The Carboxyl-terminal Nucleoplasmic Region of MAN1 Exhibits a DNA Binding Winged Helix Domain

Sandrine Caputo1, Joël Couprie1, Isabelle Duband-Goulet1, Emilie Konde1, Feng Lin3*, Sandrine Braud1, Muriel Gondry1, Bernard Gilquin1, Howard J. Worman3 and Sophie Zinn-Justin1#

Running Title: MAN1 carboxyl-terminal structure

1 DIEP/DSV, Bâtiment 152, CEA Saclay, 91191 Gif-sur-Yvette Cedex, France
2 Institut Jacques Monod - CNRS UMR 7592, Universités Paris6/7, 2 place Jussieu, 75251 Paris Cedex 05, France
3 Departments of Medicine and of Anatomy and Cell Biology, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, NY 10032, USA
* Present address: United Biomedical, Inc., 25 Davids Drive, Hauppauge, NY 11788, USA
$ S.C., J.C., I.D.G., B.G. and S.Z.J. are supported by grants from ‘Association Française contre les Myopathies’ (AFM, grants 8699, 9513 and 11591).
\(1^{t}\) H.J.W. was supported by a grant (MDA3711) from the Muscular Dystrophy Association.

# To whom correspondence should be addressed: Sophie Zinn-Justin, DIEP/DSV, Bâtiment 152, CEA Saclay, 91191 Gif-sur-Yvette Cedex, France; Phone: (33) (1) 69 08 30 26; Fax: (33) (1) 69 08 90 71; Email: szinn@cea.fr

Keywords: MAN1, NMR, winged helix motif, DNA, transcription regulation, R-Smad

MAN1 is an integral protein of the inner nuclear membrane, which interacts with nuclear lamins and emerin, thus playing a role in nuclear organization. It also binds to chromatin-associated proteins and transcriptional regulators, including the R-Smads, Smad1, Smad2 and Smad3. Mutations in the human gene encoding MAN1 cause sclerosing bone dysplasias, which sometimes have associated skin abnormalities. At the molecular level, these mutations lead to loss of the MAN1 / R-Smads interaction, thus perturbing transforming growth factor \(\beta\) superfamily signaling pathway. As a first step to understand the physical basis of MAN1 interaction with R-Smads, we here report the structural characterization of the carboxyl-terminal nucleoplasmic region of MAN1, which is responsible for Smad binding. This region exhibits an amino-terminal globular domain adopting a winged helix fold, as found in several Smad-associated sequence-specific DNA binding factors. Consistently, it binds to DNA through the positively charged recognition helix H3 of its winged helix motif. However, it does not show the predicted carboxyl-terminal U2AF homology domain in solution, suggesting that the folding and stability of such a domain in MAN1 depend upon binding to an unidentified partner. Modeling the complex between DNA and the winged helix domain shows that the regions involved in DNA binding are essentially distinct from those reported to be involved in Smad binding. This suggests that MAN1 binds simultaneously to R-Smads and their targeted DNA sequences.

The nuclear envelope separates the nucleus from the cytoplasm in eukaryotic cells. It consists of inner and outer nuclear membranes and nuclear pore complexes. The inner nuclear membrane is closely associated with the underlying chromatin and nuclear lamina. For many years, the nuclear envelope was thought to function mainly as an architectural stabilizer of the nucleus, participating in assembly and disassembly processes during mitosis. However, recent findings demonstrate that nuclear envelope proteins are involved in...
fundamental nuclear functions, such as chromatin organization and gene expression (1). Inherited or de novo mutations in genes encoding nuclear envelope proteins cause a wide range of human diseases (2). These findings emphasize the importance of understanding the functions of the nuclear envelope, in both physiologic and pathologic states.

MAN1 (also known as LEMD3) is a transmembrane protein of the inner nuclear membrane. It was originally identified as an antigen recognized by self antibodies from the serum of a patient with a collagen vascular disease (3,4). MAN1 is part of a protein complex essential for chromatin organization and cell division. It is analogous to the yeast protein SRC1, which may play a role in sister chromatid separation (5). In *Xenopus* embryos, overexpressed MAN1 induces the formation of a secondary neural axis, by binding directly to the MH2 domain of Smad1, Smad5 or Smad8, thus antagonizing bone morphogenetic protein signaling (6,7). Similarly, in humans, MAN1 binds to the MH2 domain of the R-Smads Smad1, Smad2 and Smad3, which mediate signaling by activin, bone morphogenic protein and transforming growth factor β (TGFβ1) (8,9). Heterozygous loss-of-function mutations in the human gene encoding MAN1 that disrupt this critical interaction cause sclerosing bone dysplasias characterized by increased bone density and sometimes skin abnormalities (10). Thus, mutations in a ubiquitous nuclear envelope protein give rise to relatively tissue-specific disease phenotypes, suggesting a role for MAN1 in the regulation of tissue specific gene transcription, as has similarly been proposed for nuclear lamins (11).

