*Journal of Biomolecular NMR*, **13:** 249–262, 1999.
*KLUWER/ESCOM*



# The solution structure of the N-terminal zinc finger of GATA-1 reveals a specific binding face for the transcriptional co-factor FOG

K. Kowalski, R. Czolij, G.F. King, M. Crossley & J.P. Mackay*
*Department of Biochemistry, University of Sydney, Sydney, NSW 2006, Australia*

Received 20 July 1998; Accepted 1 October 1998

*Key words:* GATA-1, protein-protein interactions, solution structure, zinc finger

**Abstract**

Zinc fingers (ZnFs) are generally regarded as DNA-binding motifs. However, a number of recent reports have implicated particular ZnFs in the mediation of protein-protein interactions. The N-terminal ZnF of GATA-1 (NF) is one such finger, having been shown to interact with a number of other proteins, including the recently discovered transcriptional co-factor FOG. Here we solve the three-dimensional structure of the NF in solution using multidimensional $^{1}H/^{15}N$ NMR spectroscopy, and we use $^{1}H/^{15}N$ spin relaxation measurements to investigate its backbone dynamics. The structure consists of two distorted β-hairpins and a single α-helix, and is similar to that of the C-terminal ZnF of chicken GATA-1. Comparisons of the NF structure with those of other $C_4$-type zinc binding motifs, including hormone receptor and LIM domains, also reveal substantial structural homology. Finally, we use the structure to map the spatial locations of NF residues shown by mutagenesis to be essential for FOG binding, and demonstrate that these residues all lie on a single face of the NF. Notably, this face is well removed from the putative DNA-binding face of the NF, an observation which is suggestive of simultaneous roles for the NF; that is, stabilisation of GATA-1 DNA complexes and recruitment of FOG to GATA-1-controlled promoter regions.

## Introduction

GATA-1 is the founding member of the GATA family of transcriptional activators, which includes GATA-2, -3, -4, -5, and -6 (Orkin, 1992; Weiss and Orkin, 1995), as well as related proteins in lower organisms, such as Pannier in Drosophila (Ramain et al., 1993) and AreA in yeast (Kudla et al., 1990). The family is characterized by the presence of two (in most cases) adjacent type IV zinc-finger domains (Harrison, 1991) which share a characteristic Cys-$X_2$-Cys-$X_{17}$-Cys-$X_2$-Cys topology, and which mediate high-affinity binding of the protein to (A/T)GATA(A/G) motifs in DNA. GATA-1 is predominantly expressed in erythroid cells, and is thought to play a role in most, if not all, genes expressed during red cell development (Orkin, 1992; Weiss and Orkin, 1995). Gene disruption experiments in mice have been used to demonstrate that GATA-1 is essential for normal erythroid development (Fujiwara et al., 1996; Pevny et al., 1991).

The two zinc-finger domains of GATA-1, termed the N-finger (NF) and the C-finger (CF) from their relative positions in the protein, share ∼50% sequence identity (Figure 1). This level of homology is not unusual within the GATA family; however, what is remarkable is that the two fingers appear to play quite different roles in the biological function of GATA-1. While it has been unequivocally shown, from a combination of biological and structural studies, that the CF domain is both necessary and sufficient for high-affinity, sequence-specific DNA binding (Martin and Orkin, 1990; Omichinski et al., 1993), the role of the NF has proven somewhat more elusive. While it is strictly required for terminal erythroid maturation (Weiss et al., 1997), the NF of GATA-1 is unable to bind DNA in isolation (Pedone et al., 1997). Despite this, recent work has shown that it increases both the stability (Martin and Orkin, 1990)

*To whom correspondence should be addressed. E-mail: j.mackay@biochem.usyd.edu.au.
Supplementary material available from the authors: $^{1}H$ and $^{15}N$ chemical shifts for NF200–243.

and specificity (Whyatt et al., 1993) of GATA-1:DNA complexes, and is capable of contacting DNA at double (A/T)GATA(A/G) motifs (Trainor et al., 1996; Yang and Evans, 1992), in conjunction with the C-finger. However, more significantly, several reports have implicated the NF in specific protein-protein interactions with other zinc-finger proteins, including Sp1 (Merika and Orkin, 1995), EKLF (Merika and Orkin, 1995), FOG (Tsang et al., 1997), and GATA-1 itself (Calligaris et al., 1995; Crossley et al., 1995; Yang and Evans, 1995; Elefanty et al., 1996; Haenlin et al., 1997). These reports, together with a number of other recent studies on unrelated zinc-finger proteins that appear to mediate protein-protein interactions (for a review, see Mackay and Crossley, 1998), suggest a second biological function for zinc finger domains. That is, in addition to their familiar role as DNA recognition motifs, zinc fingers may act as protein-protein recognition domains.

The interaction of the NF with ZnF-6 from FOG is of particular interest. FOG is a recently discovered erythroid protein (Tsang et al., 1997) containing nine zinc fingers, of which five contain an unusual CCHC arrangement of zinc-binding residues. FOG is coexpressed with GATA-1 during embryonic and hematopoietic development, and has been demonstrated both to synergistically activate transcription from a hematopoietic-specific regulatory region and to cooperate with GATA-1 in both erythroid and megakaryocytic differentiation (Tsang et al., 1997). Yeast two-hybrid, transient transfection, and GST-pulldown assays have all been used to confirm that the NF of GATA-1 and ZnF-6 from FOG are sufficient to mediate the interaction between the two proteins (Tsang et al., 1997; Fox et al., 1998), and that correct folding of the domains is crucial for the interaction to occur (Fox et al., 1998). This interaction therefore represents one of the most thoroughly characterized associations between two zinc finger domains, at a biochemical level, although a more detailed molecular picture of the interaction is still lacking.

Here we go part way to addressing this deficit; we describe the three-dimensional solution structure and backbone dynamics of the NF of murine GATA-1, and use this information to distinguish between the structural motifs required for zinc finger dimerisation and DNA-binding. We show that the backbone fold of NF is very similar to that of CF and conclude that the differences in function are due to specific contacts made by functional groups on the sidechains of the two fingers. A comparison of the sequences of CFs and NFs from a variety of species reveals that critical residues for mediating the interaction with FOG (Fox et al., 1998) are highly conserved in the NFs, while they are not conserved in the corresponding CFs. These residues map onto a single contiguous surface of the NF, a surface which is essentially non-overlapping with the putative DNA-binding face of the NF.

## Experimental procedures

*Expression and purification*

A construct encoding residues 200–243 of the N-terminal zinc finger of murine GATA-1 (NF200–243) was cloned into the *E. coli* expression vector pGEX-2T (Pharmacia), creating a C-terminal fusion with glutathione-*S*-transferase (GST), as previously described (Mackay et al., 1998). This construct was expressed in the host strain BL21 (DE3) grown in Luria broth. Cells were grown at 37 °C, and expression of NF200–243 was induced at an $OD_{600nm}$ of between 0.6 and 0.8 by the addition of isopropyl-β-D-thiogalactopyranoside (0.075 mM). After a further 3 h, cells were harvested by centrifugation, and the cell pellets stored at −80 °C prior to use.

Cell pellets were resuspended in lysis buffer containing Tris (50 mM, pH 8), EDTA (1 mM), NaCl (50 mM), phenylmethylsulfonylfluoride (PMSF, 0.5 mM), β-mercaptoethanol (2 mM), and lysozyme (1 mg $ml^{-1}$). The suspension was incubated on ice for 30 min, treated with Triton X-100 to 1%, and further incubated on ice for 30 min. Cells were lysed by sonication and centrifuged. The supernatant was loaded onto an *S*-hexylglutathione-agarose column pre-equilibrated in PBS (containing 1 mM EDTA, and 0.5 mM PMSF). The columns were washed, and the 46-residue NF200–243 peptide (containing an additional Gly-Ser at the N-terminus, arising from the thrombin recognition site) was cleaved from the column with thrombin (100 U per 1 l culture) at 37 °C for 1 h. The eluted peptide was purified to homogeneity by reverse phase HPLC, and its identity was confirmed using electrospray mass spectrometry ($M_{theor.} = 5191.0$ Da; $M_{obs.} = 5191.7$ Da).

