The GCM domain is a Zn-coordinating DNA-binding domain

Serge X. Cohen[a,1], Martine Moulin[a,1], Oliver Schilling[b], Wolfram Meyer-Klaucke[b], Jörg Schreiber[c], Michael Wegner[c], Christoph W. Müller[a,*]

[a] *European Molecular Biology Laboratory, Grenoble Outstation, B.P. 181, 38042 Grenoble Cedex 9, France*
[b] *European Molecular Biology Laboratory, Hamburg Outstation, Notkestr. 85, D-22603 Hamburg, Germany*
[c] *Institut für Biochemie, Universität Erlangen-Nürnberg, Fahrstrasse 17, D-91054 Erlangen, Germany*

Received 14 June 2002; revised 22 July 2002; accepted 1 August 2002

First published online 4 September 2002

Edited by Hans Eklund

Abstract Glial cells missing (GCM) proteins form a small family of transcriptional regulators involved in different developmental processes. They contain a DNA-binding domain that is highly conserved from flies to mice and humans and consists of approximately 150 residues. The GCM domain of the mouse GCM homolog a was expressed in bacteria. Extended X-ray absorption fine structure and particle-induced X-ray emission analysis techniques showed the presence of two Zn atoms with four-fold coordination and cysteine/histidine residues as ligands. Zn atoms can be removed from the GCM domain by the Zn chelator phenanthroline only under denaturating conditions. This suggests that the Zn ions are buried in the interior of the GCM domain and that their removal abolishes DNA-binding because it impairs the structure of the GCM domain. Our results define the GCM domain as a new type of Zn-coordinating, sequence-specific DNA-binding domain.

Key words: Transcription factor; DNA-binding domain; Zn coordination; EXAFS

1. Introduction

The *gcm* gene family codes for a small family of eukaryotic transcription factors [1]. The first family member was discovered in *Drosophila melanogaster* where it functions as a switch between neuronal and glial cell fate during early development of the nervous system [2–4]. The two murine homologs mGCMa and mGCMb do not appear to be involved in glia cell development. Mouse GCMa is required for proper formation of the labyrinth in the developing placenta where the exchange of nutrients and gases between mother and embryo takes place [5,6]. Mouse GCMb is essential for parathyroid gland development [7]. In both cases, targeted deletion leads to loss of respective tissue.

Drosophila gcm, and mouse *gcma* and *gcmb* genes encode transcription factors with lengths of 504, 436 and 504 amino acid residues, respectively. They all contain a DNA-binding domain of about 150 amino acid residues at the N-terminal end and one or two *trans*-activation domains in the C-terminal half of the molecule. Sequence homology between different species is restricted to the DNA-binding domain. This GCM domain is sufficient for DNA binding. In vitro binding site selection experiments using *Drosophila* GCM defined an octameric motif with the sequence 5′-ATGCGGGT-3′ [8,9], which is also recognized by the murine homologs mGCMa and mGCMb with apparently similar specificities [10,11]. The same target sequences are also present in potential GCM target genes [8,12]. GCM binds as a monomer to its DNA-binding site, which is rather unusual as most transcription factors bind their target sites as dimers. The GCM domain contains a large number of cysteine and histidine residues, which had led to the suggestion that the GCM domain is a Zn-containing DNA-binding domain. However, the sequence signature of the cysteine and histidine residues does not resemble any classical Zn-binding DNA-binding domains like Zn fingers or Zn clusters [13]. In addition Zn-ion chelators which effectively abolish DNA-binding of Zn-finger proteins had failed to show any effect on DNA-binding of the GCM domain, which had led to the conclusion that in GCM domains an important structural role of Zn ions is unlikely [10].

Alanine-scan mutagenesis of Arg, Lys, His and Cys residues had given some insight into residues critical for DNA-binding of GCM [10]. Further insight into the function of GCM domains has been hampered due to the lack of structural information. We have started an extensive biophysical characterization of the GCM domain including co-crystallization experiments with DNA oligonucleotides as a prerequisite for an X-ray analysis. Different crystal forms of GCM/DNA complexes were obtained and a crystallographic analysis is currently in progress. X-ray fluorescence spectroscopy of initial GCM/DNA co-crystals unexpectedly revealed the presence of Zn ions in the GCM domain that was further explored by extended X-ray absorption fine structure (EXAFS) and microPIXE (particle-induced X-ray emission) analyses. These experiments unambiguously demonstrate that the GCM domain contains two Zn ions tightly coordinated by cysteine and histidine residues and required for its DNA-binding activity. Our results define the GCM domain as a new class of Zn-containing, sequence-specific DNA-binding domain.