Sequence analysis indicates that MAN1 spans the inner nuclear membrane twice, resulting in a protein with amino-terminal and carboxy-terminal nucleoplasmic domains (4). At its amino-terminus, MAN1 contains a LEM domain that is present in several proteins including the inner nuclear membrane proteins lamina-associated polypeptide 2 and emerin (4,12-14). This domain interacts with the DNA and chromatin binding protein Barrier-to-Autointegration Factor (BAF) (15,16). The entire amino-terminal nucleoplasmic region of MAN1 also binds to the nuclear intermediate filaments lamin A and lamin B1 and to emerin (17). Thus, it mediates protein-protein interactions through contacts with the chromatin and the nuclear lamina. This domain is also necessary for efficient localization of MAN1 in the inner nuclear membrane (18).

We here examine the three-dimensional structure of the carboxy-terminal nucleoplasmic region of MAN1, which is responsible for the inhibition of physiologically important signaling pathways through an interaction with several R-Smads (8,9). This region of MAN1 has also been shown to bind to BAF and to the transcriptional regulators GCL and Btf (17). It comprises a first fragment (amino acids 655-758) showing sequence characteristics of a globular domain and a second fragment (amino acids 782-911) predicted to be an RRM-like protein interaction domain named U2AF homology motif (UHM) (19). We report that the first globular domain adopts a three-dimensional structure generally found in DNA binding regions of transcription factors. We show that indeed the entire carboxy-terminal region of MAN1 is involved in DNA binding and propose that this interaction is synergetic to the binding of MAN1 to different transcriptional regulators, particularly R-Smads.

**EXPERIMENTAL PROCEDURES**

**Sample Preparation**

All cloning procedures were performed according to standard methods (20). The region from amino acid 655 to amino acid 775 of MAN1 (MAN1Ca) was overexpressed in *E. coli* strain BL21 DE3 pLys S transformed with a construct generated in pGEX-4T-1 (Amersham Pharmacia Biotech, Inc.) that encodes glutathione-S-transferase and a thrombin cleavage site fused to MAN1Ca. The fusion protein was purified using Glutathione Sepharose 4B (Amersham Pharmacia Biotech, Inc.) and cleaved using thrombin protease. Because of the cloning strategy, the peptide resulting from the cleavage comprises additional residues from 1 to 5, MAN1 residues from 6 to 126 and again additional residues from 127 to 133. Uniformly labelled 15N protein was produced in minimum medium M9 containing 1 g.l-1 of (15NH4)2SO4 (Boehringer) as the nitrogen source. Uniformly labeled 13C/15N *Spirulina maxima* cyanobacteria. The resulting protein was
characterised by electrospray ionization mass spectroscopy and N-terminus sequencing. NMR samples (approximately 0.8 mM) were prepared in 50 mM phosphate buffer (pH 6.0) containing 150 mM NaCl in either 90% H2O/10% D2O or in 100% D2O, 1 mM ethylenediaminetetraacetic acid (EDTA), a protease inhibitor cocktail (Sigma-Aldrich), 1 mM tris(2-chloroethyl) phosphate (TCEP) and 1mM NaN3. TSP was added as a chemical shift reference. 13C and 15N chemical shifts were referenced indirectly to 3-(trimethylsilyl)[2,2,3,3-2H4] propionate (TSP), using the absolute frequency ratios.

The region of MAN1 from amino acid 658 to amino acid 910 (MAN1C B) was overexpressed in E. coli strain BL21 DE3 Star transformed with a plasmid that encodes ZZ fusion, a cleavage site for tobacco etch virus protease and MAN1CB (21). It was purified using immunoglobulin IgG Sepharose™ 6 Fast Flow (Amersham Pharmacia Biotech, Inc.) and cleaved using the catalytic domain of tobacco etch virus protease. Uniformly labeled 15N protein was produced in minimum medium M9 containing 1 gl-1 of (15NH4)2SO4 (Boehringer) as the nitrogen source. The resulting protein was characterised by electrospray ionization mass spectroscopy and amino-terminus sequencing. NMR samples of the protein at approximately 0.5 mM were prepared in 50 mM Tris buffer (pH 6.0) containing 150 mM NaCl in either 90% H2O/10% D2O or in 100% D2O, 1mM EDTA, a protease inhibitor cocktail (Sigma-Aldrich), 1 mM TCEP, and 1mM NaN3. TSP was added as a chemical shift reference. 15N chemical shifts were referenced indirectly to TSP, using the absolute frequency ratios.

MAN1CA and MAN1C0 mutants were generated with QuickChange Multi Site-Directed Mutagenesis Kit (Stratagene), following the manufacturer’s instructions.