For the preparation of uniformly $^{15}N$-labelled NF200–243, expression and purification was carried out as described above, except that cells were grown in minimal medium containing $^{15}N$-ammonium chloride (Isotec) as the sole nitrogen source. Uniform labelling was confirmed using electrospray mass spectrometry ($M_{theor.} = 5266.0$ Da; $M_{obs.} = 5265.2$ Da).

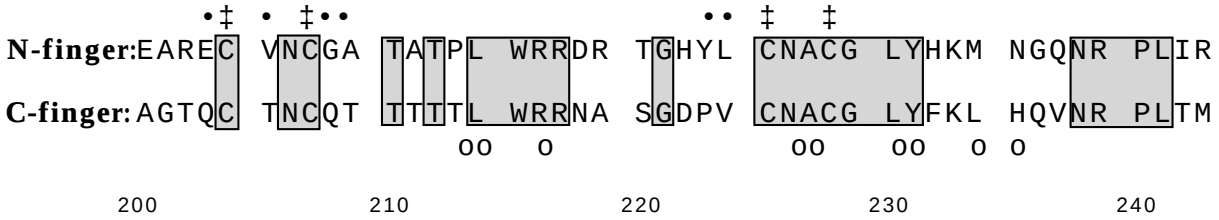


*Figure 1.* Comparison of the amino acid sequences of the N- and C-fingers of murine GATA-1. Sequence identities are boxed, and zinc-chelating residues are labelled with a ‡. N-finger residues implicated in the interaction with FOG are indicated with closed circles and C-finger residues that have shown to interact with DNA bases (Omichinski et al., 1993) are indicated with open circles.

*Sample preparation*

An unlabelled NF200-243 sample was prepared for NMR by dissolving ∼6 mg NF200-243 in $H_2O/D_2O$ (95:5, 280 μl) containing Tris-$d_{11}$ (10 mM), tris(2-carboxyethyl)phosphine (TCEP, 5 mM), and $ZnSO_4$ (5 mM), giving a ∼4 mM sample. The pH was adjusted to 5.4 using 0.1 M HCl and the sample was placed in a 5-mm-O.D., susceptibility-matched microcell (Shigemi). For experiments in $D_2O$, the sample was lyophilized and reconstituted in a microcell in 280 μl 99.96% $D_2O$ (Sigma). A $^{15}$N-labelled sample of NF200–243 was likewise prepared by dissolving ∼3 mg $^{15}$N-NF200–243 in $H_2O/D_2O$ (95:5, 280 μl) containing Tris-$d_{11}$ (20 mM), TCEP (3 mM), and $ZnSO_4$ (3 mM), yielding a ∼2 mM sample. The pH was adjusted to 5.5, and the sample was placed in a microcell.

Note that the addition of zinc to the HPLC-purified peptides required some care. It was found that the most reproducible strategy for avoiding aggregation was to add the zinc to a solution of the peptide buffered at pH 8 and containing an excess of TCEP, a reducing agent tailored for the stoichiometric reduction of disulfide bonds. Following the addition of a slight excess of $Zn^{2+}$, the pH was lowered gradually to 5.4–5.5. It was found that at higher pH values than this, substantial broadening was observed throughout the $^1$H spectrum, while at pH values lower than ∼4.5, the peptide reversibly unfolded. Prior to commencing NMR studies, the monomeric nature of GN200–243 was established using sedimentation equilibrium experiments (Mackay et al., 1998).

*NMR spectroscopy*

All NMR experiments used for determination of the structure of GN200–243 were carried out at 288 K on a 600-MHz Bruker AMX-600 spectrometer, equipped with a 5-mm broadband reverse probe and three-axis pulsed field gradients. Water suppression was achieved either using on-resonance presaturation during the relaxation delay between scans and during the mixing period in NOESY (Kumar et al., 1980) experiments or using pulsed-field gradients. When presaturation was used, pulse sequences were modified by addition of a SCUBA (Brown et al., 1988) sequence prior to the first 90° pulse to facilitate the recovery of bleached Hα resonances. The following homonuclear 2D spectra were recorded on the unlabelled sample: TOCSY (Bax and Davis, 1985) with DIPSI-2 mixing (Shaka et al., 1988; $t_m$ = 87 ms); NOESY ($t_m$ = 80, 150, and 200 ms); DQFCOSY (Piantini et al., 1982); and E-COSY (Griesinger et al., 1985). Using the $^{15}$N-labelled sample, HSQC (Bax et al., 1990; Norwood et al., 1990), 2D HSQC-TOCSY (Cavanagh et al., 1991; Norwood et al., 1990), J-modulated HSQC (Billeter et al., 1992), and 3D NOESY-HSQC (Fesik and Zuiderweg, 1988; Marion et al., 1989) ($t_m$ = 200 ms) experiments were recorded.

For the measurement of $^{15}$N relaxation parameters, $^{15}$N $T_1$ and $T_2$ experiments were recorded on a Bruker DRX-500 spectrometer, equipped with a 5-mm triple resonance probe and z-axis gradients. The spectra were recorded using pulse sequences from the Bruker library, and incorporated pulsed field gradients, a flip-back sequence for water suppression (Grzesiek and Bax, 1993), and the sensitivity improvement method of Palmer et al. (1991). The strength of the $^{15}$N spin-locking field used in the $T_2$ ($T_{1\rho}$) experiment was 1.7 kHz. The $T_1$ spectra were recorded with τ delays of 50, 100, 150, 200, 300, 400, 500, 600, 700, 800, 100, and 1500 ms, while the $T_2$ experiments used values of 18, 36, 54, 90, 144, 198, 252, 324, 414, 522, 612, 720, and 810 ms. All spectra were acquired with spectral widths of 1720 and 6010 Hz in $F_1$ and $F_2$, using acquisition times of 37 and 170 ms in $t_1$ and $t_2$, respectively. The total recycle delay between scans was set to 3.9 s for both the $T_1$ and the $T_2$ experiment.

*Data processing*

Spectra were processed using XWINNMR (Bruker) and 3D experiments were analysed using AURELIA (Bruker), both running on a Silicon Graphics $O_2$ workstation. The $^1$H frequency scale of all spectra was directly referenced to $d_4$-TSP at 0.00 ppm, while the $^{15}$N frequency scale was indirectly referenced to liquid $NH_3$ using the method described by Live et al. (1984). Two- and three-dimensional NMR spectra were processed using Lorentzian-Gaussian window functions in the directly detected dimension and shifted squared sine bell functions in the indirectly detected dimension(s), and were zero-filled and linear predicted (in the case of $^{15}$N dimensions) before Fourier transformation. Polynomial baseline corrections were applied to the processed spectra in $F_2$.

*Restraint generation*

Interproton distance restraints were derived from the intensities of crosspeaks in the NOESY spectra. Analysis of these spectra yielded 520 non-redundant interproton distance restraints which were assigned upper-distance bounds of 2.8 Å (strong), 3.5 Å (medium), or 5.0 Å (weak) based on the corresponding crosspeak intensity. Pseudoatom corrections were added to distance restraints where necessary (Wüthrich et al., 1983), and an empirical correction of 0.5 Å was added to the upper bound for restraints involving methyl groups (Clore et al., 1991).

Twenty $\phi$ dihedral angle restraints were derived from $^3J_{\mathrm{NH}\alpha}$ coupling constants measured from the $J$-modulated HSQC experiment. The $\phi$ angle was restrained to $-35 \pm 15°$ for $^3J_{\mathrm{NH}\alpha} < 3$ Hz, $-65 \pm 15°$ for $3 < {}^3J_{\mathrm{NH}\alpha} < 5.8$ Hz, $-120 \pm 30°$ for $8 < {}^3J_{\mathrm{NH}\alpha} < 9.5$ Hz, and $-120 \pm 15°$ for $^3J_{\mathrm{NH}\alpha} > 9.5$ Hz. A restraint of $-100 \pm 80°$ was applied for 11 additional residues in which the intraresidue Hα-HN NOE was clearly weaker than the NOE between HN and the Hα of the preceding residue (Clubb et al., 1994).

