin-releasing hormone) gene promoter. We have demonstrated that this pattern is induced by a structural motif that we have termed ‘NORE’ (N-Oct-3 Responsive Element).

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 DR $\alpha$  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 DR $\alpha$  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′GCTCCTGCATAAATAATAGGGCCC3′ –  
(DR $\alpha$ ) 5′AATTGATTTGCATTTTAATGGTCA3′

A 100 bp fragment encompassing the DR $\alpha$  promoter sequence was generated by PCR using the plasmid pSVODR $\alpha$ lacZ (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 12 900 M<sup>–1</sup>.cm<sup>–1</sup>. 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, R<sub>s</sub> = 35.5 Å), ovalbumin (M = 43 kDa, R<sub>s</sub> = 30.5 Å), chymotrypsinogen (M = 25 kDa, R<sub>s</sub> = 20.9 Å) and ribonuclease A (M = 13.7 kDa, R<sub>s</sub> = 16.4 Å). Hydrodynamic or Stokes radii (R<sub>s</sub>) were calculated from the plot of (–log Kav)<sup>1/2</sup> versus R<sub>s</sub>.

### Analytical ultracentrifugation

Sedimentation velocity analysis was performed using a Beckman XL-I 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  $\mu$ l were loaded into 12-mm path-length double-sector cells and centrifuged at 42 000 r.p.m. Their absorbance was recorded at 280 nm. The solvent density,  $\rho$ , and viscosity,  $\eta$ , were measured at 20°C as 1.027 g/ml and  $\eta/\eta_{\text{H}_2\text{O}} = 1.134$  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  $\eta = 1.398$  cp. The partial specific volume of the protein,  $\bar{v}$ , 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,  $s_{20,w}$ , were derived from the experimental ones ( $s$ ) using the following equation:

$$s_{20,w} = s[(1 - \bar{v}\rho_{20,w})/(1 - \bar{v}\rho)](\eta/\eta_{20,w})$$

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

$$s = M(1 - \bar{v}\rho)/(N_A 6\pi\eta R_H)$$

### 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 POU<sub>s</sub> and POU<sub>h</sub> 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° increments in the -180° to 180° range. The 441 resulting conformers were first filtered out using an energy threshold (<2.10<sup>4</sup> kcal/mol), and then divided into structural families. Each cluster was defined by conformations with similar relative orientations of the POU<sub>s</sub> and POU<sub>h</sub> 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°C at a sample-detector distance of 2.3 m covering the momentum transfer range  $0.15 < s < 3.5 \text{ nm}^{-1}$  ( $s = 4\pi\sin\theta/\lambda$ , where  $2\theta$  is the

scattering angle and  $\lambda = 0.15 \text{ nm}$  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 ( $M = 66 \text{ kDa}$ ).

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 CRY SOL (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:

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

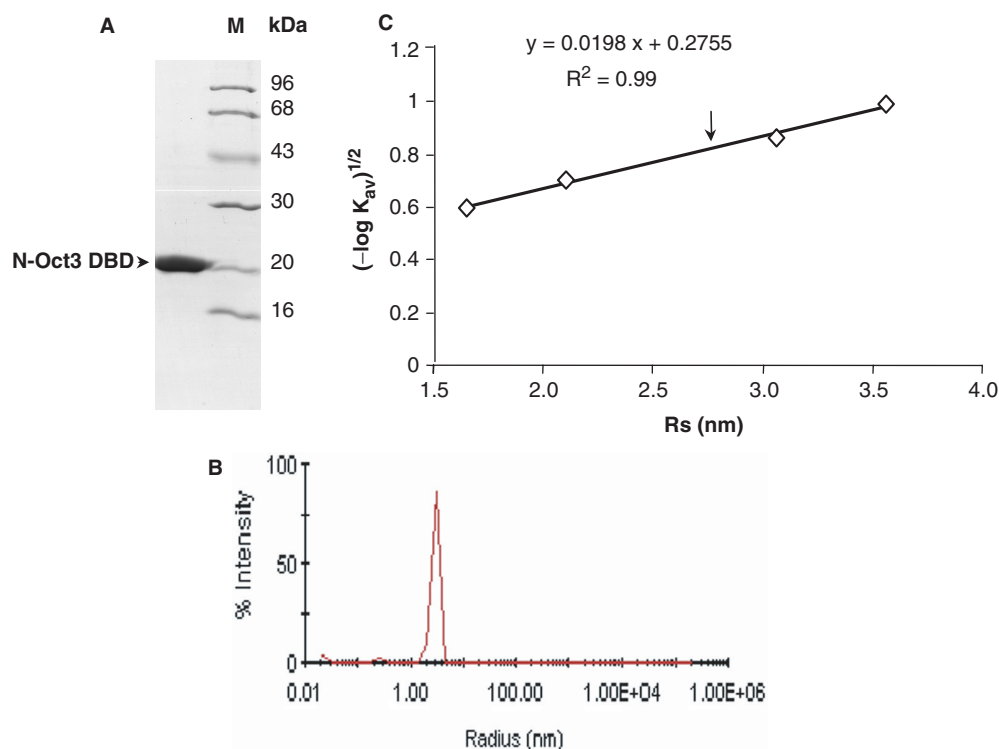
where  $N$  is the number of experimental points,  $c$  is a scaling factor,  $I_{\text{exp}}(s_j)$ ,  $I_{\text{calc}}(s_j)$  and  $\sigma(s_j)$  are the experimental and calculated intensity, and the experimental error at the momentum transfer  $s_j$ , respectively.