NMR Spectroscopy

All assignment experiments of MAN1CA were performed at 30°C on a Bruker DRX-600 or DRX-900 spectrometers equipped with a triple-resonance probe according to the previously reported procedure (22). The NMR cross-peak volumes were used for structure calculation were measured on 5 NOESY experiments (a 15N-HSQC-NOESY in H2O and a 13C-HSQC-NOESY in D2O recorded at 900 MHz, with a 100 ms mixing time, at the European Large Scale Facilities in Utrecht, Netherlands, a 15N-HSQC-NOESY in H2O, with a 150 ms mixing time, and a 13C-HSQC-NOESY in D2O, with a 200 ms mixing time, and a 13C-HSQC-NOESY in the 13C aromatic region, with a 200 ms mixing time, all three recorded on a local 600 MHz spectrometer equipped with a triple resonance TXI cryoprobe). Phi torsion angle values were deduced from the analysis of the Hn-Ha and HMBC-J experiments (23,24). Hydrogen bound restraints were derived from slowly exchanging amide protons, identified by measuring the amide proton exchange rates from 1H-15N HSQC spectra recorded at different times on a protein sample dissolved in D2O. All spectra were processed with the programs Xwinnmr (Bruker) or NMRPipe (25) and analyzed using Felix (Accelrys).

Solution Structure Determination

We solved the three-dimensional solution structure of the carboxyl-terminal domain of MAN1CA using heteronuclear double and triple resonance NMR spectroscopy and molecular modeling. Coordinates and NMR restraints were deposited at the Protein Data Bank. The solution structure was calculated on the basis of the analysis of 3581 NMR crosspeaks (765 were picked on the 15N-HSQC-NOESY 600MHz, 756 on the 15N-HSQC-NOESY 900 MHz, 1241 on the 13C-HSQC-NOESY 900 MHz, 763 on the 13C-HSQC-NOESY 600 MHz and 56 on the 13C-HSQC-NOESY aromatic 600 MHz, respectively). A semi-automated iterative assignment procedure was applied for the assignment and the construction of the three-dimensional structures (26). A force field adapted to NMR structure calculation (file parallhdg.pro in CNS 1.0 (27)) was used. On this basis, 1850 restraints were generated. Thus, the mean number of distance restraints per residue yields 19.2 for the segment R6 to I111. Furthermore, using the program TALOS (28), the analysis of the Hn-Ha and the HMBC-J experiments, 169 couples of (φ,ψ) torsion angles were derived from backbone 1H, 15N, and 13C chemical shifts. Finally, 8 hydrogen bounds were imposed during the structure calculation. At the last step, 1000 structures were calculated and the 20 best structures were selected and refined with a standard energy function (CHARMM22), including an electrostatic energy term. This term was calculated with no net charge on the side-
chain atoms and with a distance-gated dielectric constant. Analysis of the 20 final structures showed that no distance violations larger than 0.5 Å were present and that the covalent geometry was respected. The rmsd around the average structure was 1.0 Å for the backbone and 1.6 Å for the heavy atoms.

**DNA preparation**

The 211 base pair DNA fragment was generated by polymerase chain reaction with a thermostable DNA polymerase (Promega) using a PTC-100 PCR System (MJ Research, Inc.). The 211 base pair DNA fragment, obtained from the Dra I and BamHI double digest of the plasmid pUC(357.4), was used as template and the 5'-AAATAGCTTAACTTTCATCAAGCAAG-3' and 5'-CCCCGGGCGAGCTCGAATTCC-3' oligonucleotides as sense and antisense primers. 5' end labelling with 32P-ATP and T4 polynucleotide kinase was performed according to standard protocols (20).

**Protein-DNA Interactions**

Proteins were diluted to the concentrations indicated in the Fig. 5 legend in 50 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 1 mM 4-(2-aminoethyl) benzenesulfonyl fluoride, 1 mM TCEP, 50 mM NaCl and 0.1% triton X-100. They were incubated with the radioactive 211 base pair DNA fragment for 3 hours at room temperature. Protein-DNA complexes were analyzed on 5% polyacrylamide gels at an acrylamide to bis-acrylamide ratio of 29/1 (w/w) in 12.5 mM Tris-HCl (pH 8.4), 95 mM glycine and 0.5 mM EDTA. After one hour pre-electrophoresis, samples were loaded onto the gels and resolved by electrophoresis at 70V for 2 hr. DNA was detected by autoradiography of dried polyacrylamide gels at -80°C using Biomax MR films (Kodak) and an intensifying screen. For affinity measurements, dried polyacrylamide gels were exposed to a phosphor screen, and measurements of the radioactive signals were performed with a STORM 860 scanner (Amersham) using ImageQuant software (Molecular Dynamics, Inc.).