2. Materials and methods

2.1. Expression and purification of the GCM domain of mGCMa

The gene coding for residues 1–174 of murine GCMa was cloned into plasmid pET28a (Novagen) which contains a N-terminal His-tag followed by a thrombin cleavage site. For expression we used *Escherichia coli* strain BL21(DE3) codon+ (Stratagene). After induction by

*Corresponding author. Fax: (33)-476-207199.
E-mail address: mueller@embl-grenoble.fr (C.W. Müller).

[1] Both authors contributed equally.

PII: S0014-5793(02)03257-X

0.5 mM IPTG cells were grown for 4 h at 37°C. For cell lysis 8 g of bacterial wet paste was sonicated in 30 ml lysis buffer (500 mM NaCl, 20 mM Tris–HCl (pH 7.9), 1 mM β-mercaptoethanol) and subsequently centrifuged for 90 min at 250 000×*g*. The clear supernatant was incubated for about 1 h with 3 ml Ni-NTA resin (Qiagen). Subsequently the resin was transferred into a column and washed with 20 column volumes of washing buffers (20 mM Tris (pH 7.9), 8 mM imidazole, 1 M NaCl for 10 volumes, 500 mM NaCl for 5 volumes, 250 mM NaCl for the last 5 volumes) and eluted in a buffer containing 250 mM imidazole, 200 mM NaCl and 20 mM Tris (pH 7.9). The fractions containing the protein were pooled together and dialyzed against a buffer containing 20 mM Tris (pH 7.9), 200 mM NaCl and 10 mM dithiothreitol (DTT). The histidine tag was cleaved overnight with human plasma thrombin (Sigma) using 1 U per 0.15 mg of protein at room temperature. After cleavage the protein still contains six residues preceeding the mGCMa sequence. As a final step the protein was applied to a Superdex-75 column (Pharmacia) equilibrated with 200 mM NaCl, 20 mM Tris (pH 7.9), 10 mM DTT. The pooled fractions were finally concentrated to 13 mg/ml.

2.2. EXAFS experiment

GCM-domain solution of 0.6 mM (13 mg/ml) in 200 mM NaCl, 20 mM sodium phosphate (pH 7.9), 10 mM DTT was filled into sample holders covered with Kapton windows, frozen in liquid nitrogen and kept at 30 K during the experiment. X-ray absorption spectra at the Zn K-edge were gathered in fluorescence mode at the EMBL bending magnet beamline D2 (DESY, Hamburg, Germany) using a Si(111) double crystal monochromator, a focussing mirror and a 13 element Ge solid-state detector. Energy calibration was achieved by Bragg reflections of a static Si(220) crystal in back reflection geometry [14]. Data was reduced with the EXPROG software package (C. Hermes and H.F. Noilting, EMBL-Hamburg) with $E_{O,Zn} = 9660$ eV. Analysis of the EXAFS spectrum with constrained model refinement was performed with EXCURV98 [15]. Ligand types and coordination numbers were varied manually; the software refined distances, Debye–Waller factors and the Fermi energy offset.

2.3. microPIXE analysis

To determine the exact stoichiometry of the Zn ions present in the GCM domain we used PIXE analysis, where a strong proton beam induces X-ray fluorescence characteristic for the elements present in the sample. In proteins of known sequence the number of sulfurs in cysteines and methionine residues can be used as an internal calibration. In this analysis sodium and chlorine had to be removed from the sample (both these elements give overlapping X-ray emission signals with the sulfur), as well as any other source of sulfur than the protein (namely DTT or β-mercaptoethanol). Hence, the protein buffer, 200 mM NaCl, 20 mM Tris (pH 7.9), 10 mM DTT, was exchanged for a solution containing 200 mM KBr and 20 mM potassium phosphate (pH 6.5) using a centricon-10 (Millipore). The scanning proton microprobe developed at Oxford University possesses a sensitivity of 1–10 ppm per weight and a proton beam diameter of 1 μm [16]. For the experiment 1 μl of protein solution was dried on a Kimfoil film. Subsequently, the sulfur signal was used to localize the protein in the dried drop and this region of the drop was scanned for 15 min in the proton beam integrating the signal in this region.