## 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  $R_s$  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



**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.

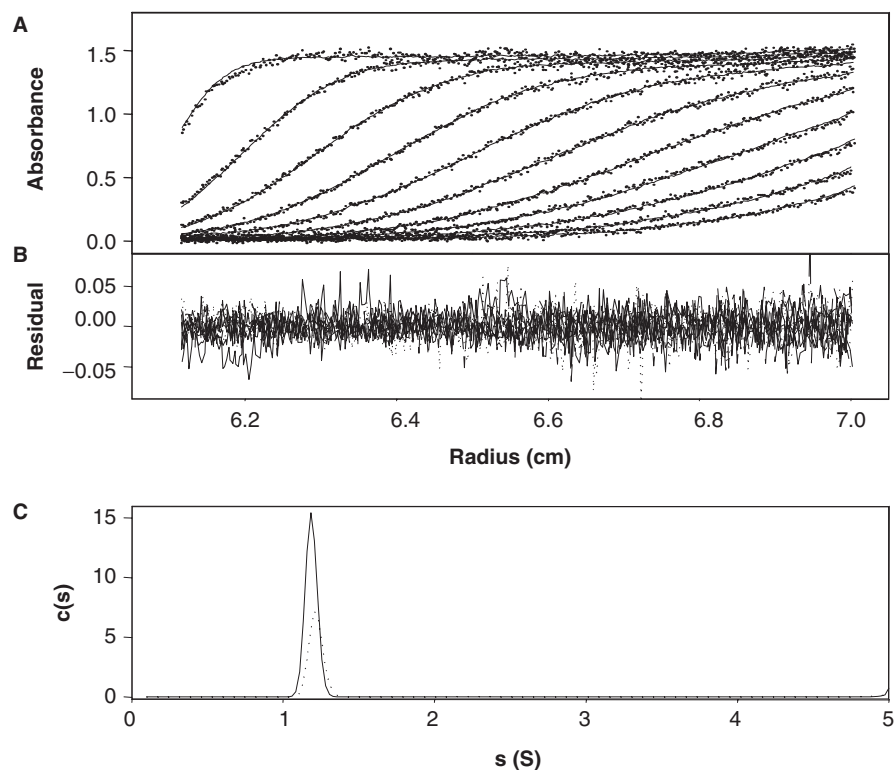
sedimentation profiles performed in the same conditions, along with their best-fits using a single component, are shown in Figure 2A, the corresponding residuals being displayed in Figure 2B. Identical sedimentation coefficients were obtained (1.84S) at the two concentrations used (1 and 2 mg/ml), and 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 homodimers which bind to a variety of DNA targets (19) do not exist prior to complex formation, but are a consequence of specific interactions with target DNAs.

The question then arises as to whether the elongated shape of the free N-Oct-3 DBD indicated by the hydrodynamic data reflects a single conformation or represents the average of a collection of conformers. In addition, we would like to determine the molecular mechanisms responsible for the transitions between the free and DNA-bound conformations. To attempt to answer these questions, we have performed a comparative analysis of two regulatory conformations of N-Oct-3

POU, either induced by the NORE motif of the CRH gene promoter (19) or by an element of the HLA DR $\alpha$  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 DR $\alpha$ gene promoter and POU domain allostery: a combined footprinting and molecular modeling approach

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 POU 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 POU 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 POU and POUh sub-domains within each monomer between the two modes. Finally, the NORE motif designates the 14-bp sequence element TNNRTAAATAATRN (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 2.** Sedimentation velocity analysis of the N-Oct-3 DBD. (A) Sedimentation velocity absorbance profiles were obtained at 12°C at a rotor speed of 42000 r.p.m. and scans were recorded at 280 nm for 21 h. The data analysis was performed using 22 regularly spaced profiles. The best-fit profiles corresponding to a single-component model are superimposed on the experimental data. For clarity, only one profile out of two is shown. (B) Corresponding residuals at a 2 mg/ml DBD concentration. (C) Continuous distributions of sedimentation coefficients obtained by considering elongated proteins of frictional ratio 1.5; protein concentration 1 mg/ml (dotted line) and 2 mg/ml (continuous line).

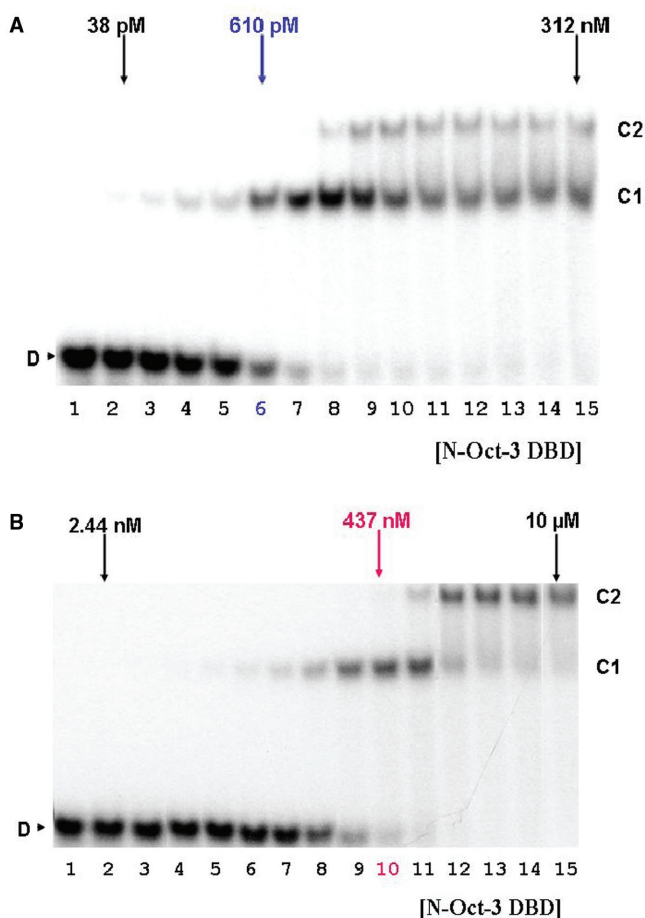
of eliciting a novel homodimerization mode exclusive to the N-Oct-3 DBD. Both the NORE and PORE motifs elicit a 'POU<sub>h</sub>-dominant' mode of N-Oct-3 DBD homodimerization with a strong anchoring into the DNA minor groove. However, in the case of the NORE mode, the two POU<sub>h</sub>-binding sites 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  $K_{d1}$  of  $5 \times 10^{-10}$  M for the first monomer (see Figure 3A legend) and an apparent dissociation constant  $K_{d2} \leq 2.6 \times 10^{-8}$  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 POU<sub>s</sub> and POU<sub>h</sub> sub-

domains within this first bound monomer must be elicited by a MORE-type motif, the only one with POU<sub>s</sub> and POU<sub>h</sub>-binding sites on both strands of the DNA.