**Molecular Modeling of the MAN1 Winged Helix / DNA Complex**

The winged helix domain of MAN1 was superimposed onto the winged-helix domain of MecI-DNA complex using the program Sybyl 6.9. The superimposition was done by fitting the Cα atoms of the following segments 6-23, 24-25, 26-37, 38-53, 60-64 and 65-68 of MecI on 15-32, 44-45, 46-57, 60-75, 82-86 and 93-96 of MAN1C

(A) (resulting rmsd 3.4 Å). These segments correspond to the secondary elements of the winged helix motif. Then, in the MecI-DNA complex, the winged helix domain of MecI was replaced by the corresponding domain of MAN1. To remove the small number of steric clashes, 1000 steps of steepest descents energy minimization were run using the program CHARMM (29). Side-chain conformations were allowed to vary only on MAN1 winged helix domain, keeping the entire DNA fixed and maintaining the MAN1 winged helix domain backbone by progressively decreasing harmonics constraints. A MAN1 winged helix / DNA complex model with no bad clashes and reasonable interaction energy was thus obtained.

**RESULTS**

The MAN1 Carboxyl-terminal Nucleoplasmic Region Contains a Well-structured Domain.

The carboxyl-terminal nucleoplasmic region of MAN1, comprising residues 655 to 910, exhibits two potential globular domains: the region from amino acid 655 to 758 is conserved in all MAN1 analogs (called MAN1 and LEM2 proteins) (30) and shows globular domain sequence characteristics (31); the region between amino acids 785 and 910 is found only in MAN1 proteins and shows an unusual RRM motif called UHM, predicted to adopt an α/β fold and to be involved in protein recognition (19). We have recorded the 1H-15N HSQC spectra of MAN1C

(A) (amino acid 655 to amino acid 775 of MAN1) and MAN1C

(B) (amino acid 658 to amino acid 910), corresponding to the first predicted globular domain and the entire carboxyl-terminal nucleoplasmic region of MAN1, respectively. The spectrum of MAN1C

(A) is well dispersed, confirming that the region from amino acid 655 to 775 adopts a globular structure in solution. Superposition of MAN1C

(A) and MAN1C

(B) spectra shows that all peaks of MAN1C

(A) are found at identical chemical shifts in MAN1C

(B) spectrum (Fig. 1). This suggests that the three-dimensional structure of MAN1C

(A) is not affected by the presence of the putative UHM
domain. Furthermore, only about 40 additional peaks are found on the MAN1C\textsuperscript{B} spectrum compared to the MAN1C\textsuperscript{A} spectrum. These peaks are mostly clustered between 7.5 and 8.5 ppm in the proton dimension. This is not consistent with the presence of a well-structured α/β UHM domain. A slow proteolysis of MAN1C\textsuperscript{B} is observed at 300K and analysis of the resulting peptides by SDS-PAGE and amino-terminal sequencing revealed that only MAN1C\textsuperscript{A} is resistant to proteolysis. Finally, expression in E. coli of the second putative domain alone (amino acid 776 to 910) yielded a rapidly aggregating protein. All these data suggest that only the first globular domain adopts a stable three-dimensional structure on the NMR time scale (millisecond).

**The Fragment from Amino Acid 666 to Amino Acid 750 of MAN1 Adopts a Well-defined α/β Structure**

The three-dimensional structure of MAN1C\textsuperscript{A} was characterized using heteronuclear NMR. In the following, residues belonging to this fragment are numbered from 6 to 126 (residues 1 to 5 and 127 to 133 are additional amino acids linked to the plasmid construction). Backbone and sidechain \(^1\)H, \(^{13}\)C and \(^{15}\)N resonance assignments were performed from residue 6 to 112 (22). Next, molecular modeling calculations were carried out in order to obtain a structure consistent with the 1811 NOESY derived proton-proton distances and the 169 dihedral angle values deduced from TALOS (Table 1). The region between residues 17 and 101 (amino acid 666 to 750, using the MAN1 numbering) adopts a well defined α/β fold. The backbone root-mean square deviation (rmsd) calculated on this fragment with respect to the mean coordinate yields 1.0 Å. The three-dimensional structure of the region between residues 17 and 101 is constituted of three α-helices H1 (residues 17 to 37), H2 (residues 47 to 54) and H3 (residues 58 to 76) and three β-strands S1 (residues 44 to 46), S2 (residues 81 to 88) and S3 (residues 92 to 99) organized into a H1-S1-H2-H3-S2-S3 topology (Fig. 2). Thus, the amino-terminal half of the domain is mainly α-helical, whereas the carboxyl-terminal half is composed of two large β-strands arranged in a twisted anti-parallel β-sheet.