Using $^3J_{\alpha\beta2}$ and $^3J_{\alpha\beta3}$ coupling constants measured from E-COSY spectra, in combination with HN-Hβ2, HN-Hβ3, Hα-Hβ2, and Hα-Hβ3 NOE intensities and the results of preliminary structure calculations, a total of 31 $\chi_1$ dihedral angle constraints (including 8 stereospecific assignments of methylene protons) were derived. $\chi_1$ angles were constrained to a range of $\pm 30°$ about a central value of either 60°, −60°, or 180°, or to a range of 180° if only one of the three possible rotamers could be excluded. The β methylene protons of the two Pro residues were stereospecifically assigned from NOESY spectra on the basis that the

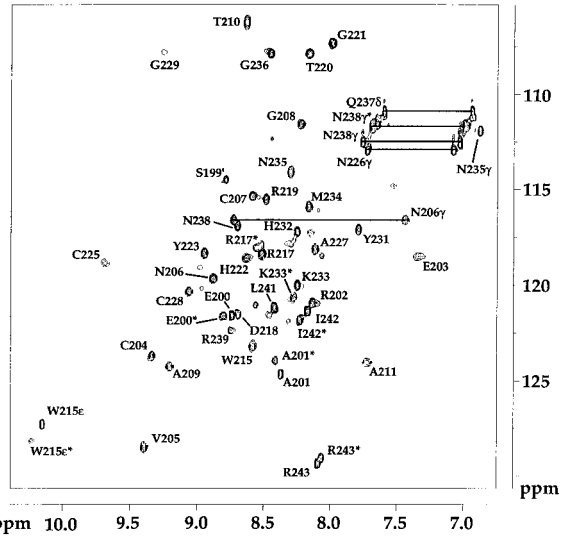


*Figure 2.* $^1$H/$^{15}$N-HSQC spectrum of uniformly $^{15}$N-labelled NF200–243 (conditions as stated in the Materials and methods). Assignments are indicated beside each peak. Signals arising from glutamine and asparagine sidechains are connected by a horizontal line, except for N235δ, for which a second signal could not be located. Assignments followed by an asterisk correspond to a minor species (see text), and S199′ is non-native. Several signals of lower intensity could not be assigned, and presumably arise from a minor species (see text).

Hα is always closer to Hβ3 than Hβ2 (Kline et al., 1988). Both X-Pro bonds were clearly identified as being in the trans conformation ($\omega \sim 180°$) on the basis of strong NOEs between the Hδ of each Pro residue and the Hα and HN protons of the preceding residues (Wüthrich, 1986). No hydrogen bond restraints were used in the structure calculations.

*Structure calculations*

Structures were calculated using a combination of distance geometry and simulated annealing. The distance geometry program DIANA was used to calculate 1000 structures from random starting conformations. DIANA calculations incorporated 2 cycles using redundant dihedral angle constraints (REDAC) in order to reduce computational time required for obtaining a family of acceptable conformers (Güntert and Wüthrich, 1991). Distance geometry calculations were performed in the absence of $Zn^{2+}$ but an approximately tetrahedral arrangement of the four Cys Sγ atoms was maintained by adding restraints of $3.55 \text{ Å} \leq d \leq 3.95$ Å for $S\gamma_i$-$S\gamma_j$.

The best 100 structures (selected on the basis of their final target function values) were refined in the program X-PLOR using a modified version of the dy-

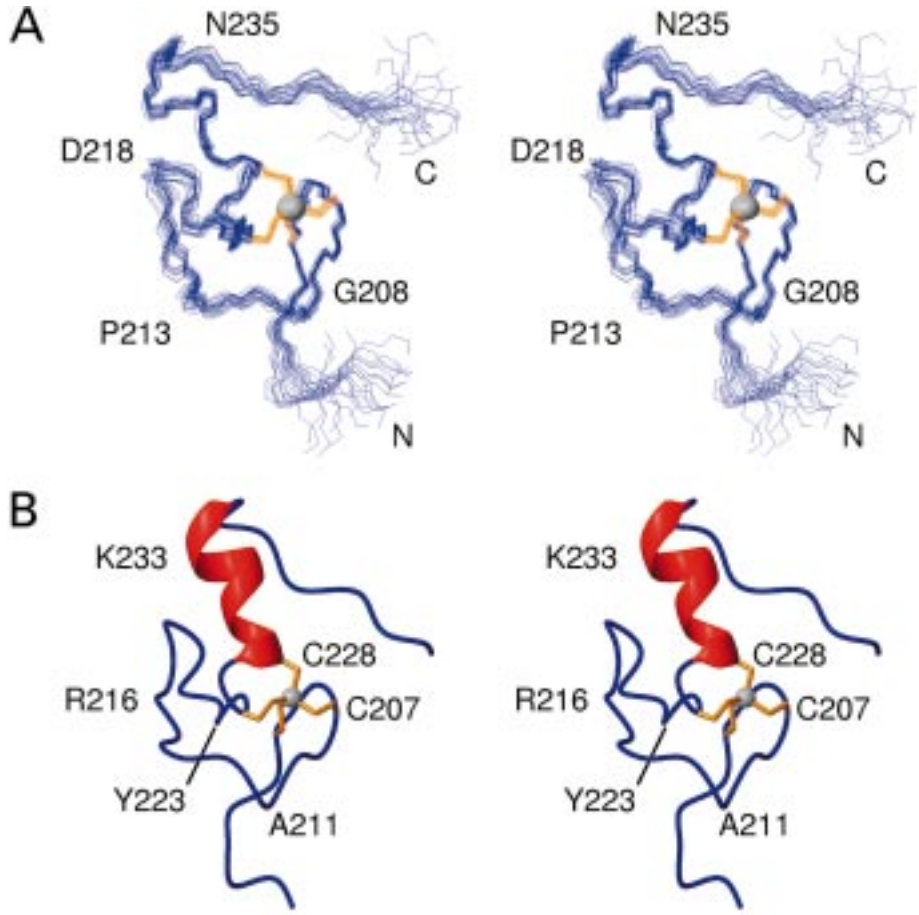


*Figure 3.* (A) Stereo view of the ensemble of 25 conformers used to represent the solution structure of GN200–243. Structures are superimposed for best fit over the backbone (Cα, C, N) atoms of residues 201–241, relative to the best structure. Cysteine sidechain heavy atoms are shown in orange and the zinc atoms are shown in grey. (B) Ribbon diagram (stereo view) of GN200–243, displaying the secondary structure recognised by the program PROMOTIF. The α-helix is shown in red, the sidechains of the cysteine residues in yellow, and the zinc atom in grey. Note that the two β-hairpins are not recognized due to the lack of regular hydrogen bonding patterns.

namical simulated-annealing protocol (Nilges et al., 1988). The $Zn^{2+}$ atom was incorporated into the calculations by introducing covalent restraints to maintain tetrahedral geometry with $S\gamma_i$-Zn bond lengths of 2.3 Å, and bond angles of 109° for the $S\gamma_i$-Zn-$S\gamma_j$ angle and 107° for the $C\beta_i$-$S_i$-Zn angle. Calculations were carried out in the simplified all-hydrogen force field. Interatomic distances and dihedral angles were constrained by experimental energy terms ($k_{noe}$ = 50 kcal mol$^{-1}$ Å$^{-2}$, $k_{dihed}$ = 200 kcal mol$^{-1}$ rad$^{-2}$) which remained constant throughout the calculations. Covalent geometry was constrained by standard X-PLOR parameters (Brünger, 1992). Initial atomic velocities were chosen from a Maxwellian distribution at 1000 K, with non-bonded interactions being modelled by a weak 'repel' function that ignores electrostatic interactions and allows atoms to pass through each other (repel = 0.9, $C_{rep}$ = 0.003 kcal mol$^{-1}$ Å$^{-4}$). The system was then cooled to 100 K over 10 ps, during which time $C_{rep}$ was increased linearly to 4 kcal mol$^{-1}$Å$^{-4}$, while repel was decreased to 0.75. The final stage consisted of 200 cycles of restrained conjugate minimization. The 25 conformers with the lowest value of $E_{tot}$ were taken to represent the final structure. The structures were visualised and analysed using the programs MOLMOL (Koradi et al., 1996), PROCHECK (Laskowski et al., 1993), and PROMOTIF (Hutchinson and Thornton, 1996).