2.4. Circular dichroism (CD)

For the CD experiments and X-ray fluorescence experiments the protein was incubated for 20 min with 5 mM 1,10-phenanthroline prior to gel filtration. The gel filtration column was equilibrated with 200 mM KF and 20 mM ammonium phosphate (pH 6.5) suitable for CD spectroscopy. The CD experiment was performed in a solution containing 0.4 mg/ml GCM-domain protein in a 0.1 cm thick cuvette (200 μl) using a Jasco J-810 spectrometer (sensitivity settings 100 millidegrees, data pitch 0.5 nm). Wavelength scans were performed from 180 to 250 nm at a speed of 50 nm/min. The signal to noise ratio was increased by averaging 10 spectra. The secondary structure content was calculated using Jasco spectrometer software. The protein melting curve was obtained with 2 mg/ml GCM-domain solution at the minimum of 224 nm using a temperature ramp from 10 to 95°C with a ramping speed of 20°C/h.

2.5. Electrophoretic mobility shift assay (EMSA) experiments

EMSAs were performed with 5–30 ng bacterially expressed, purified GCM domain and 0.5 ng of ^{32}P-labeled probe containing the consensus GCM-binding site as described [10]. In some samples, the purified GCM domain was incubated for 20 min at room temperature with 1–5 mM 1,10-phenanthroline prior to the addition of probe. In denaturation/renaturation experiments, the GCM domain was incubated for 2 min at 90°C in the presence or absence of 1,10-phenanthroline, followed by immediate placement on ice or slow cooling to 4°C at a rate of 0.1°C per 10 s. Samples were loaded onto native 4% polyacrylamide gels and electrophoresed in 0.5×TBE (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.3). After drying, gels were exposed overnight to X-ray film.

3. Results

3.1. Protein expression and purification

The DNA-binding domain of mGCMa was previously expressed in eukaryotic cells [10]. We now show that the GCM domain can be also expressed in *E. coli* and subsequently purified to high purity and concentrated to about 13 mg/ml. Using size exclusion chromatography the protein elutes as a single, symmetric peak with an elution volume corresponding to a molecular mass of 20 kDa (calculated mass: 20 996 Da), which suggests that the protein is a monomer in solution. This result is consistent with the observation that the GCM domain binds its DNA-target sites as a monomer [10]. The purification protocol yields about 1 mg of pure protein from 1 g of bacterial wet paste that binds specifically to its target sites as confirmed by EMSA. The possibility to express the GCM domain as a soluble, monomeric and active protein, which can be concentrated to high concentrations and co-crystallized with DNA oligonucleotides, confirms the GCM domain as a stable, independent folding domain and makes it a prime candidate for further structural studies. Co-crystals with DNA duplexes containing the target site have been obtained and an X-ray analysis is currently in progress.

3.2. EXAFS and microPIXE analysis of the GCM domain reveals the presence of two Zn ions

The primary sequence of the GCM domain contains a large number of conserved cysteine and histidine residues often involved in the coordination of Zn ions (Fig. 1). However, the cysteine/histidine pattern in the primary sequence of mGCMa H-X_{12}-H-X_8-C-X_5-C-X_3-C-X_{14}-C-X_{11}-C-X_{2-4}-C-X_8-C-X_2-H-X_{23}-H-X_2-H does in no way resemble classical Zn fingers or Zn_2Cys_6 binuclear clusters. Furthermore treatment with 1,10-phenanthroline and EDTA as chelating agents did not interfere with DNA binding [10]. Whether Zn is contained within the GCM domain therefore requires further investigation.

An X-ray fluorescence spectrum recorded at the MAD beamline BM14 at the ESRF showed a clear absorption signal at a wavelength of 1.28 Å (9.7 KeV) indicating the presence of Zn in GCM/DNA co-crystals. To further characterize the Zn-binding site an EXAFS experiment was performed at beamline D2 at the EMBL Outstation, Hamburg, Germany. EXAFS experiments provide information on the type and number of surrounding ligands. However, if several Zn sites are present in the same molecule EXAFS does not discriminate the different binding sites but rather integrates them. EXAFS does therefore not determine the exact number of Zn ions bound per protein molecule. The Zn K-edge EXAFS spectrum was obtained in good quality up to $k = 15.5$ Å^{-1} (Fig. 2). The double peak between 4.0 and 5.0 Å^{-1} indicates multiple backscattering by imidazole rings ligating the Zn atom. Model refinement revealed the presence of three coordinating

A

```
                          1       10        20        30        40
mGCMa   -----------------MELDDFDPEDKEILSWDINDVKLPQNVKTTDWFQEWPDSYVK
mGCMb   ----------MPADSTQDEDAVLSYGMKLTWDINDPQMPQEPTHFDHFREWPDGYVR
dGCM    MVLNGMPITMPVPMPVPMPVPSPPATKSRVAIDWDINDSKMPS-VGEFDDFNDWSNGHCR
                                    . ***** ..*     * * .*     .

          50        60        70        80        90        100
mGCMa   HIYSSDDRNAQRHLSSWAMRNTNNHNSRILKKSCLGVVVCSRDCSTEEGRKIYLRPAICD
mGCMb   FIYSSQEKKAQRHLSGWAMRNTNNHNGHILKKSCLGVVVCARACALKDGSHLQLRPAICD
dGCM    LIYSVQSDEARKHASGWAMRNTNNHNVNILKKSCLGVLLCSAKCKLPNGASVHLRPAICD
         ***     *..* * **********  *********..*.  *    *  . *******

          110               120       130       140       150       160
mGCMa   KARQKQQRKSCPN--CNGPLKLIPCRGHGGFPVTNFWRHDGRFIFFQSKGEHDHPRPETK
mGCMb   KARLKQQKKACPN--CHSPLELVPCRGHSGYPVTNFWRLDGNAIFFQAKGVHDHPRPESK
dGCM    KARRKQQGKQCPNRNCNGRLEIQACRGHCGYPVTHFWRRDGNGIYFQAKGTHDHPRPEAK
        *** *** * ***  *.  * .  **** *.***.*** **  *.**.** *******.*

          170
mGCMa   LEAEARRAMKKVHM
mGCMb   SETEGRRSALKRQM
dGCM    GSTEARRLLAGGRR
          .* **     .
```

B

GCM-domain	H-X_{12}-H-X_8-C-X_5-C-X_3-C-X_{14}-C-X_{11}-C-$X_{2\text{-}4}$-C-X_8-C-X_2-H-X_{23}-H-X_2-H
Known domain type	Consensus sequence
Zn-fingers (Cys_2His_2)	-C-$X_{2\text{-}4}$-C-X_{12}-H-$X_{3\text{-}5}$-H-
Nuclear receptor (Cys_8)	-C-X_2-C-X_{13}-C-X_2-C-X_{15}-C-X_5-C-X_{12}-C-X_4-C-
Gal4 (Cys_6)	-C-X_2-C-X_6-C-X_6-C-X_2-C-X_6-C-
GATA-1 (Cys_4)	-C-X_2-C-X_{17}-C-X_2-C-
P53 tumour suppressor	-C-X_3-H-X_{59}-C-X_3-C-

Fig. 1. Conserved cysteine and histidine residues in the GCM domain. A: Sequence alignment of the GCM domain of murine GCMa with murine GCMb and *D. melanogaster* GCM. Conserved and conservatively substituted residues are marked with asterisks and dots, respectively. Conserved cysteine and histidine residues in the GCM domain are depicted on yellow and red background. B: Sequence signatures of Zn-coordinating, sequence-specific DNA-binding domains with known 3D-structures.

sulfur atoms at 2.31 ± 0.01 Å and one coordinating imidazole group at 2.03 ± 0.02 Å (Fig. 2, Table 1). This model has an *R*-factor of 36.2%. The metal ligand distances are in good agreement with distances found for four-fold coordinated structural Zn sites in proteins and model complexes [17]. The average Zn coordination of GCM therefore consists of one ligating histidine residue and three ligating cysteine or methionine residues. Since cysteine is far more abundant in Zn coordination than methionine [18], the coordinating sulfur atoms represent most likely cysteine residues.

Table 1
Average Zn coordination of the GCM domain of mGCMa

Ligand	Atom[a]	*N*	*r* (Å)	$2\sigma^2$ (Å^2)
His[b]	N_1	1	2.03 (2)	0.006 (1)
	C_2	1	2.95 (2)	0.018 (1)
	C_2	1	3.07 (2)	0.018 (1)
	N_3	1	4.09 (2)	0.018 (1)
	C_3	1	4.18 (2)	0.018 (1)
	S_1	3	2.31 (1)	0.006 (1)

N: coordination number, *r*: mean interatomic distance to Zn center, $2\sigma^2$: Debye–Waller factor. *R*-factor: 36.2%, Fermi energy offset: −7.9 eV. Numbers in brackets indicate the uncertainty of the last digit. The *R*-factor for the EXAFS data is defined as:

$$R = \sum_i^N \frac{1}{\sigma_i}(|\chi_i^{\exp}(k) - \chi_i^{\text{model}}(k)|) \times 100\%$$

with

$$\frac{1}{\sigma_i} = \sum_j \frac{k(j)^n}{k(j)^n |\exp(j)|}$$

[a]Atoms with equal indices were grouped together for the collective refinement of the Debye–Waller factors and the distance to the metal center in the case of the first coordination shell.
[b]Histidine is represented by an imidazole group. Its structure was taken from the EXCURV98 software and used for rigid body refinement.

The number of Zn atoms per protein molecule was determined using a microPIXE analysis (Table 2). The analysis showed the presence of 1.7 (±0.15) Zn atoms per protein molecule using as internal standard the 12 sulfurs present in seven cysteines and five methionines. The estimated error of the Zn content results from the errors in the sulfur and Zn concentration and from the error in fitting the thickness of the sample (E. Garman, personal communication). The analysis also showed no detectable amounts of other metals present in the protein sample.

Taken both microPIXE and EXAFS results together we conclude that the GCM domain contains two Zn sites, either both with a 3Cys/1His coordination or one site with a 4Cys and the second with a 2Cys/2His coordination. According to

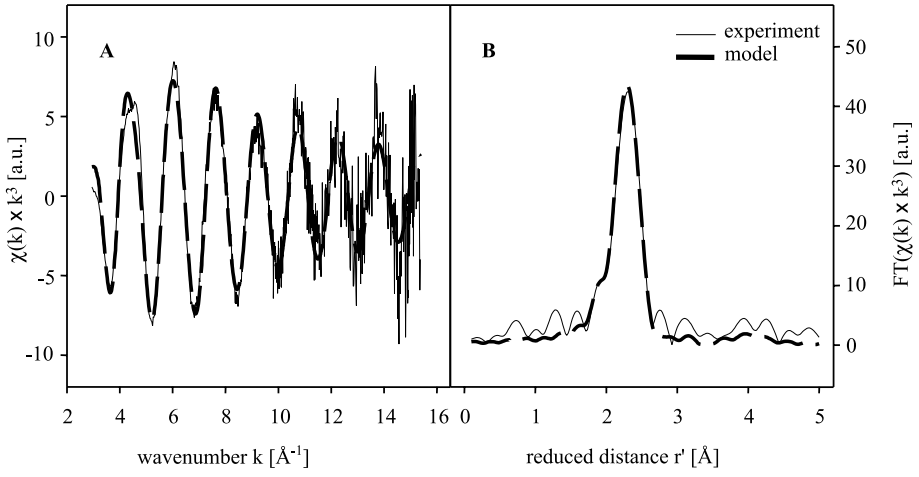

Fig. 2. EXAFS spectrum of mGCMa. A: EXAFS spectrum, B: Fourier transformation. Data at the Zn K-edge was gathered as described in Section 2. Experimental data are represented by thin lines, the theoretical spectrum of the model given in Table 1 is represented by thick, gray lines. $\chi(k)$: EXAFS amplitude, r′: metal-ligand distance corrected for first shell phase shifts, a.u.: arbitrary units.

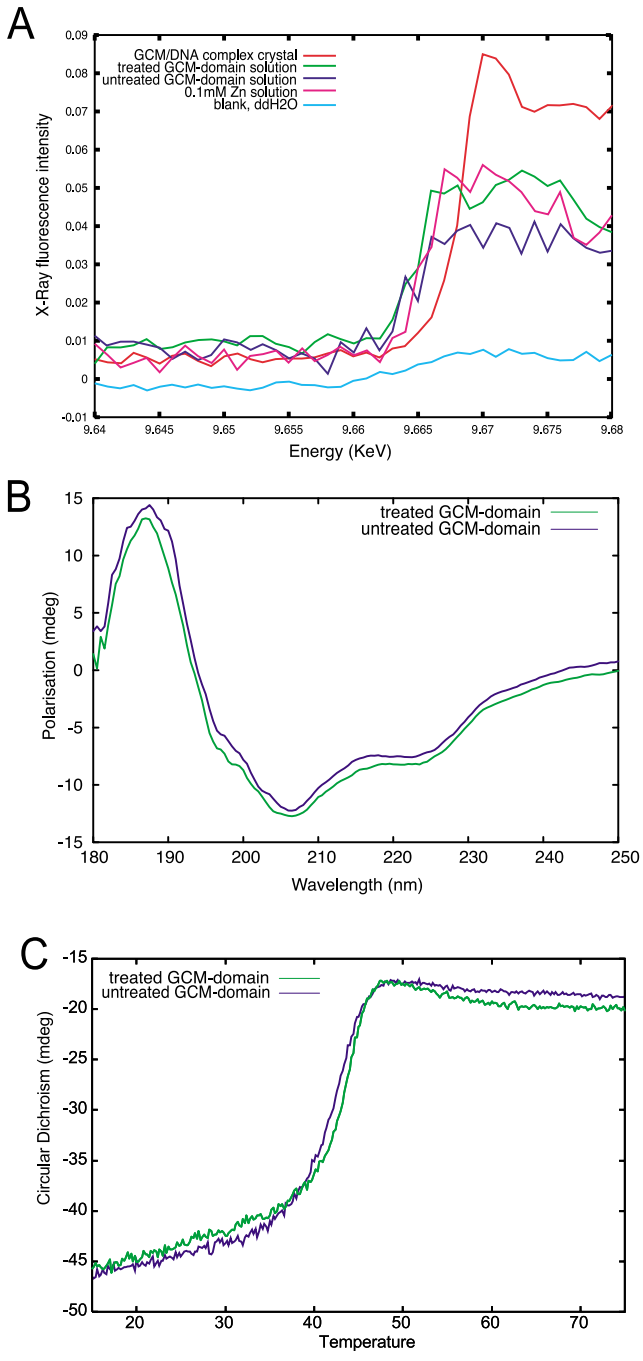

Fig. 3. Comparison of GCM-domain protein before and after 1,10-phenanthroline treatment. A: Fluorescence absorption spectra of GCM-domain/DNA co-crystals recorded at ESRF beamlines BM14 and protein solutions (0.1 mM, 2 mg/ml) with and without 1,10-phenanthroline treatment recorded at ESRF beamline ID29. The concentration of the protein in the crystal is approximately 25 mM. The deviations from the theoretical value of the Zn absorption at 9.659 KeV (1.2836 Å) are most likely caused by calibration errors of the monochromators. Measurement performed on beamline ID29 were multiplied by a factor of 5 to compare them with the spectrum of the crystal recorded at BM14. B: CD spectrum of mGCMa GCM domain treated with 5 mM 1,10-phenanthroline compared with untreated sample. Both samples were recorded at 0.4 mg/ml protein concentration and no normalization was applied. C: Thermal stability of mGCMa GCM domain treated with 5 mM 1,10-phenanthroline compared with untreated sample. The CD melting curve was taken at 224 nm corresponding to the second minimum of spectrum B. No normalization was applied to both curves.

Table 2
microPIXE spectrum analysis

Element	Conc. by weight (ppm or %)	Error estimate (±%)
Al	17.6%	0
P	10.4%	0
S	3.0%	1
K	31.0%	0
Ca	2517.0 ppm	29
Mn	81.0 ppm	57
Fe	186.0 ppm	40
Zn	84522.0 ppm	5
Br	44.6%	1

For the above listed elements intensities of the X-ray fluorescence at the K-edge were extracted by peak fitting and integration of the counts in the peaks to obtain the weight in part per million (ppm) or as percentage. Phosphate was present in the protein buffer solution. Potassium was present both in the buffer and in the matrix. Bromine was present in the matrix, the L-edge of bromine overlaps with the aluminum K-edge resulting in an aluminum signal even in its absence.

the microPIXE analysis one site could be fully occupied and the second occupied by 70% or both sites partially occupied. Our results identify six cysteines of seven present in the GCM domain of mGCMa as direct Zn ligands. A model containing three partially occupied Zn atoms can be excluded because of the limited number of cysteine residues.

3.3. *Zn ions of the GCM domain can be only removed under denaturating conditions*

EXAFS and microPIXE analysis show the presence of two Zn atoms in the GCM domain. Unlike other Zn-containing transcription factors, treatment of native GCM with the high-affinity Zn chelator 1,10-phenanthroline did not abolish nor reduce DNA-binding activity in EMSA when tested at final concentrations of both 1 mM and 5 mM (Fig. 4A, and data not shown).

To further investigate this apparent discrepancy the GCM domain was incubated for 20 min with 5 mM 1,10-phenanthroline before the last gel filtration step (see Section 2). After

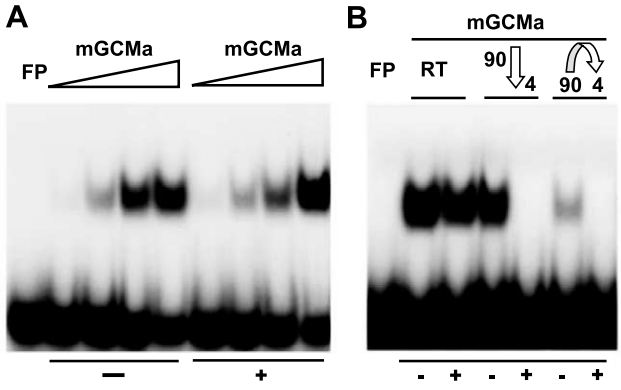

Fig. 4. Zn dependent DNA-binding of the GCM domain. A: Binding of increasing amounts of the murine GCMa domain (5, 10, 20 and 30 ng, indicated by triangles) to its cognate radiolabeled site was analyzed by EMSA after preincubation in the absence (−) or presence (+) of 1 mM 1,10-phenanthroline. B: EMSA of the murine GCMa domain (20 ng) on its binding site following various pretreatment protocols. Protein was kept at room temperature (RT) or was heated to 90°C in the presence (+) or absence (−) of 1 mM 1,10-phenanthroline followed by fast (straight arrow) or slow (bent arrow) cooling to 4°C. FP, free probe with no protein added.

this treatment we used X-ray fluorescence to test the presence of Zn in the GCM domain. In addition, we compared the CD spectra of native GCM-domain protein and protein treated with phenanthroline and the melting curves of the two samples. The results of the X-ray fluorescence experiments are presented in Fig. 3A. Protein treated with phenanthroline and native protein show very similar X-ray fluorescence spectra, which indicates that under our experimental conditions phenanthroline is not able to remove the Zn ions. The CD spectrum of native GCM protein (Fig. 3B) is typical for a protein with high β-sheet content of 37% and an α-helix content of 10%. The melting temperature for the native protein was calculated to 41°C (Fig. 3C). As expected from the X-ray fluorescence experiments no difference of the CD spectra and thermal melting curves between native and phenanthroline-treated proteins can be observed.

In further experiments, the GCM domain was heat-denatured in the presence and absence of 1,10-phenanthroline and renatured by slow cooling or immediately put on ice. As evident from EMSAs, the GCM domain was still able to bind to its octameric consensus motif after this treatment provided that 1,10-phenanthroline was absent during heating. However, when 1,10-phenanthroline was present during denaturation, no DNA-binding activity was recovered (Fig. 4B) suggesting that 1,10-phenanthroline effectively removes Zn ions from a denatured GCM domain, thereby interfering with proper refolding.

4. Discussion

Our EXAFS and microPIXE analyses show the presence of two Zn ions in the GCM domain which are tetrahedrally coordinated by cysteine and histidine residues. This is an unexpected result because in earlier experiments the strong metal chelators 1,10-phenanthroline and EDTA showed no effect on the DNA-binding activity of mGCMa which made any important role of Zn ions in the GCM domain unlikely [10]. The two Zn ions present in the GCM domain coordinate six out of the seven conserved cysteines and two out of five conserved histidines (Fig. 1A). Results of earlier alanine substitution experiments indicate which cysteine residues might coordinate the Zn ions [10]. Alanine substitutions of all cysteine residues except residue Cys101 in the GCM domain of mGCMa led to strongly reduced expression levels in transiently transfected COS cells, whereas alanine mutation of Cys101 only changed the redox sensitivity of the DNA-binding activity but did not affect its expression level. It is therefore likely that Cys101 is the only cysteine residue not directly involved in Zn binding. Among the histidine residues, only alanine substitution of residues His152 and His154 led to reduced GCMa expression, which makes both residues prime candidates for the Zn-ligating histidine residues.

Given the presence of Zn ions in the GCM domain, the resistance of its DNA-binding activity to Zn chelators can be either a consequence of the tight binding of the Zn ions by the protein or the absence of any role of the Zn ions in the DNA-binding activity (or even both). Our X-ray fluorescence and CD spectroscopy experiments demonstrate that the Zn ions are indeed tightly bound and that at room temperature phenanthroline is not able to remove the Zn ions from the GCM domain. Such tight binding of the Zn ions is also supported by the observation that both Zn positions are almost fully occupied in the bacterially expressed GCM domain despite the facts that no Zn ions were present during bacterial growth and purification and that large amounts of DTT were used during the last steps of the purification. Only under heat-denaturating conditions as demonstrated by our EMSA experiments the GCM domain becomes sensitive to metal chelators and treatment with 1,10-phenanthroline can abolish DNA binding (Fig. 4). This indicates that the Zn ions in the GCM domain are deeply buried in the protein core where they appear to play an essential, structural role in the native protein or during the folding process. The removal of the Zn ions impairs the structure of the GCM domain and therefore abolishes DNA binding. An important structural role of the Zn ions in the GCM domain is also supported by a recent survey: accordingly, in most DNA-binding domains Zn ions play structural roles where they are generally four-fold coordinated with cysteine and histidine residues as observed in the GCM domain. In Zn-containing enzymes Zn ions show more flexible coordination geometries with coordination numbers varying between four and six and with residues His, Asp and Glu as preferred Zn ligands [17]. The tight binding of Zn ions by the GCM domain differs from the situation in the Zn-finger transcription factor Krox-24 where treatment with phenanthroline at room temperature is sufficient to abolish DNA binding [10]. Similarly, in the tumor suppressor p53 a Zn ion in tetrahedral coordination ligated by three cysteines and one histidine can be removed by a Zn chelator. Metallothioneines can also modulate the transcriptional activity of p53 and it has therefore been suggested that Zn ions regulate the p53 DNA-binding affinity in vivo [19]. For the GCM domain the strong binding of the Zn ions and the drastic treatment required for their removal makes any regulatory role unlikely.

Our results add the GCM domain as an additional prototype of Zn-coordinating, sequence-specific DNA-binding domain. For five different classes structural information is available [20,21]. The classical Zn-finger proteins of the Cys_2His_2-type form the largest class of Zn-containing DNA-binding domains with more than 700 members being identified in the human genome. Other classes of Zn-containing DNA-binding domains are the Gal4-type transcription factors where six cysteines and two Zn ions form a bi-nuclear Zn_2Cys_6 cluster, the transcription factors of the nuclear receptor family with two tandem Zn-binding modules, the GATA transcription factors with a Zn-binding module resembling the N-terminal Zn-binding module of the nuclear receptor DNA-binding domain and the tumor suppressor p53. The different classes can be generally recognized by relatively simple Cys/His sequence signatures corresponding to the ability of the Zn ion to stabilize small, discrete protein domains except for p53 and the GCM domain where no such simple signatures can be derived (Fig. 1B). It has been argued that Zn is used to stabilize domains, which are not sufficiently large to form a stable hydrophobic core. However, this argument is probably less valid for p53 and the GCM domain where the DNA-binding domains comprise 200 and 150 residues, respectively. In p53 the Zn ion has acquired an additional regulatory role, whereas the exact role of the two Zn ions in the GCM domain has to await a detailed crystallographic analysis.

Acknowledgements: We thank members of the EMBL/ESRF Joint Structural Biology Group for support at different ESRF beamlines

and Elspeth F. Garman for the microPIXE analysis. M.W. acknowledges support by the Sonderforschungsbereich 473 of the DFG.

References

[1] Wegner, M. and Riethmacher, D. (2001) Trends Genet. 17, 286–290.
[2] Hosoya, T., Takizawa, K., Nitta, K. and Hotta, Y. (1995) Cell 82, 1025–1036.
[3] Jones, B.W., Fetter, R.D., Tear, G. and Goodman, C.S. (1995) Cell 82, 1013–1023.
[4] Vincent, S., Vonesch, J.L. and Giangrande, A. (1996) Development 122, 131–139.
[5] Schreiber, J., Riethmacher-Sonnenberg, E., Riethmacher, D., Tuerk, E.E., Enderich, J., Bosl, M.R. and Wegner, M. (2000) Mol. Cell. Biol. 20, 2466–2474.
[6] Anson-Cartwright, L., Dawson, K., Holmyard, D., Fisher, S.J., Lazzarini, R.A. and Cross, J.C. (2000) Nat. Genet. 25, 311–314.
[7] Gunther, T. et al. (2000) Nature 406, 199–203.
[8] Akiyama, Y., Hosoya, T., Poole, A.M. and Hotta, Y. (1996) Proc. Natl. Acad. Sci. USA 93, 14912–14916.
[9] Schreiber, J., Sock, E. and Wegner, M. (1997) Proc. Natl. Acad. Sci. USA 94, 4739–4744.
[10] Schreiber, J., Enderich, J. and Wegner, M. (1998) Nucleic Acids Res. 26, 2337–2343.
[11] Tuerk, E.E., Schreiber, J. and Wegner, M. (2000) J. Biol. Chem. 275, 4774–4782.
[12] Miller, A.A., Bernardoni, R. and Giangrande, A. (1998) EMBO J. 17, 6316–6326.
[13] Garvie, C.W. and Wolberger, C. (2001) Mol. Cell 8, 937–946.
[14] Pettifer, R.F. and Hermes, C. (1985) J. Appl. Cryst. 18, 404.
[15] Binsted, N., Strange, R.W. and Hasnain, S.S. (1992) Biochemistry 31, 12117–12125.
[16] Garman, E. (1999) Struct. Fold Des. 7, R291–R299.
[17] Alberts, I.L., Nadassy, K. and Wodak, S.J. (1998) Protein Sci. 7, 1700–1716.
[18] Auld, D.S. (2001) Biometals 10, 5227–5236.
[19] Meplan, C., Richard, M.J. and Hainaut, P. (2000) Oncogene 19, 5227–5236.
[20] Luscombe, N.M., Austin, S.E., Berman, H.M. and Thornton, J.M. (2000) Genome Biol., REVIEWS001.
[21] Berg, J.M. and Shi, Y. (1996) Science 271, 1081–1085.