The helices form a three helix bundle. They are amphipathic, and their hydrophobic core is constituted by M20, V21, I24, I25, V27, and L28 (H1), V50, L54 (H2) and W67, A70, V71, L74 (H3). The three stranded β-sheet is packed onto the three helix bundle. The α/β interface is mainly hydrophobic. It is constituted by L28 (H1), I47 and V50 (H2), W67 and V71 (H3) that contact the hydrophobic face of the three-stranded β-sheet composed of M45 (S1), V81 (S2) and W96, W98 (S3). This interface also involves a hydrogen bond network between the backbone of L28 (H1), the side chain of N32 (H1), the side chain of W98 (S3) and the backbone of R80 (S2).

**The α/β Structure of MAN1 Corresponds to a Winged Helix Domain**

The three-dimensional structure of the fragment between residues 17 and 101 of MAN1C\textsuperscript{A} was submitted to the DALI server. Its structure is close to the three-dimensional structure of numerous proteins belonging to the winged helix superfamily, as defined by SCOP. Winged helix domains are mainly used for DNA recognition (32). The winged helix domain of MAN1 is structurally similar to several DNA binding domains belonging to transcription factors (PDB code = 1OKR, Z-score=4.0, sequence identity=5%, Fig. 3A; 1P4X, Z-score=3.8, sequence identity=9%) and to histone H5 (1HST, Z-score=3.5, sequence identity=9%). However, recently, winged helix domains were also described as protein-RNA (19,33,34) and protein-protein (4,35) interaction modules. The MAN1 winged helix domain is also structurally similar to a protein-protein interaction module belonging to the ESCTR-II endosomal trafficking complex (1U5T; Z-score=4.0, sequence identity=5%, Fig. 3B). In this structure, contacts between different winged helices are mediated, on one side, by the concave surface formed by helices H2, H3 and β-sheet S2-S3, and on the other side, by the loop connecting H3 to S2.

**Both the Winged Helix Domain and the Entire Carboxyl-terminal Nucleoplasmic Region of MAN1 Interact with DNA**

Calculation of the electrostatic potential at the surface of the MAN1 winged helix domain shows that the amino-terminal region of H3 and the tip of the β-sheet are mainly positively charged (Fig. 4). In particular, the sequence RKKMKKVWDR found in H3, which corresponds to the recognition helix in known
complexes of winged helix domains with DNA (32), presents six positively charged residues (Fig. 4B). We tested the importance of these six residues for the binding of MAN1CA and MAN1CB to DNA by gel shift retardation assay. Therefore, we produced three types of mutants R60A-K61A-K62A, K64A-K65A and R69A of the recognition helix H3. Interaction with a 211 base pair linear DNA fragment was tested by electrophoresis in a 5% polyacrylamide gel. Fig. 5 shows that the DNA forms complexes with wild type proteins as demonstrated by delayed migration compared to naked DNA (Fig. 5A and 5B, lanes 2 to 5). The apparent affinities of the two peptides for DNA are 150 ± 13 nM for MAN1CA and 50 ± 18 nM for MAN1CB. The appearance of discrete bands with MAN1CB as well as the higher affinity of MAN1CB for DNA suggests that the second domain may play a role in the stability of the complex. Consistently, for all mutants of MAN1CA, a complete loss of DNA binding is observed (Fig. 5A, lanes 6 to 17), whereas the mutant of MAN1CB, which exhibits a single mutation R69A shows only a 10-fold decrease of affinity for DNA (Fig. 5B, lanes 14 to 17). Yet the two mutants (R60A-K61A-K62A and K64A-K65A) of MAN1CB present a complete loss of DNA binding (Fig. 5B, lanes 6 to 13). Thus, several positively charged residues of the recognition helix H3 are involved in the binding of the carboxyl-terminal region of MAN1 to DNA. The poorly structured UHM domain also contributes to the affinity of MAN1 for DNA.

Modeling of the Complex of MAN1 Winged Helix Domain with DNA

Nine crystal structures of winged helix domains in complex with B-DNA are available. We analyzed these structures in order to get insight into the structural determinants of the DNA recognition by the winged helix motif. Fig. 6A shows the structural alignment of the recognition helix residues in these complexes. Interestingly, the arginine residue that makes multiple hydrogen bonds with a guanine base is roughly at the center of helix H3. This residue is flanked on both sides by hydrophobic regions. Before and after these two hydrophobic regions, several hydrophilic residues of the H3 helix make contacts with DNA. The MAN1 sequence presents all these characteristics: a central R69, flanked by two hydrophobic regions (V66-W67 and A70-V71) with hydrophilic residues on both sides. In order to construct a structural model of the complex between the MAN1 winged helix domain and DNA, we selected the Mec1/DNA complex (36) because of the high DALI score found for Mec1 (PDB code=1OKR). Our domain was globally fitted onto the winged helix structure of Mec1 in complex with DNA using the DALI alignment, and R69 of MAN1 was adjusted to R51 of Mec1. In such a model, the winged helix domain of MAN1 is in close contact with the DNA fragment (Fig. 6B). In particular, the side chain of R69 interacts with the Gua0 (our DNA reference point) by two H-bonds. Moreover, at the beginning of helix H3, K62 and K65 form a salt bridge with the backbone phosphates at Ade+2 and Ade+3, respectively. Therefore, in this complex, one residue of each of three mutated segments of MAN1 interacts with DNA. Finally, as in other complexes of winged helix domains with DNA, the wing1 (the turn between β2 and β3) of MAN1’s winged helix makes contact with the minor groove: F93 at the tip of the β-sheet is positioned in this groove. Thus, the calculated model is consistent with our experimental data as well as with the published structures of winged helix transcription factors bound to DNA.