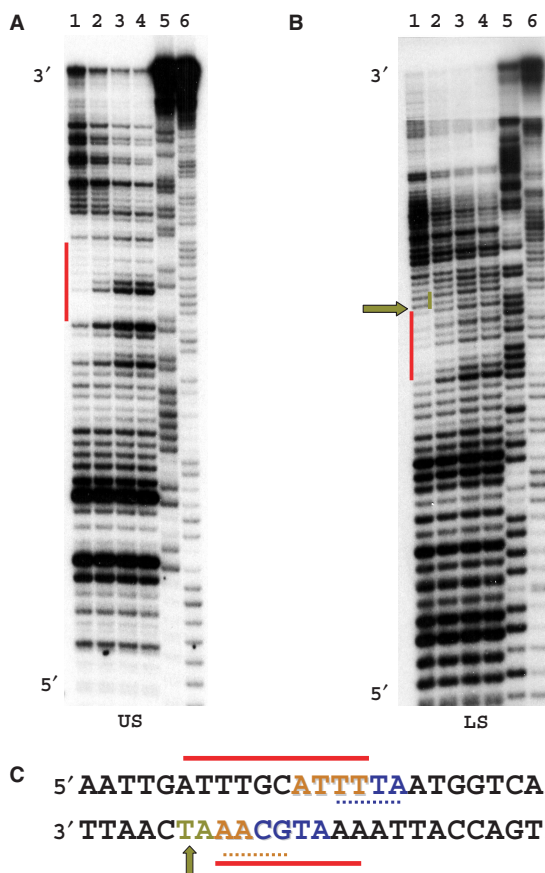
A MORE motif is characterized by two strong POU<sub>s</sub> anchoring sites on opposite DNA strands and on either side of the pseudo-dyad axis. The sequence of these binding sites is most often ATG(/A)C, but an ATNN motif is sufficient to establish the highly specific set of interactions with the conserved Gln and Thr residues of the POU<sub>s</sub> 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 POU<sub>s</sub>-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 POU<sub>s</sub>-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 POU<sub>s</sub>-binding sites as A12T13T14T15 on the upper strand and A12BT13BG14BC15B on the lower strand, the two corresponding POU<sub>h</sub>-binding sites can now be predicted as G14BC15BA16BA17B and T14T15T16A17 respectively, based on the known MORE motif organization



**Figure 3.** EMSA analysis of the interaction between the N-Oct-3 DBD and the DR $\alpha$  DNA. (A) Experiments were performed as previously described (22), except that the DNA concentration was set to 200 pM. Lane 1 corresponds to free DNA ('D'). The protein concentration was increased by 2-fold step increments starting from 38 pM (lanes 2–15). The assay at 610 pM DBD concentration (lane 6) resulted in 50% equimolecular C1 complex formation, indicating an apparent dissociation constant  $K_d$  of 0.61 nM. Accurate calculation gave an effective  $K_d$  of 0.5 nM. (B) In these assays, radiolabeled DR $\alpha$  DNA was mixed with an excess of cold probe, to a final 400 nM concentration. Lane 1 corresponds to free DNA ('D'). The protein concentration was increased by 2-fold step increments starting from 2.44 nM (lanes 2–9 and 11–15). An additional assay using the 437 nM intermediate protein concentration (lane 10) resulted in 100% equimolecular N-Oct-3 DBD/DR $\alpha$  DNA complex ('C1') formation, indicating an interaction stoichiometry of 400 nM. Note the non-cooperative mode of the N-Oct-3DBD homodimerization on the DR $\alpha$  DNA, as revealed by sequential 1:1 ('C1') and 2:1 ('C2') complex assembly. As the 2.560  $\mu$ M protein concentration induces 100% C2 complex formation (lane 13), it must be  $\geq 100$ -fold the apparent dissociation constant  $K_{d2}$  for the second site (35).

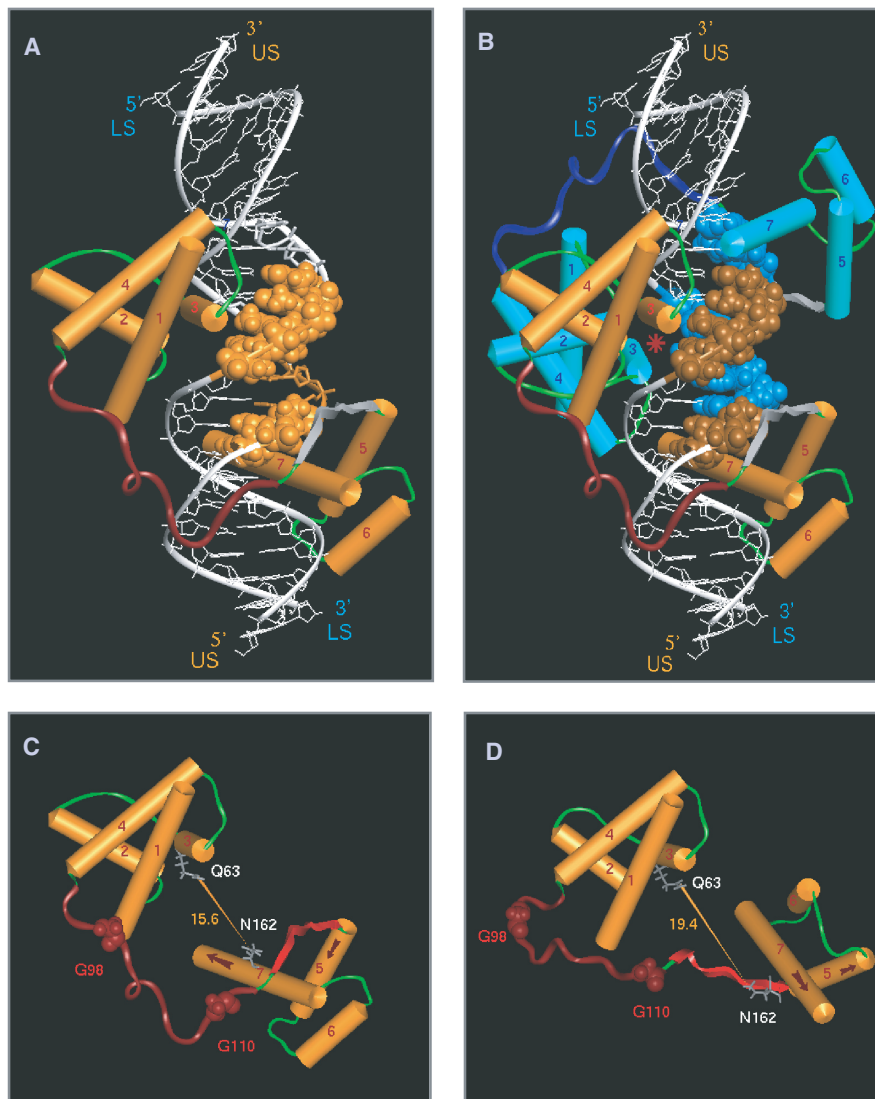
(11,12). In this mode, each POUh-binding site overlaps the POU<sub>s</sub>-binding site of the other monomer on the same strand (see Figure 4C and its legend). The extent of the DNase I footprint on the lower strand as a consequence of the first monomer binding designates G14BC15BA16BA17B as the first POUh-binding site (lane 1 in Figure 4B), and hence the A12T13T14T15 as the first POU<sub>s</sub>-binding site. This implies that the A12BT13BG14BC15B tetramer on the lower strand