*Analysis of $^{15}N$ relaxation data*

Crosspeak volumes were extracted from the relaxation spectra using the program XEASY (Bartels et al.,

1995), and used to determine relaxation parameters for each resolved crosspeak, as previously described (Mackay et al., 1996). Inertial moment tensor analysis suggested that the protein was sufficiently globular to be well approximated by a sphere (principal moments 1.35:1.2:1). The data was analysed using the simplified method of Habazettl and Wagner (1995).

## Results and discussion

*NMR resonance assignments*

$^{1}$H resonance assignments for GN200–243 were obtained using the standard homonuclear sequential assignment methodology (Wüthrich, 1986), supplemented with information from the $^{15}$N-separated experiments. It was noticed early on in the assignment process that a second series of crosspeaks could be observed for several residues, with intensities of ∼10–25% of the main peaks. This is seen clearly in the $^{15}$N-$^{1}$H HSQC spectrum (Figure 2), where peaks labelled with an asterisk correspond to the minor species. With the exception of the W215 sidechain amide proton, this phenomenon was basically restricted to residues near the two termini of the protein, viz. S199′, E200, A201, L241, I242, and R243 (note that the non-native Gly–Ser sequence at the N-terminus is left behind after the thrombin cleavage, and is numbered as G198′–S199′). In two cases (E200 and I242), an even smaller third set of signals (∼5% of main peaks) could be observed. These minor signals were much less prominent in the NOESY spectra than the scalar correlated spectra (e.g., TOCSY, $^{15}$N-$^{1}$H HSQC), and did not give rise to any observable non-sequential NOEs. Given that the mass spectrum of the purified GN200–243 peptide did not indicate the presence of minor species, it was concluded that the signals corresponded either to misfolded forms of the protein, or to chemical exchange conformers in which one or both of the termini exist in an alternative conformation. In the latter case, the rate of interconversion between these conformers would be $< \sim 100\ s^{-1}$, taking account of the chemical shift differences observed. It is notable that similar observations have previously been reported for zinc-binding peptides (Hammarström et al., 1996), and the cause of the heterogeneity has never been resolved.

In addition, the resonances of several residues within the sequence were significantly broadened, particularly those of L214, R216, and N226. Again, this is most likely due to a conformational exchange process; however, because of the low signal-to-noise observed for these residues in all spectra collected, the nature of the conformational exchange could not be identified. This and the peak doubling noted above are discussed further below. In total, 93% of backbone $^{1}$H/$^{15}$N and 95% of sidechain assignments were obtained.

*Structure determination*

The final structure calculations were based on a total of 520 non-redundant interproton distance restraints derived from 2D and 3D ($^{15}$N-separated) NOESY spectra and 31 $\phi$ and 31 $\chi_1$ dihedral angle constraints derived from coupling constant and NOE measurements. Thus, the total number of experimentally determined restraints was 582, representing an average of 12.6 restraints per residue.

The geometry of the zinc binding site was defined using standard interatomic distances and angles (see Experimental procedures section); the tetrahedral nature of the coordination site, as well as confirmation that the zinc atom is ligated by four cysteine (and no histidine) residues, was obtained previously from UV-visible spectrophotometry studies (Mackay et al., 1998).

The structures were calculated using a hybrid distance geometry simulated annealing protocol (Nilges et al., 1988) in the programs DIANA (Güntert et al., 1988) and X-PLOR (Brünger, 1992). A family of 25 structures with the lowest residual restraint violations was used to represent the solution structure of GN200–243 (Figure 3A). The structures display good covalent geometry, judging from the small deviations from ideal bond lengths and angles, and good non-bonded contacts, as shown by the low value of the mean Lennard-Jones potential (Table 1). There are no violations of distance or angle constraints greater than 0.23 Å and 2.4°, respectively. In addition, a PROCHECK (Laskowski et al., 1993) analysis shows that, for the residues which exhibit $\phi$ angle order parameters $\geq 0.6$ (i.e., excluding S199′, L214, R216, and L241), over 99% of non-glycine/non-proline residues fall into the most favoured or additionally allowed regions of the Ramachandran plot. Excluding the terminal residues G198′–E200, I242, and R243, the atomic root mean square differences (RMSDs) for the final 25 structures with respect to the mean coordinate positions are $0.56 \pm 0.17$ Å for the backbone atoms (N, C$\alpha$, and C′) and $0.95 \pm 0.11$ Å for all heavy atoms. These figures, together with the number of restraints used to generate the structures and the good definition of many of the amino acid sidechains, correspond to

*Table 1.* Structural statistics for the family of 25 GN200–243 structures

| | |
|---|---|
| **Distance restraints** | |
| Intraresidue ($i - j = 0$) | 125 |
| Sequential ($\|i - j\| = 0$) | 153 |
| Medium range ($\|i - j\| < 5$) | 77 |
| Long range ($\|i - j\| > 5$) | 165 |
| Total | 520 |
| **Dihedral angle restraints** | |
| $\phi$ | 31 |
| $\chi_1$ | 31 |
| Total | 62 |
| **Mean RMSDs from experimental restraints** | |
| NOE (Å) | $0.0180 \pm 0.0007$ |
| Dihedral angles (°) | $0.21 \pm 0.05$ |
| **Mean RMSDs from idealized covalent geometry**[a] | |
| Bonds (Å) | $0.002 \pm 0.00005$ |
| Angles (°) | $0.505 \pm 0.004$ |
| Impropers (°) | $0.369 \pm 0.006$ |
| **Restraint violations** | |
| Mean NOE violations/structure $> 0.1$ Å | $7.1 \pm 1.2$ |
| Maximum NOE violation (Å) | 0.23 |
| Mean angle violations / structure $> 0.5°$ | $2.9 \pm 1.5$ |
| Maximum angle violation (°) | 2.4 |
| **Mean energies (kJ mol$^{-1}$)** | |
| $E_{\mathrm{NOE}}$ | $12.4 \pm 0.9$ |
| $E_{\mathrm{cdih}}$ | $0.16 \pm 0.11$ |
| $E_{\mathrm{vdW}}$ | $2.5 \pm 0.4$ |
| $E_{\mathrm{bond}}$ | $2.85 \pm 0.14$ |
| $E_{\mathrm{improper}}$ | $8.01 \pm 0.27$ |
| $E_{\mathrm{angle}}$ | $50.1 \pm 0.7$ |
| $E_{\mathrm{total}}$ | $76.1 \pm 1.2$ |
| **Atomic RMS differences versus mean (Å)**[b] | |
| Backbone atoms (4–44) | $0.56 \pm 0.17$ |
| Heavy atoms (4–44) | $0.95 \pm 0.11$ |

[a] Idealised geometry is defined by the CHARMM force field as implemented in XPLOR.
[b] Atomic differences are given as the average RMS against the mean coordinate structure. All energies, violations and RMSDs are given as the mean $\pm$ standard deviation.

a third-generation NMR structure, according to the convention outlined by Clore and Gronenborn (1991).

*Description of the structure*

The structure of GN200–243 (Figure 3A, B) essentially comprises two highly distorted β-hairpins followed by an α-helix. The first hairpin is formed by residues R202–T210, and displays backbone hydrogen bonds between the NH proton of G208 and the carbonyl oxygen of C204, and between C204-NH and A209-CO (Figure 4A), resulting in a twisting of the hairpin (G208-CO would accept this H-bond in a typical β-hairpin). In addition, there are three hydrogen bonds to the Sγ atom of C204: from the backbone amide protons of N206, C207, and C225. All of these hydrogen bonds, as well as the values of the backbone dihedral angles for residues C204–C207 (Pérez-Alvarado et al., 1996), are characteristic of the 'rubredoxin knuckle' (Figure 4B), a motif named af-

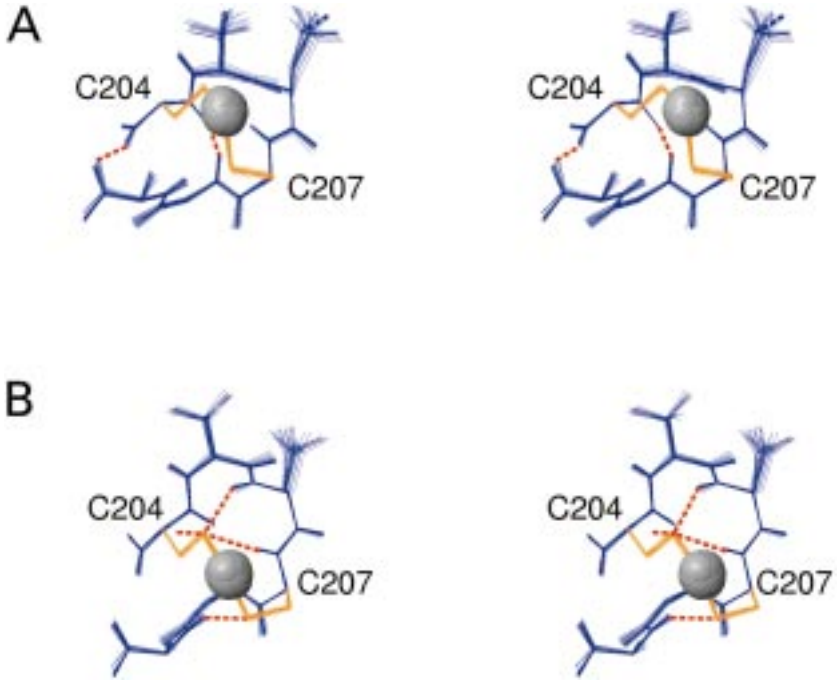


*Figure 4.* The rubredoxin knuckle of GN200–243. (A) Stereo view showing the backbone hydrogen bonds (in red) which stabilize the first β-hairpin of GN200–243. (B) Stereo view showing the hydrogen bonds (in red) made to Sγ atoms. The donor atom for the hydrogen bond between Sγ of C7 and HN of C28 is not shown.

ter the iron-binding domains of rubredoxins (Adman et al., 1975), and which is common in zinc-binding domains (Schwabe and Klug, 1994). In addition, a fourth hydrogen bond involving a sidechain Sγ atom is found between C207 and the NH proton of A209. Hairpin 1 exhibits the lowest RMSDs of any region within the structure (0.10 Å for backbone atoms in the region R202–T210).

Residues T212–W215, which are recognized as either a type IV or a type VIII turn by the program PROMOTIF (Hutchinson and Thornton, 1996; depending on the particular member of the ensemble), then lead to the second hairpin (residues R216–Y223), which contains a classical type I β-turn formed by residues D218–G221. Three of the four residues in this turn (D218, T220, and G221) display very high positional turn potentials (Hutchinson and Thornton, 1994), and the sidechain carboxyl group of D218 forms hydrogen bonds with the backbone NH protons of T220, H222, and D218 itself. Aside from the hydrogen bonds in the turn region though, no backbone hydrogen bonds could be discerned for this hairpin. In part, this is due to the broadness of the NMR signals from several residues in this region (T212, L214, and R216, in particular), an observation which is indicative of the presence of some conformational exchange process. However, there are numerous cross-strand NOEs (between such pairs of residues as T212 and C225, and P213 and N226) which serve to juxtapose the two strands of the hairpin.

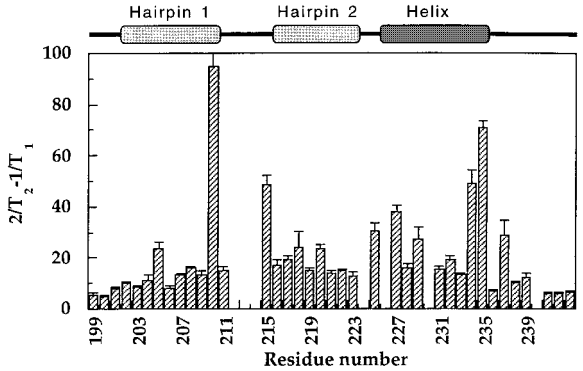


*Figure 5.* Summary of the $^{15}N$ relaxation data analysis for GN200–243. $(2/T_2 - 1/T_1)$ is plotted against residue number. Error bars represent the standard error in the value of $(2/T_2 - 1/T_1)$. Shown above the graph is an outline of the secondary structural elements of NF.

Residues N226 to N235 form an α-helix, although the N-terminal residue of the helix is ill-defined (ranging from N226 to C228 in the ensemble of structures).

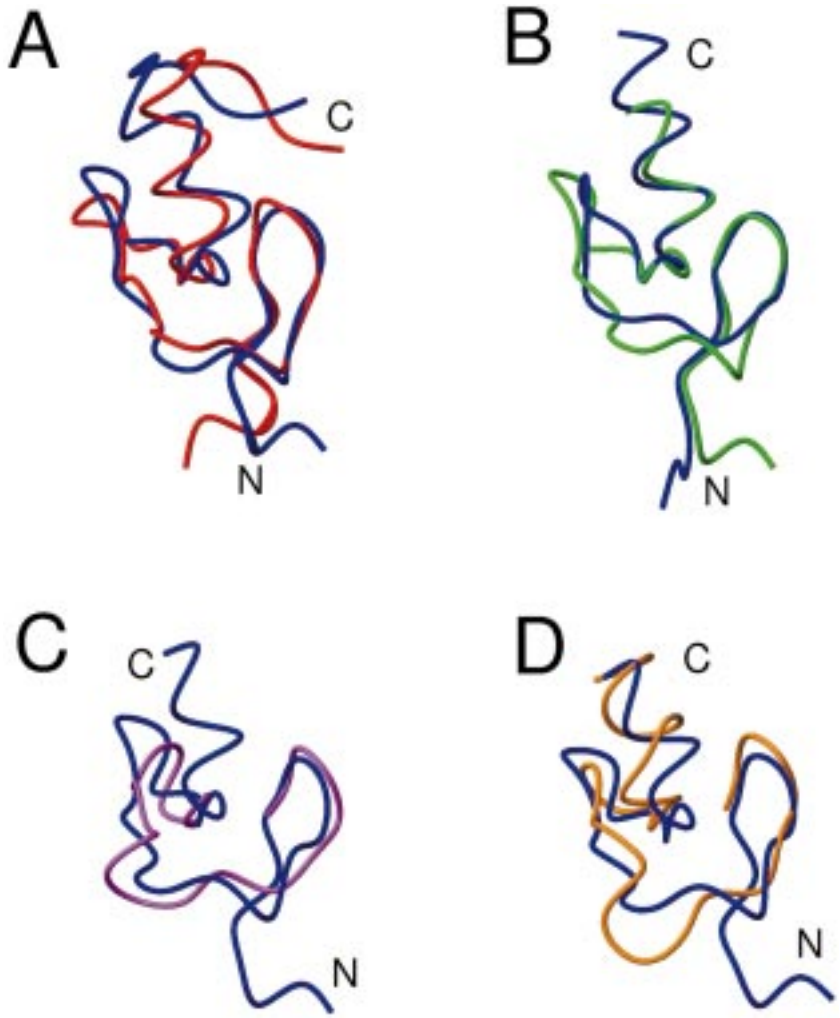


*Figure 6.* Comparison of the structures of the N-finger (blue) and several other zinc-binding domains with homologous structures. The proteins are shown as Cα traces. (A) Comparison with the C-finger (red, 1gat) of GATA-1. Residues 202–235 (Cα, C, and N atoms) superimpose with an RMSD of 1.8 Å. (B) Comparison with the N-terminal zinc-binding module of the estrogen receptor (green) (Schwabe et al., 1993). The two structures are shown superimposed over residues 203–209/225–231 (NF) and 6–12/24–30 (ER) (Cα, C, N atoms), giving an RMSD of 0.71 Å. Comparison with (C) the N- (magenta) and (D) C-terminal (orange) zinc-binding units of CRIP (Pérez-Alvarado et al., 1996). The regions used for superimposition are: 3–23 (N-terminal unit of CRIP) and 204–224 (GN200–243), and 30–60 (C-terminal unit of CRIP ) and 204–234 (GN200–243), and their RMSDs are 2.7 Å and 3.1 Å, respectively.

The pitch and number of residues per turn are 5.5 ± 0.5 Å and 3.73 ± 0.17, respectively, for the ensemble. Although these values are normal within the errors limits given, and the region is classified by PROMOTIF as an α-helix, the hydrogen bonding pattern is somewhat non-classical at several positions, being intermediate between α- and $3_{10}$-helix. Thus, $(i, i + 3)$ backbone hydrogen bonds are often observed in the same structures as $(i, i + 4)$ hydrogen bonds. This could partly be caused by a dearth of key medium-range NOEs characteristic of α-helix (e.g., $d_{\alpha N}(i, i + 4)$), due to spectral overlap. Transitions from α- to $3_{10}$-helix, however, have been observed previously for zinc finger domains (see, for example, Wuttke et al. (1997) and references therein). Asparagine, the classical N-terminal helix-capping residue, is found at the first position of the helix (N226), although its sidechain amide group does not appear to form a hydrogen bond with any backbone NH protons in the first turn of the helix. This could also be a consequence of a lack of experimental constraints for this residue, due to the low intensity of its backbone amide proton in the NMR

spectra. N226 is also found to adopt a positive $\phi$ angle in all of the structures; asparagine is the only residue other than glycine for which such angles are commonly observed (Fasman, 1989). It should be noted, however, that the density of restraints in this region of the structure is not high, and the positive $\phi$ angle could therefore possibly be an artefact of this paucity of restraints. This is commonly observed in NMR structure calculations for residues which are poorly defined by experimental restraints.

A glycine residue (G236), also with a positive $\phi$ angle, breaks the helix at the C-terminus and causes the chain to change direction, as is observed in around one third of all $\alpha$-helices (Schellman, 1980). It is also noticeable that a glycine residue is observed in the interior of the helix, namely G229. The remaining residues (Q237– R243) are essentially in an extended conformation, and form some contacts with residues in the helix and hairpin 1.

The zinc coordination sphere consists of the S$\gamma$ atoms of C204, C207, C225, and C228, and it is clear that the metal coordination contributes substantially towards the stability of the folded domain, as well as positioning the elements of secondary structure relative to one another. In addition, a number of hydrophobic clusters are observed. The single tryptophan residue (W215) is involved in interactions with the sidechains of C204, T212, and Y223, while Y231 forms the centre of a second cluster, involving residues N235, Q237, R239, and P240. A cluster comprising the sidechains of V205, H222, and L224 helps to establish the orientation of the second hairpin relative to the helix, and finally, interactions between L241 and N206/C207 provide some directionality to the final extended portion of the chain.

The measured $^{15}$N $T_1$ and $T_2$ data were interpreted using a simplified approach in which the parameter $(2/T_2 - 1/T_1)$ is calculated from the raw relaxation data (Habazettl and Wagner, 1995); a plot of $(2/T_2 - 1/T_1)$ versus residue number is shown in Figure 5. This parameter may be used as a guide to the relative order parameters for each residue, and, because it can be expressed in terms of only $J(0)$ (to a reasonable approximation), it is particularly sensitive to slow internal motions (on the millisecond to microsecond timescale). Thus, the elevated values of $(2/T_2 - 1/T_1)$ observed for residues A210 and W215 are consistent with the presence of a slow conformational exchange process, as proposed above. Elevated values are also observed for residues at either end of the $\alpha$-helix, in keeping with the large linewidth of N226 in homonuclear spectra and the presence of hydrogen bonding patterns intermediate between $\alpha$- and $3_{10}$-helix (especially at the C-terminal end of the helix).

*Comparison with the GATA-1 C-finger*

The structure of the C-finger of chicken GATA-1 bound to a 16-base-pair double-stranded oligonucleotide containing a single GATA site has previously been determined using NMR spectroscopy (Omichinski et al., 1993). An overlay of the backbone of the corresponding portion of this structure with that of the N-finger determined in the current study is shown in Figure 6A. The RMSD for the overlay (over the C$\alpha$, C, and N atoms of residues 202–235 of the N-finger) is 1.8 Å; clearly the two domains adopt the same fold, with the only substantial difference being the relative orientations of the second hairpin and the helix. The angle between these two elements of secondary structure is somewhat smaller in the N-finger structure, as evidenced by the NOEs observed between D218 and L224. The resolutions of the two structures are also comparable; RMSDs for the C-finger domain are 0.51 Å and 1.11 Å over backbone and all heavy atoms, respectively (residues 201–241, using the N-finger numbering). It is notable that the elements of secondary structure, especially the $\alpha$-helix, are somewhat more regular in the C-finger structure. This is likely to be a consequence of the presence of the DNA, which interacts directly with the helical region of the C-finger, thereby probably stabilizing the protein structure to some extent. The region corresponding to T212–R217 (N-finger numbering) is also more ordered in the C-finger structure; all of these residues are seen to directly contact DNA in the solution structure.

*Homology with other structures*

The structures of a number of other proteins containing CCCC-type zinc binding domains have been elucidated. These fall into two classes: the DNA-binding hormone receptor proteins and the so-called LIM domains, whose function at a molecular level is somewhat unclear. Members of both classes of proteins contain two independent zinc-binding modules, which pack against each other, forming extensive contacts. Three structures of hormone-receptor proteins bound to their cognate DNA sequence have been solved (Luisi et al., 1991; Schwabe et al., 1993; Rastinejad et al., 1995), and these domains share a high level of structural homology with each other (overlays of backbone atoms have RMSDs of $< 1$ Å). The N-terminal of

the two zinc-binding modules in each structure bears some resemblance to NF, although there are several gaps in the sequence alignment (the hormone receptor topology in this region is $CX_2CX_{13}CX_2C$, while the N-finger is $CX_2CX_{17}CX_2C$), which have a substantial effect on the orientation of the second hairpin and the helix. Despite this, Figure 6B shows that the overall fold is very similar, with an RMSD between NF and the homologous region from the estrogen receptor (Schwabe et al., 1993) of 0.71 Å (for backbone atoms of residues 203–209 and 225–231 in the N-finger). The C-terminal zinc binding module, however, is somewhat different, with more widely spaced Cys ligands and only a single distorted hairpin. A comparison of the CF structure with the glucocorticoid receptor (Omichinski et al., 1993) revealed a comparable degree of structural homology.

In contrast, the four LIM domain structures which have been solved (Hammarström et al., 1996; Konrat et al., 1997; Pérez-Alvarado et al., 1994, 1996) all share essentially the same Cys-to-Cys spacing as the N-finger, although there is almost no other sequence homology at all. Overlays of both the N- and C-terminal zinc-binding modules of CRIP (Pérez-Alvarado et al., 1996) with NF are shown in Figure 6C and 6D, respectively (RMSDs for the displayed regions are 2.7 Å and 3.1 Å, respectively). Note that the N-terminal module of LIM domains lacks a helical sequence, but otherwise, both modules shows a strong structural homology with NF.

*The GATA-1:FOG interaction*

The lack of independent DNA-binding ability in the NF of GATA-1 (in contrast to the NFs of GATA-2 and -3; Pedone et al. (1997)) is attributable not to sequence differences between the three fingers (in fact there are very few sequence differences between the N-finger regions of GATA-1, -2, and -3), but rather to GATA-1's lack of an N-terminal basic region, which borders the N-fingers of both GATA-2 and GATA-3. Because of the lack of independent DNA-binding capacity, the precise role of the GATA-1 N-finger has remained somewhat ambiguous. Significantly, however, this finger has recently been implicated in mediating protein-protein interactions with other finger domains, including finger 6 from the transcriptional co-factor FOG (Tsang et al., 1997). Through mutagenesis experiments, a number of residues in the N-finger have been implicated in mediating this interaction (Fox et al., 1998), and it is noticeable that all of these residues are conserved across species, and are not present in C-finger sequences. This is consistent with the observation that the C-finger is unable to bind FOG (Tsang et al., 1997). In addition, these residues are conserved in the N-fingers of GATA-2 and -3, which are also capable of interacting with FOG (Tsang et al., 1997), demonstrating that their presence is critical for formation of the GATA-FOG complex. It is also notable that GATA-4 and -5 share these same conserved residues, suggesting that they too may bind FOG or a FOG homologue in the same way.

Figure 7A shows the residues implicated in FOG binding (E203, V205, G208, A209, H222, and Y223) mapped onto the structure of the N-finger, and it can clearly be seen that they form a single contiguous surface, which presumably forms specific contacts with FOG finger 6. Cys204, which clearly could not be mutated without complete disruption of the structure, is highlighted in orange, and may also be involved in contacting FOG. The C-finger structure is also shown in the same orientation (Figure 7B), with the residues involved in contacting the DNA highlighted in cyan. It is notable that the DNA-binding surface of the C-finger and the FOG-binding surface of the N-finger are essentially on opposite faces. Thus, it would appear that NFs may be able to simultaneously bind both DNA and FOG. Given the likelihood that FOG is also capable of binding DNA through one or more of its eight other zinc finger domains, this raises the possibility that the GATA-FOG interaction could be involved in the establishment of active chromatin conformations, thereby priming the associated loci for transcription.

*Finger-finger interfaces*

Given that a single face on the NF which is critical for the interaction with FOG finger 6 has been delineated, and shown to be essentially non-overlapping with the putative DNA-binding face, it is of interest to compare these molecular recognition surfaces with analogous surfaces from related structures. The nuclear hormone receptor proteins contain both protein-DNA and protein-protein interfaces. The primary interaction with DNA occurs via the helix in the first zinc-binding module, in a similar fashion to the CF-DNA interaction. Contacts between the N- and C-terminal zinc binding units, however, appear to largely be mediated by hydrophobic interactions between the two helices. This contrasts with the NF-FOG interaction, for which much of the helix in the NF is dispensable (Fox et al., 1998), and the remaining residues are expected to contact DNA. Further, two of the six residues which

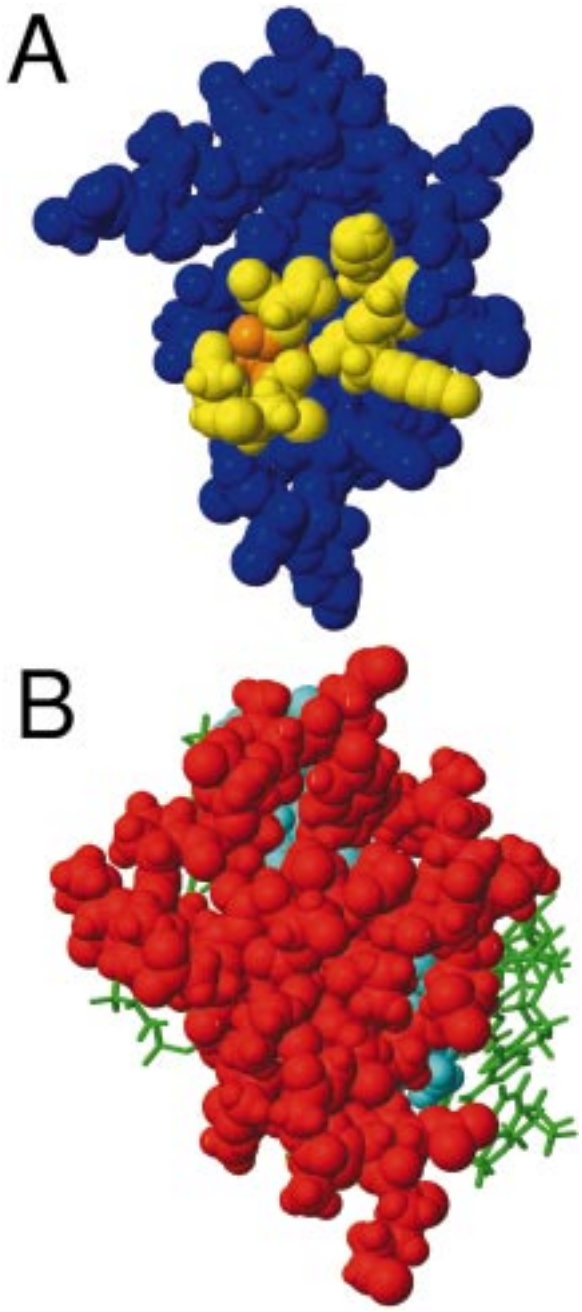


*Figure 7.* Functional surfaces of the C- and N-fingers of GATA-1. (A) CPK representation of the structure of GN200–243. The protein is rotated ca 180° about the vertical axis in the page relative to Figure 3. The residues which are implicated by mutagenesis experiments (Fox et al., 1998) are shown in yellow, and C204 is shown in orange. (B) CPK representation of the structure of the C-finger of chicken GATA-1 (red) bound to DNA (green) (Omichinski et al., 1993), shown in the same orientation as GN200–243 in (A). The residues which make direct contact with DNA bases are depicted in cyan.

have been shown to mediate the NF-FOG interaction are charged (viz., E203 and H222). The interface between the two hormone receptor monomers (Luisi et al., 1991; Rastinejad et al., 1995; Schwabe et al., 1993) likewise cannot be easily related to the NF–FOG situation. The hormone receptor zinc-finger domains do not dimerize in the absence of DNA (see, for example, Schwabe et al. (1990) and Hard et al. (1990)), and the dimer interface in the DNA-bound structures is solely contained within the C-terminal zinc-binding unit, which has little structural homology with NF.

Although LIM domains are regarded as protein recognition epitopes, no structural information detailing such interactions is available at present. The interface between the two zinc-binding units of single LIM domains is dominated by hydrophobic interactions, and mapping the NF onto, for example, either of the zinc-binding elements of CRIP (Pérez-Alvarado et al., 1996) reveals little overlap between the FOG-binding site of NF and the LIM interfacial residues.

## Conclusions

In summary, we have determined the solution structure of the N-finger of murine GATA-1 and shown that the residues identified as critical for mediating the interaction between NF and finger 6 of FOG form a single contiguous surface, the FOG-binding face. This surface is non-overlapping with the face of the finger which would be expected to contact DNA bases in a GATA-1:DNA complex, where the DNA sequence contains, for example, a double GATA site. Thus it appears that the NF carries more than one distinct function; it can both modulate the DNA-binding properties of GATA-1, by affecting specificity and affinity, and can simultaneously bind to the transcriptional co-factor FOG, thereby recruiting it to the promoter. Comparisons with other proteins that contain zinc-finger:zinc-finger interfaces reveal different modes of interaction, suggesting that protein-protein contacts between zinc-binding motifs may take many forms and are most likely rather widespread (for a review, see Mackay and Crossley, 1998).

## References


Adman, E., Watenpaugh, E.D. and Jensen, L.H. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 4854–4858.

Bartels, C., Xia, T., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR*, **6**, 1–10.

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **65**, 355–360.

Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. and Tschudin, R. (1990) *J. Magn. Reson.*, **86**, 304–318.

Billeter, M., Neri, D., Otting, G., Qian, Y.Q. and Wüthrich, K. (1992) *J. Biomol. NMR*, **2**, 257–274.

Brown, S.C., Weber, P.L. and Mueller, L. (1988) *J. Magn. Reson.*, **77**, 166–169.

Brünger, A.T. (1992) *X-PLOR Manual, Version 3.1*, Yale University Press, New Haven, CT.

Calligaris, R., Bottardi, S., Cogoi, S., Apezteguia, I. and Santoro, C. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 11598–11602.

Cavanagh, J., Palmer III, A.G., Wright, P.E. and Rance, M. (1991) *J. Magn. Reson.*, **91**, 429–433.

Clore, G.M. and Gronenborn, A.M. (1991) *Science*, **252**, 1390–1399.

Clore, G.M., Wingfield, P.T. and Gronenborn, A.M. (1991) *Biochemistry*, **30**, 2315–2323.

Clubb, R.T., Ferguson, S.B., Walsh, C.T. and Wagner, G. (1994) *Biochemistry*, **33**, 2761–2772.

Crossley, M., Merika, M. and Orkin, S.H. (1995) *Mol. Cell. Biol.*, **15**, 2448–2456.

Elefanty, A.G., Antoniou, M., Custodio, N., Carmo-Fonseca, M. and Grosveld, F.G. (1996) *EMBO J.*, **15**, 319–333.

Fasman, G.D. (1989) *Prediction of Protein Structure and the Principles of Protein Conformation*, Plenum Press, New York, NY, p. 65.

Fesik, S.W. and Zuiderweg, E.R.P. (1988) *J. Magn. Reson.*, **78**, 588–593.

Fox, A.H., Kowalski, K., King, G.F., Mackay, J.P. and Crossley, M. (1998) *J. Biol. Chem.*, **273**, 33595–33603.

Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C. and Orkin, S.H. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 12355–12358.

Griesinger, C., Sørensen, O.W. and Ernst, R.R. (1985) *J. Am. Chem. Soc.*, **107**, 6394–6396.

Grzesiek, S. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 12593–12594.

Güntert, P., Braun, W. and Wüthrich, K. (1988) *FEBS Lett.*, **229**, 317–324.

Güntert, P. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 447–456.

Habazettl, J. and Wagner, G. (1995) *J. Magn. Reson.*, **B109**, 100–104.

Haenlin, M., Cubadda, Y., Blondeau, F., Heitzler, P., Lutz, Y., Simpson, P. and Ramain, P. (1997) *Genes Dev.*, **11**, 3096–3108.

Hammarström, A., Berndt, K.D., Sillard, R., Adermann, K. and Otting, G. (1996) *Biochemistry*, **35**, 12723–12732.

Harrison, S.C. (1991) *Nature*, **353**, 715–719.

Hutchinson, E.G. and Thornton, J.M. (1994) *Protein Sci.*, **3**, 2207–2216.

Hutchinson, E.G. and Thornton, J.M. (1996) *Protein Sci.*, **5**, 212–220.

Kline, A.D., Braun, W. and Wüthrich, K. (1988) *J. Mol. Biol.*, **204**, 675–724.

Konrat, R., Weiskirchen, R., Kräutler, B. and Bister, K. (1997) *J. Biol. Chem.*, **272**, 12001–12007.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–55.

Kudla, B., Caddick, M.X., Langdon, T., Martinez-Rossi, N.M., Bennett, C.F., Sibley, S., Davies, R.W. and Arst Jr, H.N. (1990) *EMBO J.*, **9**, 1355–1364.

Kumar, A., Ernst, R.R. and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1–6.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J Appl. Crystallogr.*, **26**, 283–291.

Live, D.H., Davis, D.G., Agosta, W.C. and Cowburn, D. (1984) *J. Am. Chem. Soc.*, **106**, 1939–1941.

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L. P., Yamamoto, K.R. and Sigler, P.B. (1991) *Nature*, **352**, 497–505.

Mackay, J.P. and Crossley, M. (1998) *Trends Biochem. Sci.*, **265**, 1–4.

Mackay, J.P., Kowalski, K., Czolij, R., King, G.F. and Crossley, M. (1998) *J. Biol. Chem.*, **273**, 30560–30567.

Mackay, J.P., Shaw, G.L. and King, G.F. (1996) *Biochemistry*, **35**, 4867–4877.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) *J. Am. Chem. Soc.*, **111**, 1515–1517.

Martin, D.I.K. and Orkin, S.H. (1990) *Genes Dev.*, **4**, 1886–1898.

Merika, M. and Orkin, S.H. (1995) *Mol. Cell. Biol.*, **15**, 2437–2447.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) *FEBS Lett.*, **229**, 317–324.

Norwood, T.J., Boyd, J., Heritage, J.E., Soffe, N. and Campbell, I.D. (1990) *J. Magn. Reson.*, **87**, 488–501.

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J. and Gronenborn, A.M. (1993) *Science*, **261**, 438–446.

Orkin, S.H. (1992) *Blood*, **80**, 575–581.

Palmer III, A.G., Cavanagh, J., Wright, P.E. and Rance, M. (1991) *J. Magn. Reson.*, **93**, 151–170.

Pedone, P.V., Omichinski, J.G., Nony, P., Trainor, C., Gronenborn, A.M., Clore, G.M. and Felsenfeld, G. (1997) *EMBO J.*, **16**, 2874–2882.

Pérez-Alvarado, G.C., Miles, C., Michelsen, J.W., Louis, H.A., Winge, D.R., Beckerle, M.C. and Summers, M.F. (1994) *Nat. Struct. Biol.*, **1**, 388–398.

Pérez-Alvarado, G., Kosa, J.L., Louis, H.A., Beckerle, M.C., Winge, D.R. and Summers, M.F. (1996) *J. Mol. Biol.*, **257**, 153–174.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.-H., D'Agati, V., Orkin, S.H. and Costantini, F. (1991) *Nature*, **349**, 257–260.

Piantini, U., Sørensen, O.W. and Ernst, R.R. (1982) *J. Am. Chem. Soc.*, **104**, 6800–6801.

Ramain, P., Heitzler, P., Haenlin, M. and Simpson, P. (1993) *Development*, **119**, 1277–1291.

Rastinejad, F., Perlmann, T., Evans, R.M. and Sigler, P.B. (1995) *Nature*, **375**, 203–211.

Schellman, C. (1980) In *Protein Folding* (Ed., Jaenicke, R.) Elsevier, Amsterdam, pp. 53–61.

Schwabe, J.W.R., Chapman, L., Finch, J.T. and Rhodes, D. (1993) *Cell*, **75**, 567–578.

Schwabe, J.W.R. and Klug, A. (1994) *Nat. Struct. Biol.*, **1**, 345–349.

Shaka, A.J., Lee, C.J. and Pines, A. (1988) *J. Magn. Reson.*, **77**, 274–293.

Trainor, C.D., Omichinski, J.G., Vandergon, T.L., Gronenborn, A.M., Clore, G.M. and Felsenfeld, G. (1996) *Mol. Cell. Biol.*, **16**, 2238–2247.

Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M. and Orkin, S.H. (1997) *Cell*, **90**, 109–119.

Weiss, M.J. and Orkin, S.H. (1995) *Exp. Hematol.*, **23**, 99–107.

Weiss, M.J., Yu, C. and Orkin, S.H. (1997) *Mol. Cell. Biol.*, **17**, 1642–1651.
Whyatt, D.J., de Boer, E. and Grosveld, F. (1993) *EMBO J.*, **12**, 4993–5005.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.
Wüthrich, K., Billeter, M. and Braun, W. (1983) *J. Mol. Biol.*, **169**, 949–961.
Wuttke, D.S., Foster, M.P., Case, D.A., Gottesfeld, J.M. and Wright, P.E. (1997) *J. Mol. Biol.*, **273**, 183–206.
Yang, H.-Y. and Evans, T. (1992) *Mol. Cell. Biol.*, **12**, 4562–4570.
Yang, H.-Y. and Evans, T. (1995) *Mol. Cell. Biol.*, **15**, 1353–1363.