DISCUSSION

The Carboxyl-terminal Nucleoplasmic Region of MAN1 Contains a Well-folded Winged Helix Domain but Does Not Show the Predicted Stable UHM domain

NMR analysis of the solution structure of the carboxyl-terminal nucleoplasmic region of MAN1 showed that this region is composed of a DNA binding winged helix domain followed by a poorly stable/folded peptide fragment. Superimposition of the one-dimensional 1H NMR spectra of MAN1CA and MAN1CB shows that the two spectra are similar (data not shown). The low contribution of the second putative UHM domain to the one-dimensional 1H NMR spectrum of MAN1CB indicates that this domain is not completely disordered. Gel filtration experiments carried out on MAN1CB do not suggest extensive oligomerization or aggregation of this fragment, which could be responsible for the lack of NMR
signal. More likely, this lack of signal could be a consequence of a partial unfolding or conformational exchange process. However, the carboxyl-terminal nucleoplasmic region of MAN1 is at least partially functional in the conditions used in this study, as it is capable of binding to the MH2 domain of Smad2 and Smad3 (data not shown). Within the three structurally characterized UHM domains, two of them (from proteins U2AF35 (37) and U2AF65 (38)) are in complex with a peptide ligand. Furthermore, the UHM domain of U2AF35 adopts its stable three-dimensional structure upon binding to its ligand (39). The third domain, belonging to the protein TgDRE (40), was characterized alone and was rapidly aggregating (K. Frénal, personal communication). Similarly, we suggest that the folding and the stability of the UHM domain of MAN1 depend upon binding to a yet unknown biological partner.

The consequence of the unusual behaviour of the predicted UHM domain is that most winged helix peaks can be easily identified in the HSQC spectrum of the entire carboxyl-terminal region of MAN1. Essentially no chemical shift perturbations were found between the spectrum of the winged helix domain alone and the spectrum of the winged helix domain within the entire carboxyl-terminal region. This suggests that there is no extensive contact between the winged helix and the predicted UHM domains of MAN1.

The Carboxyl-terminal Nucleoplasmic Region of MAN1 Binds to Both DNA and R-Smads

We carried out gel retardation assays using a 211 base pair linear DNA fragment and increasing quantities of either MAN1C A or MAN1CB. Clearly, both domains bind to DNA. The recognition helix H3 of the winged helix domain is critical for this binding, as it is in other published winged helix domain/DNA complexes. The apparent affinity for DNA is only three-fold higher for MAN1C B as compared to MAN1C A, showing that the contribution of the poorly folded MAN1C B region to DNA binding is not essential.

MAN1C B also shows a significant affinity for the MH2 domain of R-Smads (8,9). MAN1C A, which spans from amino acids 655 to 775 of MAN1, or the predicted UHM domain alone (amino acid 776 to amino acid 910 of MAN1) are not sufficient for R-Smad binding. However, the region from amino acid 730 to amino acid 910 of MAN1, comprising the β2 and β3 strands of the winged helix domain, the linker between this domain and the putative UHM domain and the UHM domain itself, binds to Smad2 and Smad3 (9). Interestingly, it has been shown that several homeodomain and winged helix transcription factors recruit activated Smad2 to distinct promoter elements through an interaction between the MH2 domain of Smad2 and a common Smad binding motif (41). This motif is located after the DNA binding domain of these transcription factors and comprises the following sequence consensus: P-P-N-K-T/S-I/V-X3-h-X4-h, where h is a hydrophobic residue (42). In the case of MAN1, an analogous MH2 binding motif is found within the linker, between P777 and L785, two residues after MAN1C A and a few residues before the first β-strand of the putative UHM domain. This motif P-P-N-S-L-T-X2-L could participate to the MAN1/Smad2 interaction. Such a hypothesis is consistent with the reported unfolded structure of several free MH2 binding ligands. More generally, as both the two last β-strands of the winged helix domain or the linker region and the putative UHM domain are necessary and sufficient for R-Smad binding, and as these regions are essentially distinct from those suggested as critical for DNA binding, i.e. the recognition helix and the tip of the β-hairpin (Figure 6B), we propose that MAN1C B can bind simultaneously to DNA and the MH2 domain of R-Smads.