**Figure 4.** Footprinting analysis of N-Oct-3 POU bound to the HLA DR $\alpha$  gene promoter. (A and B) Autoradiograms of 12% polyacrylamide denaturing gels showing the DNase I footprints on the upper ('US') and lower ('LS') strands of the DR $\alpha$  promoter fragment. Lanes 1: total footprint generated by the first POU binding (red color-coding). Lanes 2: cleavage products of a mixture comprising 75% complex and 25% free DNA. Lanes 3: cleavage products of a mixture comprising 25% complex and 75% free DNA. Lanes 4: free DNA cleavage products (in the absence of protein). Lanes 5–6: Maxam-Gilbert chemical sequencing references (cleavage after purine and pyrimidine residues, respectively). (C) Assignment of the POU<sub>s</sub> and POUh tetrameric binding sites deduced from the footprints (see text). The respective display codes for the first and the second N-Oct-3 POU domains binding sites are brown and blue. The first and second POUh tetrameric sub-sites are underlined in brown and blue, respectively, to compensate for the overlap with the POU<sub>s</sub>-binding sub-sites. The green marking in (A) and (B) points to an AT motif which does not interact with the DBD. The nucleotide numbering of the upper and lower strands in the 5'-3' direction is respectively 1–24 and 1B-24B.

is the second POU<sub>s</sub>-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<sub>H</sub> (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 POU<sub>s</sub>-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).



**Figure 5.** Modeling of the N-Oct-3 POU binding to the HLA DR $\alpha$  gene promoter. (A and B) Predicted structures of the 1:1 (A) and 2:1 (B) complexes between the N-Oct-3 DBD and the 24bp DR $\alpha$  DNA, based on the footprinting analysis. The nucleotides in contact with the first and second POU monomers are displayed in Van der Waals surface mode, using the same color-coding as in Figure 4C. The N-Oct-3 display code is: brown- or turquoise-colored cylinders for the  $\alpha$ -helices of the first or second POU respectively, a dark-brown or dark-blue colored coil for the linker of the first or second POU, and a gray-colored ribbon for the POUH N-terminal extension.. (C and D) Comparative analysis of the DR $\alpha$ -induced N-Oct-3 POU conformation (C) with the previously identified CRH-induced conformation (D). The two bound conformations can be interconverted by rotation around a virtual hinge Gly 98 – Gly 110 axis, taking the POU's orientation as a fixed reference. In (C) and (D), the two brown-colored arrows mark the direction of the first and third helices of POUH. The distance between the amide groups of two critical residues, Gln 63 and Asn 162, in the respective POU's and POUH DNA recognition helices ( $RH_{dist}$ ) is monitored in Å.

Now that the POU's 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 POU's insertion sites. For example the 'MORE+2' mode, corresponds to a 2bp 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 POU's 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 POU's as a fixed reference, reveals that

the two POUh sub-domain orientations can be superimposed by an  $\sim 180^\circ$  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  $\alpha$ -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  $\alpha$ -helical structure, whereas the replacement of the Oct-1 linker by that of N-Oct-3 resulted in the typical increase in the  $\alpha$ -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  $\alpha$ -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  $\alpha$ -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  $\beta$ -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  $\phi$  and  $\psi$  dihedral angles of Gly 98 and Gly 110, using the CRH-bound conformation as a starting structure. We found the combination of Gly 98 $\phi$  and Gly 110 $\psi$  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

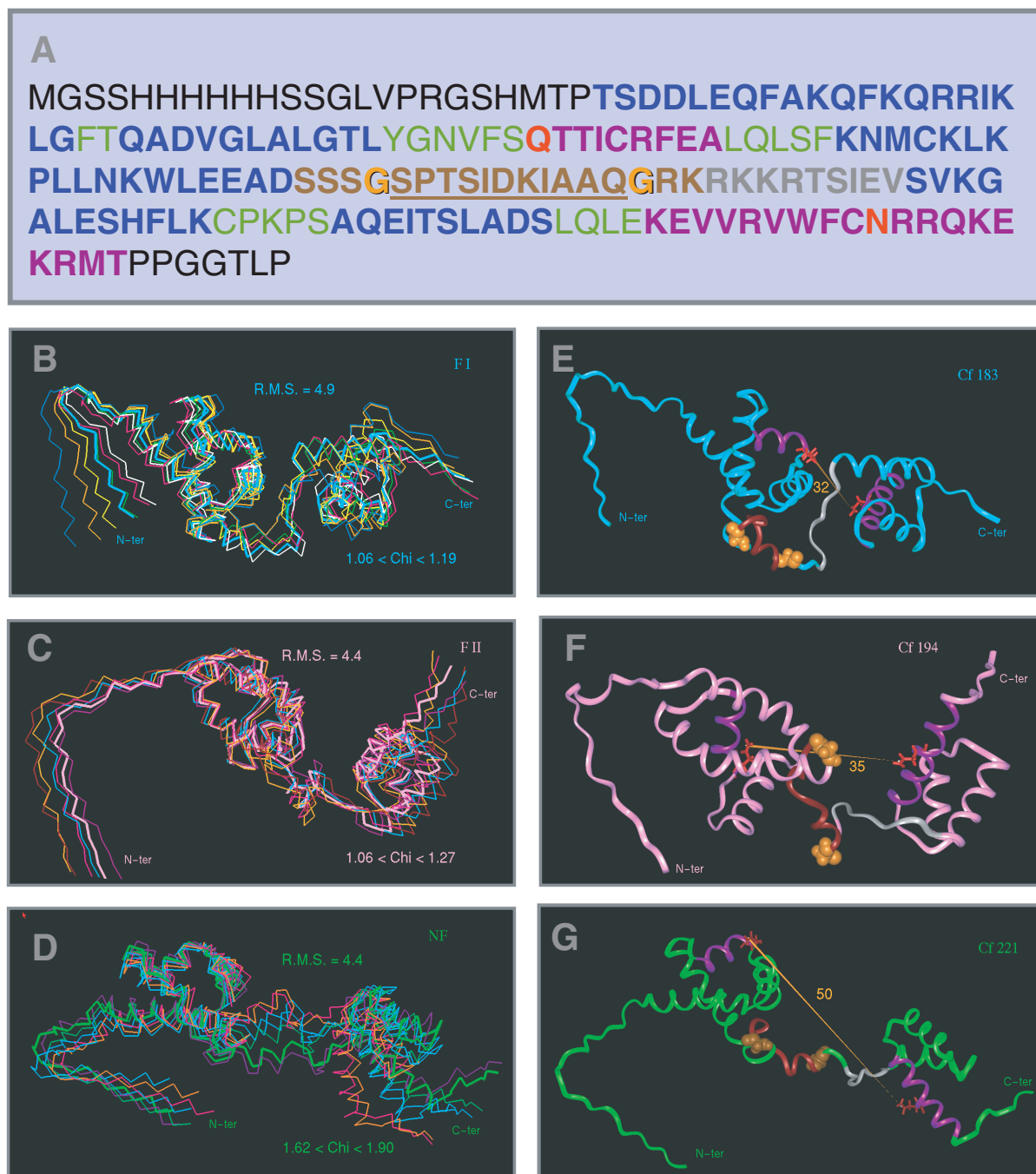
corresponding to different relative orientations of the POU and POUh sub-domains such as those displayed in Figure 6B–D. In order to identify potential free forms amongst these structures, we first compared their calculated radius of gyration ( $R_g$ ) to the free N-Oct-3 DBD hydrodynamic radius. To 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 equimolar N-Oct-3 DBD/DNA complexes (data groups 3). The structural parameters computed from the experimental data, including the radius of gyration ( $R_g$ ) and maximum particle dimension ( $D_{max}$ ), are displayed in Table 1. The estimated effective mass ( $M_{eff}$ ) of the free N-Oct-3 DBD agrees within experimental error with the value expected from the sequence ( $M_{seq}$ ), 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 CRY SOL and then compared to the experimental data. The accuracy of the fit was assessed by the discrepancy value  $\chi$  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 DR $\alpha$  DNA fragment and the N-Oct-3 DBD/DR $\alpha$  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 DR $\alpha$ -bound conformations with the experimental data for the free N-Oct-3 DBD yields slightly higher discrepancy values (see respective  $\chi$  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  $\chi$  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, 'FI' and 'FII',

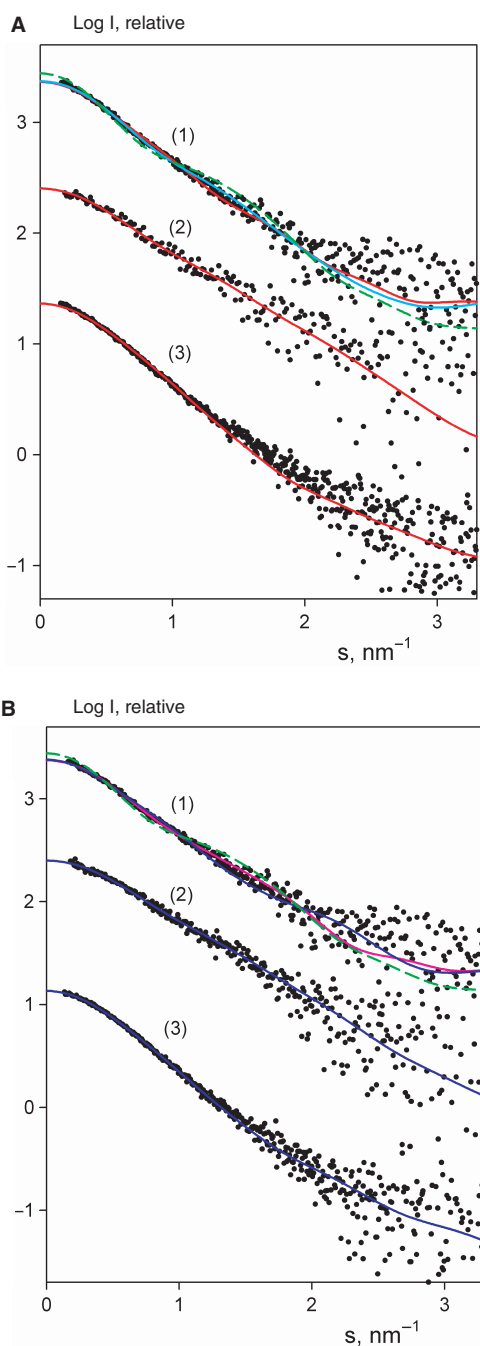




**Figure 6.** Conformational search by torsion driving. (A) Location of the linker (brown-coded) within the sequence of the N-Oct-3 DBD: the Gly 98 and Gly 110 residues (highlighted) flank the SPTSIDKIAAQ undecapeptide (underlined). Other critical features are the Gln 63 and Asn 162 residues (red-coded) in the respective POU 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  $C\alpha$  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 POU and POUh recognition helices in purple.  $RH_{\text{dist}}$  is monitored in Å.

defined by respective overall R.M.S. values of 4.9 Å (Figure 6B) and 4.4 Å (Figure 6C). Importantly, the  $\chi$  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 POU and POUh recognition helices ( $RH_{\text{dist}}$ ) within the 18–35 Å range. Interestingly, the conformers with the lowest  $RH_{\text{dist}}$  (Figure S3B) tend to be less energetically stable



**Figure 7.** Small angle X-ray scattering patterns. (A) (1) Experimental scattering pattern for the free N-Oct-3 DBD (dots), and computed scattering curves for the CRH-bound conformation (solid red line), the Cf 183 conformer (solid turquoise line) and the Cf 221 conformer (dashed green line). (2,3) Experimental (dots) and computed (color-coded) scattering patterns corresponding to the free CRH DNA (2) and the equimolecular N-Oct-3 DBD/CRH complex (3). (B) (1) Experimental scattering pattern for the free N-Oct-3 DBD (dots), and computed scattering curves for the DR $\alpha$ -bound conformation (solid blue line), the Cf 194 conformer (solid magenta line) and the Cf 221 conformer (dashed green line). (2,3) Experimental (dots) and computed (color-coded) scattering patterns corresponding to the free DR $\alpha$  DNA (2) and the equimolecular N-Oct-3 DBD/DR $\alpha$  complex (3). The scattering patterns have been offset in the logarithmic scale for better visualization.

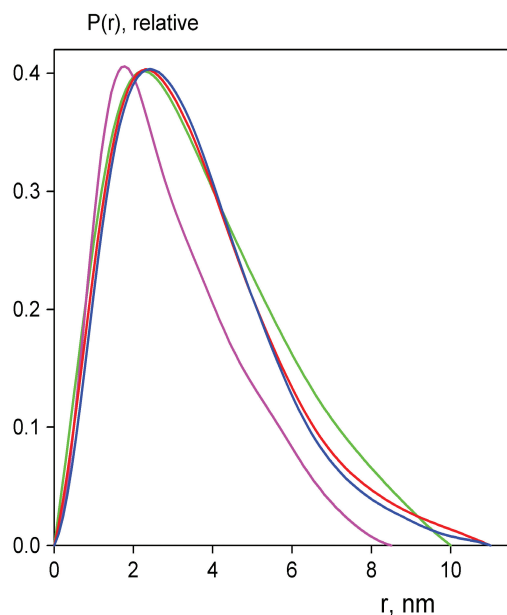
(Figure S3A), but are closer to the respective CRH- and DR $\alpha$ -bound conformations for which  $RH_{\text{dist}}$  is comprised within the 15–20 Å range (Figure 5C and D). Taken together, these results imply that the two populations of putative free forms, F1 and FII, most likely coexist, and also 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  $\chi$  values of 1.18 and 1.09 in Table 1), Cf 183 being the best FI representative (Figure 6E). Similarly, the fitted scattering curve of the DR $\alpha$ -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  $\chi$  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  $\chi$  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  $R_g$  (32 Å) and  $RH_{\text{dist}}$  (50 Å) values, cannot represent the free form and belongs to a large conformational family of extended structures, characterized by  $RH_{\text{dist}}$  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 DR $\alpha$ - 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).

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

Sample	$R_g$ (nm)	$D_{max}$ (nm)	$M_{eff}$ (kDa)	$M_{seq}$ (kDa)	$\chi$
N-Oct-3 DBD	$2.93 \pm 0.05$	$10.0 \pm 0.5$	$17 \pm 3$	20.0	1.18 (CRH-bound)
					1.23 (DR $\alpha$ -bound)
					1.09 (Cf 183)
					1.08 (Cf 194)
					1.90 (Cf 221)
CRH DNA	$2.37 \pm 0.04$	$8.5 \pm 0.5$	$14 \pm 2$	15.0	1.05
DR $\alpha$ DNA					0.82
N-Oct-3/CRH	$2.89 \pm 0.03$	$11.0 \pm 0.5$	$36 \pm 4$	35.0	1.09
N-Oct-3/DR $\alpha$	$2.85 \pm 0.03$	$11.0 \pm 0.5$	$34 \pm 4$	35.0	1.09

$R_g$ ,  $D_{max}$  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.  $\chi$  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  $\chi$  values have been calculated for the bound conformations, induced by the CRH or DR $\alpha$  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/DR $\alpha$  (blue) complexes.

## CONCLUDING REMARKS

Initially structural studies performed on the POU-type DNA-binding domain showed that individual POU 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 DR $\alpha$ -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<sub>dist</sub>') 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 DR $\alpha$ -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

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° RII3/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

- Chua,G., Robinson,M.D., Morris,Q. and Hughes,T.R. (2004) Transcriptional networks: reverse-engineering gene regulation on a global scale. *Curr. Opin. Microbiol.*, **7**, 638–646.
- Lardinois,A., Chalmel,F., Bianchetti,L., Sahel,J.A., Leveillard,T. and Poch,O. (2006) PromAn: an integrated knowledge-based web server dedicated to promoter analysis. *Nucleic Acids Res.*, **34**, W578–583.
- Sinha,S., Liang,Y. and Siggia,E. (2006) Stubb: a program for discovery and analysis of cis-regulatory modules. *Nucleic Acids Res.*, **34**, W555–W559.
- Gunewardena,S. and Zhang,Z. (2006) Accounting for structural properties and nucleotide co-variations in the quantitative prediction of binding affinities of protein-DNA interactions. *Pac. Symp. Biocomput.* 379–390.
- Garvie,C.W. and Wolberger,C. (2001) Recognition of specific DNA sequences. *Mol. Cell*, **8**, 937–946.
- DeWeese-Scott,C. and Moulton,J. (2004) Molecular modeling of protein function regions. *Proteins*, **55**, 942–961.
- Janin,J., Henrick,K., Moulton,J., Eyck,L.T., Sternberg,M.J., Vajda,S., Vakser,I. and Wodak,S.J. (2003) CAPRI: a critical assessment of predicted interactions. *Proteins*, **52**, 2–9.
- Latchman,D.S. (1999) POU family transcription factors in the nervous system. *J. Cell Physiol.*, **179**, 126–133.
- Klemm,J.D., Rould,M.A., Aurora,R., Herr,W. and Pabo,C.O. (1994) Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell*, **77**, 21–32.
- Herr,W. and Cleary,M.A. (1995) The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev.*, **9**, 1679–1693.
- Jacobson,E.M., Li,P., Leon-del-Rio,A., Rosenfeld,M.G. and Aggarwal,A.K. (1997) Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. *Genes Dev.*, **11**, 198–212.
- Remenyi,A., Tomilin,A., Pohl,E., Lins,K., Philippsen,A., Reinbold,R., Scholer,H.R. and Wilmanns,M. (2001) Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Mol. Cell*, **8**, 569–580.
- Scully,K.M., Jacobson,E.M., Jepsen,K., Lunyak,V., Viadiu,H., Carriere,C., Rose,D.W., Hooshmand,F., Aggarwal,A.K. *et al.* (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science*, **290**, 1127–1131.
- Fujii,H. and Hamada,H. (1993) A CNS-specific POU transcription factor, Brn-2, is required for establishing mammalian neural cell lineages. *Neuron*, **11**, 1197–1206.
- Eisen,T., Easty,D.J., Bennett,D.C. and Goding,C.R. (1995) The POU domain transcription factor Brn-2: elevated expression in malignant melanoma and regulation of melanocyte-specific gene expression. *Oncogene*, **11**, 2157–2164.
- Eisen,T.G. (1996) The control of gene expression in melanocytes and melanomas. *Melanoma Res.*, **6**, 277–284.
- Thomson,J.A., Murphy,K., Baker,E., Sutherland,G.R., Parsons,P.G., Sturm,R.A. and Thomson,F. (1995) The brn-2 gene regulates the melanocytic phenotype and tumorigenic potential of human melanoma cells. *Oncogene*, **11**, 691–700.
- Goodall,J., Wellbrock,C., Dexter,T.J., Roberts,K., Marais,R. and Goding,C.R. (2004) The Brn-2 transcription factor links activated BRAF to melanoma proliferation. *Mol. Cell Biol.*, **24**, 2923–2931.
- Alazard,R., Blaud,M., Elbaz,S., Vossen,C., Icre,G., Joseph,G., Nieto,L. and Erard,M. (2005) Identification of the ‘NORE’ (N-Oct-3 responsive element), a novel structural motif and composite element. *Nucleic Acids Res.*, **33**, 1513–1523.
- Thompson,R.C., Seasholtz,A.F. and Herbert,E. (1987) Rat corticotropin-releasing hormone gene: sequence and tissue-specific expression. *Mol. Endocrinol.*, **1**, 363–370.
- Wright,K.L. and Ting,J.P. (1992) In vivo footprint analysis of the HLA-DRA gene promoter: cell-specific interaction at the octamer site and up-regulation of X box binding by interferon gamma. *Proc. Natl Acad. Sci. USA*, **89**, 7601–7605.
- Millevoi,S., Thion,L., Joseph,G., Vossen,C., Ghisolfi-Nieto,L. and Erard,M. (2001) Atypical binding of the neuronal POU protein N-Oct3 to noncanonical DNA targets. Implications for heterodimerization with HNF-3 beta. *Eur. J. Biochem.*, **268**, 781–791.
- Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.
- Schuck,P. (1998) Sedimentation analysis of noninteracting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys. J.*, **75**, 1503–1512.
- Schuck,P. (1999) Sedimentation equilibrium analysis of interference optical data by systematic noise decomposition. *Anal. Biochem.*, **272**, 199–208.
- Koch,M.H.J. and Bordas,J. (1983) X-ray diffraction and scattering on disordered systems using synchrotron radiation. *Nucl. Instrum. Methods*, **208**, 461–469.
- Boulin,C.J., Kempf,R., Gabriel,A. and Koch,M.H.J. (1988) Data acquisition systems for linear and area X-ray detectors using delay line readout. *Nucl. Instrum. Meth. A*, **269**, 312–320.
- Gabriel,A. and Dauvergne,F. (1982) The localization method used at EMBL. *Nucl. Instrum. Meth.*, **201**, 223–224.
- Konarev,P.V., Volkov,V.V., Sokolova,A.V., Koch,M.H.J. and Svergun,D.I. (2003) PRIMUS – a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **36**, 1277–1282.
- Feigin,L.A. and Svergun,D.I. (1987) *Structure Analysis by Small-angle X-ray and Neutron Scattering* Plenum Press, New York.
- Svergun,D.I. (1993) A direct indirect method of small-angle scattering data treatment. *J. Appl. Crystallogr.*, **26**, 258–267.
- Guinier,A. (1939) La diffraction des rayons X aux tres petits angles; application a l’etude de phenomenes ultramicroscopiques. *Ann. Phys. (Paris)*, **12**, 161–237.
- Svergun,D.I. (1992) Determination of the regularization parameter in indirect transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 503.
- Svergun,D.I., Barberato,C. and Koch,M.H.J. (1995) CRYSOLO – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.

35. Long, K.S. and Crothers, D.M. (1995) Interaction of human immunodeficiency virus type 1 Tat-derived peptides with TAR RNA. *Biochemistry*, **34**, 8885–8895.
36. Tomilin, A., Remenyi, A., Lins, K., Bak, H., Leidel, S., Vriend, G., Wilmanns, M. and Scholer, H.R. (2000) Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. *Cell*, **103**, 853–864.
37. Dugast-Darzacq, C., Egloff, S. and Weber, M.J. (2004) Cooperative dimerization of the POU domain protein Brn-2 on a new motif activates the neuronal promoter of the human aromatic L-amino acid decarboxylase gene. *Brain Res. Mol. Brain Res.*, **120**, 151–163.
38. Blaud, M., Vossen, C., Joseph, G., Alazard, R., Erard, M. and Nieto, L. (2004) Characteristic patterns of N Oct-3 binding to a set of neuronal promoters. *J. Mol. Biol.*, **339**, 1049–1058.
39. Cheung, M.S., Klimov, D. and Thirumalai, D. (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl Acad. Sci. USA*, **102**, 4753–4758.
40. Rosgen, J., Pettitt, B.M. and Bolen, D.W. (2005) Protein folding, stability, and solvation structure in osmolyte solutions. *Biophys. J.*, **89**, 2988–2997.
41. George, R.A. and Heringa, J. (2002) An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng.*, **15**, 871–879.
42. Suzuki, M. (1989) SPXX, a frequent sequence motif in gene regulatory proteins. *J. Mol. Biol.*, **207**, 61–84.
43. Suzuki, M. and Yagi, N. (1991) Structure of the SPXX motif. *Proc. R. Soc. Lond. B. Biol. Sci.*, **246**, 231–235.
44. Kumaki, Y., Matsushima, N., Yoshida, H., Nitta, K. and Hikichi, K. (2001) Structure of the YSP TSPS repeat containing two SPXX motifs in the CTD of RNA polymerase II: NMR studies of cyclic model peptides reveal that the SPTS turn is more stable than SPSY in water. *Biochim. Biophys. Acta.*, **1548**, 81–93.
45. Gerstein, M., Anderson, B.F., Norris, G.E., Baker, E.N., Lesk, A.M. and Chothia, C. (1993) Domain closure in lactoferrin. Two hinges produce a see-saw motion between alternative close-packed interfaces. *J. Mol. Biol.*, **234**, 357–372.
46. Gerstein, M., Lesk, A.M. and Chothia, C. (1994) Structural mechanisms for domain movements in proteins. *Biochemistry*, **33**, 6739–6749.
47. Olah, G.A., Mitchell, R.D., Sosnick, T.R., Walsh, D.A. and Trewthella, J. (1993) Solution structure of the cAMP-dependent protein kinase catalytic subunit and its contraction upon binding the protein kinase inhibitor peptide. *Biochemistry*, **32**, 3649–3657.
48. Wriggers, W., Chakravarty, S. and Jennings, P.A. (2005) Control of protein functional dynamics by peptide linkers. *Biopolymers*, **80**, 736–746.
49. Nollmann, M., Byron, O. and Stark, W.M. (2005) Behavior of Tn3 resolvase in solution and its interaction with res. *Biophys. J.*, **89**, 1920–1931.
50. Augustus, A.M., Reardon, P.N., Heller, W.T. and Spicer, L.D. (2006) Structural basis for the differential regulation of DNA by the methionine repressor MetJ. *J. Biol. Chem.*, **281**, 34269–34276.
51. Choy, W.Y., Mulder, F.A., Crowhurst, K.A., Muhandiram, D.R., Millett, I.S., Doniach, S., Forman-Kay, J.D. and Kay, L.E. (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *J. Mol. Biol.*, **316**, 101–112.
52. Heller, W.T. (2005) Influence of multiple well defined conformations on small-angle scattering of proteins in solution. *Acta Crystallogr. D. Biol. Crystallogr.*, **61**, 33–44.
53. Svergun, D.I., Petoukhov, M.V. and Koch, M.H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.*, **80**, 2946–2953.
54. Cox, M., Dekker, N., Boelens, R., Verrijzer, C.P., van der Vliet, P.C. and Kaptein, R. (1993) NMR studies of the POU-specific DNA-binding domain of Oct-1: sequential <sup>1</sup>H and <sup>15</sup>N assignments and secondary structure. *Biochemistry*, **32**, 6032–6040.
55. Cox, M., van Tilborg, P.J., de Laat, W., Boelens, R., van Leeuwen, H.C., van der Vliet, P.C. and Kaptein, R. (1995) Solution structure of the Oct-1 POU homeodomain determined by NMR and restrained molecular dynamics. *J. Biomol. NMR*, **6**, 23–32.
56. van Leeuwen, H.C., Strating, M.J., Rensen, M., de Laat, W. and van der Vliet, P.C. (1997) Linker length and composition influence the flexibility of Oct-1 DNA binding. *EMBO J.*, **16**, 2043–2053.
57. Herr, W. and Cleary, M.A. (1995) The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev.*, **9**, 1679–1693.
58. Ahmad, S., Kono, H., Arauzo-Bravo, M.J. and Sarai, A. (2006) ReadOut: structure-based calculation of direct and indirect readout energies and specificities for protein-DNA recognition. *Nucleic Acids Res.*, **34**, W124–W127.
59. De Vuyst, G., Aci, S., Genest, D. and Culard, F. (2005) Atypical recognition of particular DNA sequences by the archaeal chromosomal MC1 protein. *Biochemistry*, **44**, 10369–10377.
60. Becker, N.B., Wolff, L. and Everaers, R. (2006) Indirect readout: detection of optimized subsequences and calculation of relative binding affinities using different DNA elastic potentials. *Nucleic Acids Res.*, **34**, 5638–5649.
61. Aeling, K.A., Opel, M.L., Steffen, N.R., Tretyachenko-Ladokhina, V., Hatfield, G.W., Lathrop, R.H. and Senechal, D.F. (2006) Indirect recognition in sequence-specific DNA binding by Escherichia coli integration host factor: the role of DNA deformation energy. *J. Biol. Chem.*, **281**, 39236–39248.
62. Swain, J.F. and Gierasch, L.M. (2006) The changing landscape of protein allostery. *Curr. Opin. Struct. Biol.*, **16**, 102–108.
63. Flores, S., Echols, N., Milburn, D., Hespeneheide, B., Keating, K., Lu, J., Wells, S., Yu, E.Z., Thorpe, M. et al. (2006) The Database of Macromolecular Motions: new features added at the decade mark. *Nucleic Acids Res.*, **34**, D296–301.
64. Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. and Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.*, **129**, 5656–5664.