Conclusions

In the nucleus, Smads bind to transcriptional coactivators and promoter regions and play a role in the transcription of numerous genes regulated by TGF-β superfamily members (43). The MH1 domain of Smads interacts with DNA. However, the affinity of Smads for DNA is relatively low and Smads require other sequence-specific binding factors to bind efficiently to the promoters of certain responsive genes (44). The MH2 domain of Smads interacts with several such factors, which include a plethora of non-homologous proteins including FoxH1, Mixer, TGIF, CBP, AML1, Ski and SIP1 (43). MAN1C A adopts a three-dimensional structural motif found in several of these proteins and binds to DNA with an apparent affinity of 10^-7 M through its helix H3. This same helix mediates interaction with DNA in classical DNA binding winged helix domains.
Furthermore, it was recently proposed that several Smad-associated proteins share a common MH2-binding motif (41). This motif is partially found in MAN1 25 amino acid residues after the carboxyl-terminus of the winged helix domain at the amino-terminus of the putative UHM domain. If MAN1 uses this motif to interact with R-Smads, then it can be predicted that the highly homologous human LEM2 protein, which shares the LEM domain, the two transmembrane segments and the winged helix domain with MAN1, but lacks the linker and putative UHM regions, does not act as an antagonist of the Smad-mediated signaling pathways activated by BMP, TGF-ß or activin. Consistently, Gotzmann and co-workers recently reported that they have not been able to detect antagonism of R-Smad-mediated signaling activity by LEM2 (30). As LEM2 exhibits a highly positively charged helix H3 completely similar to MAN1 helix H3 in its winged helix domain, it could potentially regulate MAN1 DNA binding by interacting with the same specific DNA sequences. Finally, MAN1C B binds to others transcriptional regulators as GCL and Btf (17). A competition between the different transcription regulators associated to MAN1 might also play a role in the complex regulation of the transcription of genes potentially targeted by the winged helix domain of MAN1.

REFERENCES

FOOTNOTES

1Abbreviations used in this paper: TGFβ: transforming growth factor-β; UHM: U2AF homology motif; EDTA: ethylenediaminetetraacetic acid; TCEP: tris(2-chloroethyl) phosphate; TSP: 3-(trimethylsilyl)[2,2,3,3-2H4] propionate.

ACKNOWLEDGEMENTS

We gratefully acknowledge Roger Genet, Marie Courcon, Mireille Moutiez and Cedric Masson for their help during protein purification. We are grateful to Philippe Savarin and Flavio Toma who kindly lent us their 600 MHz spectrometer. The 900 MHz spectra were recorded at the SON NMR Large Scale Facility in Utrecht, which is funded by the ‘Access to Research Infrastructures’ program of the European Union.
FIGURE LEGENDS

Fig. 1 Superimposition of $^1$H-15N HSQC spectra of MAN1C$^A$ and MAN1C$^B$. Overlay of the $^1$H-15N HSQC spectra obtained at 300°C for MAN1C$^A$ (blue) and for MAN1C$^B$ (red). Both protein fragments were prepared in 50 mM Phosphate/Tris buffer and 150 mM NaCl at pH 6.0.

Fig. 2 Three-dimensional structure of the MAN1 region (666-750). Cartoon representation of residues 666-750 where the secondary structures are colored in green for $\alpha$-helix H1, magenta for $\alpha$-helix H2, red for $\alpha$-helix H3 and yellow for the three $\beta$-strands.

Fig. 3 Superimposition of the WH domain of MAN1 ($\alpha$-helices in blue and $\beta$-strands in yellow) with those of (A) a domain of the bacterial transcriptional repressor MecI (PDB code = 1OKR, $\alpha$-helices in cyan and $\beta$-strands in red) and (B) a module of the ESCRT-II endosomal trafficking complex (PDB code = 1U5T, $\alpha$-helices in cyan and $\beta$-strands in red).

Fig. 4 Electrostatic properties at the surface of MAN1 WH domain. (A) Surface representation of the electrostatic potential at the surface of MAN1 WH domain (from positively charged, in blue, to negatively charged, in red). (B) Cartoon representation of the backbone (in grey) and the positively charged side chains (in blue) of the MAN1 WH domain.

Fig. 5 Comparative DNA binding of MAN1C$^A$ and MAN1C$^B$ to DNA. (A) Increasing concentrations of MAN1C$^A$ wild-type and mutants peptides were incubated with a 211 bp DNA fragment at a concentration of 18 nM in 50 mM NaCl: 10-fold (lanes 2, 6, 10 and 14), 20-fold (lanes 3, 7, 11 and 15), 40-fold (lanes 4, 8, 12 and 16) and 80-fold (lanes 5, 9, 13 and 17) molar excess of MAN1C$^A$ (655-775). Lane 1 indicates the mobility of naked DNA. (B) Increasing concentrations of MAN1C$^B$ wild-type and mutants peptides were incubated with a 211 bp DNA fragment at a concentration of 18 nM in 50 mM NaCl: 5-fold (lanes 2, 6, 10 and 14), 10-fold (lanes 3, 7, 11 and 15), 20-fold (lanes 4, 8, 12 and 16) and 40-fold (lanes 5, 9, 13 and 17) molar excess of MAN1C$^B$ (658-910). Lane 1 indicates the mobility of naked DNA.

Fig. 6 Modeling the complex between the WH domain of MAN1 and DNA. (A) Structural alignment of $\alpha$-helices H3 of winged helix domains binding to B-DNA (PDB code is indicated in parenthesis). Black boxes indicate the limits of H3 sequences. Underlined residues contact DNA through hydrogen bonds; residues in bold interact with bases. The central arginine residue is boxed in a red square; on both sides of this arginine, blue and orange boxes surround hydrophobic and hydrophilic residues, respectively. (B) Model of the complex between the WH domain of MAN1 and DNA based on the analysis the MecI/DNA complex (1SAX).
**Table 1 Structural statistics for the human MAN1 655-775 fragment.**

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of NOE assigned</td>
<td>3581</td>
</tr>
<tr>
<td>Number of constraints</td>
<td></td>
</tr>
<tr>
<td>Distance constraints</td>
<td>1811</td>
</tr>
<tr>
<td>Dihedral constraints</td>
<td>169</td>
</tr>
<tr>
<td>Number of violations</td>
<td></td>
</tr>
<tr>
<td>NOE distance &gt; 0.5Å</td>
<td>0.1 ± 0.3</td>
</tr>
<tr>
<td>Dihedral restraints &gt; 10°</td>
<td>0.05 ± 0.2</td>
</tr>
<tr>
<td>Rms deviation from idealised covalent geometry</td>
<td></td>
</tr>
<tr>
<td>Bond (Å)</td>
<td>0.018 ± 0.00024</td>
</tr>
<tr>
<td>Angle (°)</td>
<td>3.5 ± 0.05</td>
</tr>
<tr>
<td>Improper (°)</td>
<td>2.4 ± 0.2</td>
</tr>
<tr>
<td>Rms deviation on restraints</td>
<td></td>
</tr>
<tr>
<td>Distance (Å)</td>
<td>0.043 ± 0.0016</td>
</tr>
<tr>
<td>Dihedral (°)</td>
<td>1.3 ± 0.09</td>
</tr>
<tr>
<td>Energy (kcal/mol)</td>
<td></td>
</tr>
<tr>
<td>Van der Waals</td>
<td>182 ± 13</td>
</tr>
<tr>
<td>Electrostatic</td>
<td>-379 ± 25</td>
</tr>
<tr>
<td>Ramachandran plot (%) (residues 17:101)</td>
<td></td>
</tr>
<tr>
<td>Most favoured region</td>
<td>90.0</td>
</tr>
<tr>
<td>Additionally allowed region</td>
<td>9</td>
</tr>
<tr>
<td>Coordinate precision (residues 17:101) (Å)</td>
<td></td>
</tr>
<tr>
<td>Backbone</td>
<td>1.0 ± 0.1</td>
</tr>
<tr>
<td>Heavy atoms</td>
<td>1.6 ± 0.16</td>
</tr>
</tbody>
</table>

The van der Waals energy is calculated with a Lennard-Jones potential. The electrostatic energy is calculated with no net charge on side-chain atoms and a distance-gated dielectric constant. CHARMM22 parameters are used.
Fig. 1 Superimposition of $^1$H-$^{15}$N HSQC spectra of MAN1C$^A$ and MAN1C$^B$
Fig. 2 Three-dimensional structure of the MAN1 region (666-750)
Fig. 3 Superimposition of WH domain of MAN1 on similar domain found by DALI server
Fig. 4 Calculation of the electrostatic potential at the surface of MAN1 WH domain
Fig. 5 Interaction of the C-terminal nucleoplasmic region of MAN1 with DNA
Fig. 5 Interaction of the C-terminal nucleoplasmic region of MAN1 with DNA
Fig. 6 Modeling of the complex between the WH domain of MAN1 and DNA
Fig. 6 Modeling of the complex between the WH domain of MAN1 and DNA
The carboxyl-terminal nucleoplasmic region of MAN1 exhibits a DNA binding winged helix domain
Sandrine Caputo, Joëlle Couprie, Isabelle Duband-Goulet, Emilie Konde, Feng Lin, Sandrine Braud, Muriel Gondry, Bernard Gilquin, Howard J. Worman and Sophie Zinn-Justin

J. Biol. Chem. published online April 28, 2006

Access the most updated version of this article at doi: 10.1074/jbc.M601980200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts