Characterization of the Conserved Interaction between GATA and FOG Family Proteins

Received for publication, May 13, 2002, and in revised form, July 2, 2002
Published, JBC Papers in Press, July 10, 2002, DOI 10.1074/jbc.M204663200

Kasper Kowalski‡, Chu Kong Liew‡, Jacqueline M. Matthews§, David A. Gell¶, Merlin Crossley,
and Joel P. Mackay

From the School of Molecular and Microbial Biosciences, University of Sydney, Sydney New South Wales 2006, Australia

The N-terminal zinc finger (ZnF) from GATA transcription factors mediates interactions with FOG family proteins. In FOG proteins, the interacting domains are also ZnFs; these domains are related to classical CCHH fingers but have an His→Cys substitution at the final zinc-ligating position. Here we demonstrate that different CCHC fingers in the GATA family protein U-shaped contact the N-terminal ZnF of GATA-1 in the same fashion although with different affinities. We also show that these interactions are of moderate affinity, which is interesting given the presumed low concentrations of these proteins in the nucleus. Furthermore, we demonstrate that the variant CCHC topology enhances binding affinity, although the His→Cys change is not essential for the formation of a stably folded domain. To ascertain the structural basis for the contribution of the CCHC arrangement, we have determined the structure of a CCHH mutant of finger nine from U-shaped. The structure is very similar overall to the wild-type domain, with subtle differences at the C-terminus that result in loss of the interaction in vivo. Taken together, these results suggest that the CCHC zinc binding topology is required for the integrity of GATA-FOG interactions and that weak interactions can play important roles in vivo.

Zinc binding domains (often referred to as zinc fingers, or ZnFs\(^1\)) are extremely common in eukaryotes; for example, perhaps 3–4% of human genes encode proteins that contain such domains. At least 20 different classes of zinc binding domains have been identified, differing in the number of zinc ions they bind and the identity and spacing of the ligating amino acids.

Since the initial discovery of the classical zinc fingers in TFIIIA (1), it has become clear that many zinc binding domains function as sequence-specific DNA (or RNA) binding motifs. However, it has also become apparent that many classes of zinc binding domains, including LIM domains and PHD domains, have roles as mediators of protein-protein interactions. GATA and classical ZnFs have also been shown to be capable of contacting other proteins, especially in the context of the regulation of gene expression (for review, see Ref. 2).

It has been demonstrated, both in Drosophila melanogaster and in mammals, that interactions between ZnFs of GATA and FOG family transcription factors are essential for normal development. In mammals, GATA-1 and FOG-1 cooperate to drive erythroid development (3), whereas the Drosophila proteins Pannier (a GATA family protein; Ref. 4) and U-shaped (a FOG family protein; Refs. 5 and 6) combine to direct the development of both the heart (7) and sensory bristles (6).

GATA proteins (Fig. 1A; Ref. 8) typically contain two ZnFs with a C\(_2\)-C\(_2\)-C\(_2\)-C consensus. The more C-terminal of these two domains (CF) binds DNA sequences containing GATA sites, whereas the N-terminal zinc finger (NF) may contact promoters with different motifs (9). In addition, the zinc finger domains have been shown to mediate interactions with other proteins, including CREB-binding protein (10), erythroid Kruppel-like factor (11), and members of the FOG family (12–15). FOG family proteins (Fig. 1B) contain either eight or nine ZnFs that are related to the classical CCHH ZnFs (which have a conserved C-X\(_n\)-C-X\(_n\)-H-X\(_n\)-H sequence). Several of the fingers in each FOG protein, however, have an altered consensus sequence in which the final zinc binding histidine is replaced with a cysteine.

Interestingly, it is only these variant CCHC ZnFs that mediate interactions with GATA NFs. The structures of two of these domains (fingers 1 and 9 from D. melanogaster U-shaped) have recently been determined (16) and were found to resemble classical CCHH fingers in structure. The only notable difference is a short extended portion of the polypeptide backbone before the fourth zinc ligand (Fig. 1C). The GATA binding surface of one CCHC finger (finger 1 from U-shaped; USH-F1) has been determined (16, 17). Two of the residues implicated in GATA binding are located on the extended portion of the backbone, immediately preceding the final cysteine, and mutation of the cysteine to histidine (changing the finger to a classical CCHH ZnF) is sufficient to abolish the GATA-FOG interaction in vivo (18). The CCHC topology therefore appears to play a key role in facilitating the interaction, but the molecular explanation for this is currently unknown.

The residues that are implicated in GATA binding are largely conserved across FOG family CCHC ZnFs (Fig. 1D), although three of the residues that affect the binding affinity show some variation. This variation gives rise to strong and weak interac-
The zinc finger domains of GATA-1 and U-shaped. A, schematic diagram of GATA-1. Zinc finger domains (NF and CF) are shown in black. B, schematic diagrams of FOg-1 and U-shaped. CCHH ZnFs are shown in gray, and CCHC ZnFs shown in black; ZnFs that can bind the GATA-1 NF are indicated with a star. C, ribbon diagram of the solution structure of USH-F9 (16). The extended region just before the fourth zinc ligand is indicated with a bracket. D, alignment of the amino acid sequences of murine FOg-1 and D. melanogaster U-shaped. Fingers that interact with GATA-1 NF contain a conserved motif. Residues essential for the interaction, judging from mutagenesis studies, are shown in a dotted box; residues that are important (but not essential) for the interaction are shown in a dashed box. Zinc ligands are underlined. The numbering from the native proteins is indicated beside the sequences, with the numbering system used in the text shown at the bottom.

**Strong Interactors**

FOG-1 F1: VFCPCDKCSWTQNYRSGSLNQLSFLGKYSGEASR-C811
FOG-1 F6: RTLGEACETFRRHETTVKHYNLGKYPDQESR-C721
FOG-1 F9: HRYCRCLILIPFSLSLFIHAKKQCPPY-C899
UsH-1 F1: RFLCPGCQASDPSLEHAQQACR231

**Weak Interactors**

FOG-1 F5: GATCFCEETTPNIVNYVHKRLSNGS-C811
UsH-5 F5: QLIACAGGKKTSDLNLHADQNYCPPKPP-822
UsH-9 F9: KYKSCDITISFSNYVTLLKAVHKNCR-NK1145

**Non-Interactors**

FOG-1 F7: YHECTACRVSFHSLALYAHKLYCSPA-C864
UsH-3 F7: DRYSHC0D1FBRNKTAAKHYSSR-367

**EXPERIMENTAL PROCEDURES**

Expression and Purification—USH-F1, USH-F9, and murine GATA-1 NF were expressed and purified as previously described (16, 23). Pannier NF (Pannier (165–208)) was subcloned, overexpressed, and purified using the same protocol as that used for the GATA-1 NF, except that the template for PCR of the construct was obtained from a D. melanogaster cDNA library. The C1142H mutant of U-shaped finger 9 (C32H), encoding residues 1113–1146 of U-shaped, was produced by single-primer mismatch PCR using a 3′ primer encoding the Cys → His substitution. The amplified oligonucleotide was inserted into the E. coli expression vector pGEX-2T (Amersham Biosciences). The peptide (with residues numbered 1–36, including Gly-1 and Ser-2, which are the product of the thrombin recognition sequence) was overexpressed and purified by glutathione-affinity chromatography and reverse-phase high performance liquid chromatography as previously described (23). The purification yielded ~2 mg of >95% pure peptide per liter of culture. The identity of the peptide was confirmed using electrospray mass spectrometry (Mobs. = 4201.8 Da; Mtheor. = 4200.0 Da). Overexpression of 15N-labeled C32H was performed using a bacterial fermentor as previously described (16).

Sample Preparation for NMR—Samples for HSQC titrations or structure determination were prepared by dissolving 0.4 mg of 15N-labeled USH-F9, 2.6 mg of Pnr-NF, or 2.1 mg of C32H in H2O/D2O (95:5, 500 μl) containing 1.5 molar eq of both tris(2-carboxyethylphosphine)/HCl and ZnSO4; this gave sample concentrations of 200–250 μM for USH-F9 and Pnr-NF/C32H, respectively. The pH was adjusted to 5.0 using 0.1 and 0.01 M NaOH.

NMR Spectroscopy—NMR experiments were performed on a Bruker DRX-600 spectrometer equipped with a 5-mm triple- resonance gradient probe. Spectra were acquired at 293 K using spectral widths of 12 ppm for 1H and 30 ppm for 15N. The following homonuclear two-dimensional spectra were recorded on the unlabeled sample: total correlation spectroscopy (24) with MLEV mixing (T1 = 70 ms), NOE spectroscopy (T1 = 50 and 250 ms; Ref. 25), double quantum-filtered COSY (26), and E.COSY (27). Using the 15N-labeled sample, HSQC (28, 29) and three-dimensional HNHA (30) experiments were recorded. For the detection of two- and three-bond scalar couplings involving the histidine side chain nitrogen, an HSSQC was recorded in which the dephasing delay 1/4Δ was set to 11 ms (31), and the carrier frequency and spectral width...
were set to 200 ppm and 6000 Hz, respectively. Solvent suppression was achieved using either pulsed-field gradients or presaturation (double quantum-filtered-COSY and E.COSY). NMR data were processed as previously described (25) and analyzed using XEASY (32). The 1H frequency scale of all spectra was directly referenced to sodium 3-trimethylsilyl[2,2,3,3-2H]propanoate (d4-TSP) at 0.00 ppm, whereas the 15N frequency scale was indirectly referenced to liquid NH4, via the 1H frequency of the d4-TSP resonance (33).

**HSQC Titrations—**Samples were prepared as described above. [15N]C32H and [15N]USH-F9 were prepared at concentrations of 200 μM, and unlabeled Pnr-NF was prepared at a concentration of 1 mM. [15N]-Labeled USH-F9 was titrated with unlabeled Pnr-NF, and [15N]HSQC spectra were acquired at Pnr-NF:USH-F9 molar ratios of 0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, 4.0, and 5.0. The same titration was performed using [15N]-labeled C32H and unlabeled Pnr-NF.

**NMR Line-shape Analysis for USH-F9-Pnr-NF Complex—**The binding of USH-F9 (U) and Pnr-NF (P) to form a complex (UP) can be expressed as [U] + [P] ⇌ [UP], where the association constant, Kassoc, is given by

$$K_{assoc} = \frac{[UP]}{[U][P]}$$  

(Eq. 1)

where $[U]$ and $[P]$ are the total concentrations of USH-F9 and Pnr-NF, respectively, at a given point in the titration, then $K_{assoc}$ and [UP] can be expressed as

$$K_{assoc} = \frac{[UP]}{([U] - [UP])[P] - [UP]}$$  

(Eq. 2)

$$[UP] = \frac{1}{2}x - \frac{1}{4}\frac{[U][P]}{2UP}$$  

(Eq. 3)

where $x = ([U] + [P] + 1/K_{assoc})$. If the product of the lifetime of the complex (τup) and the difference in chemical shift between the free and bound states ($\delta_{up} - \delta_{f}$) is much less than one (i.e., $\tau_{up}\Delta \delta \ll 1$), then the system is considered to be in fast exchange. Under these conditions, a single resonance is observed at the population-weighted average chemical shift ($\delta_{1av}$) of the nuclei in the free and bound states,

$$\delta_{1av} = \delta_{f}u_{0} + \delta_{u}p_{0}$$  

(Eq. 4)

where $\delta_{u}$ is the chemical shift of a nucleus in free USH-F9, $\delta_{up}$ is the chemical shift of the same nucleus in the USH-F9-Pnr-NF complex, and $f_{u}$ and $f_{up}$ are the mole fractions of the free and bound species, respectively (given by $f_{u} = [U]/[U]+[UP]$). Rearranging gives

$$\delta_{1av} = \delta_{f}u_{0} + (\delta_{up} - \delta_{f})[UP]/[U]$$  

(Eq. 5)

If (3) is substituted into (5), the resulting expression can be used to fit a plot of $\delta_{1av}$ against $P_{0}$ using nonlinear least-squares methods. Data from the titration of USH-F9 with Pan-NF were fitted using the non-linear least squares package provided in Origin (Microcal, MA); the frequency at which the signal amplitude was a maximum was determined. This simulated binding curve was compared with the experimental data by calculating the sum of the squares of the differences between the simulated and actual data ($\chi^2$). The values of $\tau_{up}/\tau_{up}^{0}$, $K_{assoc}$, and $k_{off}$ were varied in a grid search to find the best fit of the data to the model.

### Isothermal Titration Calorimetry (ITC)

Lyophilized peptides were dissolved in H2O/D2O (95:5) to a final concentration of 150 μM and treated with both 1.5 molar equivalents of ZnSO4 and 2 molar eq of tris(2-carboxyethylphosphine)HCl. The pH was adjusted to 5.0 (using 10 mM NaOH), and sodium acetate (pH 5.0) was added to a final concentration of 20 mM. The samples were then concentrated by vacuum centrifugation (at 55 °C) to the required concentrations (see Table 1). One-dimensional 1H NMR spectra of each sample were acquired to confirm that they were correctly folded, and samples were dialyzed overnight into a buffer containing sodium acetate (20 mM, pH 5.0), 10 mM ZnSO4, 1.5 mM, and tris(2-carboxyethylphosphine)HCl (100 mM). Experiments at pH 7 were conducted in 20 mM MOPS buffer. Samples were degassed for 5 min immediately before the titrations. Titrations were carried out on a Microcal VP-ITC microcalorimeter. Of the two proteins samples to be used in each titration, 1.8 mM of the less concentrated protein sample was transferred into the cell, whereas the automatic injector was filled with the more concentrated protein (the titrant). Titrations involving USH-F1 and USH-F9 were conducted at 278 and 293 K, respectively. The titrant was automatically injected into the cell containing either its binding partner or buffer alone. For each titration, 15–20 injections were made at 5-min intervals, with each injection consisting of 10–15 μl of titrant. The reference power was set at 5 μcal s⁻¹, and the cell was stirred continuously at 310 rpm. Data were analyzed using the Origin isothermal titration calorimetry analysis software (Microcal Software, Northampton, MA). A linear fit of the isotherm of the protein injected into buffer alone was subtracted from the experimental isotherms to account for the heat of dilution and mixing. A nonlinear least squares fit to a single binding site model was used to obtain values for the binding constant, stoichiometry, and heat of binding.

### Structure Determination for C32H—**Resonance assignment was carried out using the sequential assignment method. The three-dimensional HNHA experiment was used to assign the 1H, 15N HSQC of C32H. Cross-peaks in the two-dimensional NOE spectra were integrated in XEASY and converted to upper distance limits using the CALIBA module of DYANA (35) and the default DYANA parameters. 1H,15N coupling constants were measured by analysis of the three-dimensional HNHA experiment. In combination with distance restraints, a set of allowable φ dihedral angles was generated using the GRIDSEARCH module of DYANA. Stereoscopic assignments for 13 pairs of methylene protons were obtained from an analysis of E.COSY and short mixing time NOE spectra. Stereoscopic assignments in conjunction with interproton distance restraints were used to generate a set of allowable ψ dihedral angle restraints in the GRIDSEARCH module of DYANA. DYANA was used to calculate 50 structures from random starting conformers. Calculations were performed in the absence of zinc ligation restraints. The conformer with the lowest penalty function value was used as input for refinement by restrained molecular dynamics/simulated annealing calculations in CNS (36). Zinc was incorporated into calculations by the introduction of covalent restraints to maintain tet-

#### Table I

| Titration          | Concentration of protein (μM) |
|--------------------|-------------------------------|
| GATA-NF into USH-F1| 58                            |
| USH-F1 into GATA-NF| 700                           |
| Pan-NF into USH-F1 | 18                            |
| USH-F1 into Pan-NF | 15                            |
| Pan-NF into USH-F9 | 400                           |
| GATA-NF into USH-F9| 8–17                          |
| PAN-NF into USH-F9 | 500–680                       |
Fig. 2. The GATA binding face of USH-F9. A, 1H,15N HSQC titration of 15N-labeled USH-F9 with Pan-NF. The 1H,15N HSQC spectrum of USH-F9 alone is shown in black, whereas the 1H,15N HSQC spectrum of USH-F9 in the presence of 5 molar eq of Pan-NF is shown in red. B, plots of average chemical shift change versus residue number for USH-F9 and USH-F1. A dotted line indicates the lower limit for residues defined as having undergone substantial change. Average chemical shift changes were calculated using the H15 and N atoms of USH-F9 and the H15, N, Cα, and C atoms of USH-F1. C, space-filling model of USH-F9 with residues identified in B, highlighted in red (except for Tyr-24, which lies on the rear surface of the structure as shown). Additional residues that are thought to be important for binding (16, 17) are shown in yellow.

RESULTS

The GATA Binding Face of Different FOG Family CCHC Fingers Is Conserved—We first sought to determine the surface—rhedral geometry, S·Zn and N2·Zn bond lengths were constrained to 2.3 and 2.0 Å, respectively, with force constants of 250 kcal mol−1 Å−2. Bond angles defining the zinc coordination site were constrained to the following values: 112° for S·Zn·S·Zn bonds, 108° for Cα·S·Zn angles, 111° for S·Zn·N2·Zn angles, and 102° for N2·Zn·N2·Zn angles. These values were applied with a force constant of 50 kcal mol−1 deg−2. The standard protein all-hydrogen force field was used, and covalent geometry was constrained using standard CNS parameters. Calculations were performed in torsion-angle space, with randomized initial atomic velocities. The first stage consisted of a search of torsion-angle space at high temperature (50000 K), with 1000 time steps of 0.015 ps and a low weight on the repulsive energy term (wrep = 0.1). Force constants for calculation of the square-well NOE and dihedral angle potentials were 150 kcal mol−1 Å−2 and 100 kcal mol−1 deg−2, respectively. The second stage consisted of slow cooling torsion-angle dynamics, with 1000 time steps of 0.015 ps, in which the temperature decreased in 250-K steps. The weight on the repulsive energy term was increased to one, and force constants remained unaltered. The third and final stage consisted of 10 cycles of conjugate gradient minimization (for 1000 steps), with force constants for NOE and dihedral angle potentials changed to 50 kcal mol−1 Å−2 and 300 kcal mol−1 deg−2, respectively. A total of 200 structures were calculated, and the 20 conformers with the lowest total energy were used to represent the solution structure of C32H. The structures were visualized and analyzed using the programs MOLMOL (37) and PROCHECK-NMR (38).

The chemical shift changes observed during the titration were summarized by calculating an average chemical shift change (averaged over HN and N atoms; Fig. 2B). Residues were judged to have undergone a substantial chemical shift change if the average chemical shift change was >0.3 ppm and these (Tyr-10, Phe-18, Thr-23, His-27, and Lys-33) are mapped onto the structure of USH-F9 in Fig. 2C (red). Some residues that were expected to be involved in the interaction (Ile-16, Phe-30, and Tyr-31), judging from mutagenesis studies using
FOG-F1, did not show a significant change, perhaps because their bulky hydrophobic side chains shield the H\textsuperscript{N} and N from the effects of Pannier binding. A similar result was observed for USH-F1 (Fig. 2B; Ref. 16), where the backbone resonances of residues Phe-18, Pro-21, Thr-23, Leu-24, His-27, and Tyr-31 each showed significant chemical shift changes. In addition, the side chain resonances of Ile-16 and Tyr-30 in USH-F1 underwent substantial shifts.

The additional GATA-interacting residues that were identified in the FOG-F1-GATA-1 NF mutagenesis study (Ile-16, Asn-19, Phe-30, and Tyr-31) are highlighted in yellow in Fig. 2C. The commonality of residues involved in NF binding of USH-F1, USH-F9, and FOG-F1 (17) suggests that a common mechanism of interaction between GATA NFs and FOG CCHC ZnFs exists. Note that the changes observed for Tyr-10 and Tyr-24 during the USH-F9-Pnr-NF titration probably reflect conformational differences that arise as a result of NF binding, given that the side chains of these residues lie on the opposite face of USH-F9 to the majority of the affected residues.

Quantifying the USH-F9-Pnr-NF Interaction Using NMR Methods—Sedimentation equilibrium experiments performed on a USH-F1 and GATA-NF mixture (16) suggested that the interaction between them might be of moderate affinity. We therefore sought to estimate the affinity of the \textsuperscript{15}N-labeled USH-F9-Pnr-NF interaction by analyzing our \textsuperscript{1}H,\textsuperscript{15}N HSQC titration data. Several USH-F9 resonances (Cys-14, Val-21, and Tyr-31) were in fast exchange, and independent fitting of their chemical shift changes using a simple heterodimerization model (Equation 5 and Fig. 3A) gave an average \(K_a\) of \((1.3 \pm 0.6) \times 10^4\) M\(^{-1}\).

Some line broadening was observed for residues that appeared to be in fast exchange during the course of the titration. This suggested that the lifetime of the complex approached the chemical shift difference between the free and bound states for these resonances. It is also evident that the binding curves for these resonances deviate from the rectangular hyperbolic binding curves, characteristic of a system in fast exchange (Fig. 3B). These observations are indicative of intermediate chemical exchange (40), and the model described by Equations 7–9 was therefore fitted to these data. In the case of Val-21, fitting of the change in \textsuperscript{15}N chemical shift against \(P_0\) gave a \(K_a\) of \(2.5 \times 10^4\) M\(^{-1}\), a \(\nu_{IP} - \nu_{l}\) of 64 Hz, and a \(k_{eff}\) of 185 s\(^{-1}\). It is clear from the fitted curves that the intermediate exchange model is more appropriate for this resonance (Fig. 3C). For this fit, values of 32 and 20 ms were used for \(T_{2UP}\) and \(T_{2UP}\), respectively. These were determined from \textsuperscript{15}N line widths of the Val-21 signal in the free and complexed forms in high resolution HSQC spectra.

Comparing the Affinities of Different GATA-FOG Complexes—To obtain an independent estimate of the affinity of the USH-F9-Pnr-NF complex and to compare this to the affinities of the USH-F1-GATA-1 NF complex (16) and the USH-F1-Pnr-NF complex, titrations were carried out in an isothermal titration calorimeter. Analysis of the data from USH-F9-Pnr-NF (Fig. 4A) and USH-F1-Pnr-NF (Fig. 4B) titrations yielded association constants of \((1.9 \pm 0.1) \times 10^4\) M\(^{-1}\) and \((1.45 \pm 0.1) \times 10^4\) M\(^{-1}\), respectively. Titrations of the same two USH fingers with the murine GATA-1 NF that was used for previous studies (16) showed that the affinities for this domain are similar to those for the “same organism” interactions \((K_{ass,USH-F1-GATA-1,NF} = (2.3 \pm 0.1) \times 10^5\) M\(^{-1}\) and \(K_{ass,USH-F9-GATA-1,NF} = (1.9 \pm 0.1) \times 10^4\) M\(^{-1}\)). Furthermore, it appeared that pH had a relatively small effect on the \(K_a\) in the range 5.0–7.0 (data not shown). This is an important result given that all structural studies on this system have been conducted in the pH range 5.0–5.5.

The Structure of the C32H Mutant of USH-F9—It has previously been shown in a yeast two-hybrid assay that mutants of murine FOGERG fingers 1 and 9 in which the final zinc-ligating cysteine was mutated to a histidine (to simulate a classical CCHH configuration) were unable to interact with the GATA-1 NF (18). This result, together with the observation that all known GATA-interacting fingers in FOG family proteins possess the CCHC topology, led us to ask about the role of this distinctive ligand arrangement. To do this, we mutated the final zinc-ligating cysteine of USH-F9 to histidine (Cys-1142 → His or Cys-32 → His in the current numbering), creating a CCHH domain similar to classical ZnFs. Circular dichroism and one-dimensional \textsuperscript{1}H NMR data confirmed that the peptide was folded (data not shown).

We then determined the structure of C32H using homo- and heteronuclear NMR methods. The positions of most resonances
were essentially unchanged compared with wild-type USH-F9 (16); changes were restricted to the loop between the two ligating histidines (His-27 and His-32) and to residues Thr-13 and Cys-14, which are in the immediate vicinity of Cys-32 in wild-type USH-F9. The tetrahedral nature of the zinc coordination site and the identity of the ligands were confirmed by calculating structures in the absence of any zinc coordination restraints. On the basis of these structures, it was evident that the thiol sulfurs of Cys-11 and Cys-14 and the N\textsuperscript{\textmd{2}} atoms of His-27 and His-32 were the zinc binding atoms and that the coordination sphere was approximately tetrahedral in nature. The identities of the zinc binding nitrogen atoms were further confirmed by the patterns observed for the histidine side chains.

Fig. 4. Analysis of GATA-FOG complexes using isothermal titration calorimetry. A, raw (upper traces) and integrated (lower traces) data from a titration of Pan-NF into USH-F9. The fit to a simple 1:1 binding model is also shown. B, raw and integrated data from a titration of Pan-NF into USH-F1. The fit to a simple 1:1 binding model is also shown.

The identities of the zinc binding nitrogen atoms were further confirmed by the patterns observed for the histidine side chains in an HSQC spectrum optimized for long range (\(^2\)J and \(^3\)J) couplings. In the final structure calculations, the geometry of the zinc binding site was defined using standard inter-atomic distances and angles. Table II summarizes the experimental constraints used in this final round of structure calculations. No hydrogen-bonding constraints were used in the calculations.

A total of 200 structures were calculated in CNS, and the 20 structures with lowest overall energies were used to represent the solution structure of C32H (Fig. 5A). 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 II). There are no violations of distance or angle constraints greater than 0.11 \(\text{Å}\) and 1.5°, respectively. A PROCHECK-NMR analysis shows that for residues Tyr-10–Cys-11, 98% of backbone \(\phi/\psi\) pairs lie within the most favored or additionally allowed regions of the Ramachandran plot. Over the same region, atomic root mean square differences for the final 20 structures with respect to the mean coordinate positions are 0.20 ± 0.08 \(\text{Å}\) for the backbone atoms and 0.80 ± 0.13 \(\text{Å}\) for all heavy atoms. The atomic coordinates for this family of conformers have been deposited with the Protein Data Bank (accession code 1JN7).

| TABLE II |
| --- |
| Structural statistics for the family of 20 C32H structures |

| Experiment constraints | Value |
| --- | --- |
| Intraresidue distances (\(\|i - j\| = 0\)) | 91 |
| Sequential distances (\(\|i - j\| = 1\)) | 168 |
| Medium range distances (\(\|i - j\| < 5\)) | 148 |
| Long range distances (\(\|i - j\| > 5\)) | 139 |
| Mean root mean square differences from idealized covalent geometry\(^a\) | 0.0013 ± 0.00008 |
| Mean root mean square differences from experimental restraints | 0.352 ± 0.013 |
| Mean root mean square differences from NOE | 0.149 ± 0.009 |
| Mean root mean square differences from dihedral angles | 0.011 ± 0.0006 |
| Mean root mean square differences from improper angles | 0.133 ± 0.032 |
| Mean root mean square differences from restraint violations | 0.6 ± 0.7 |
| Maximum NOE violation | 0.11 |
| Maximum angle violation \(^b\) | 2.8 ± 1.2 |
| Mean angle violation | 1.5 |
| Mean energies (kJ \cdot mol\(^{-1}\)) | 3.2 ± 0.4 |
| | 0.239 ± 0.093 |
| | 5.9 ± 0.6 |
| | 0.50 ± 0.05 |
| | 0.83 ± 0.07 |
| | 12.20 ± 0.34 |
| | 23.5 ± 1.2 |
| Atomic root mean square differences versus mean \(\text{Å}\)\(^b\) | 0.20 ± 0.08 |
| All heavy atoms (9–33) | 0.80 ± 0.13 |

\(^a\) Idealized geometry is defined by the CHARMM force field as implemented in CNS.

\(^b\) Atomic differences are given as the average root mean square distance against the mean coordinate structure. All energies, violations and root mean square differences are given as the mean ± S.D.

The structure of C32H (Fig. 5B) comprises a short \(\beta\) hairpin (residues Tyr-10–Cys-11 and Ile-16–Ser-17) followed by an \(\alpha\)-helix (residues Val-21–Phe-30) and is essentially the same as
As for USH-F9, the strands of the hairpin in C32H are connected by a type I β turn, with a type IV turn connecting the hairpin and the helix. The expected backbone hydrogen bonds are observed throughout the helix (unlike USH-F9, where hydrogen bonds are not observed in some positions) and across the hairpin. However, whereas in USH-F9 the carbonyl group of Cys-11 forms hydrogen bonds with the amide protons of both Asp-15 and Ile-16, the Cys-11 carbonyl group only forms hydrogen bonds with the amide of Ile-16 in C32H. This is most probably because of a slight
change in conformation of the two zinc binding cysteines, a consequence of the cysteine to histidine substitution, which alters the zinc coordination site. This is the only change in this region of the structure.

A comparison of the zinc binding sites shows that there are subtle changes in the position of the zinc binding atoms or the zinc ion itself (Fig. 6B). Cys-11, Cys-14, and His-27 adopt the same conformation in both USH-F9 and C32H, whereas a slight change in the conformation of His-32 is apparent. In the mutant, His-32 takes up a similar position to the corresponding histidine of classical CCHH ZnFs, with the more helical conformation of this region of C32H causing the histidine ring to approach the zinc ion from a different direction compared with USH-F9. The incorporation of the histidine does not affect the conformations of residues in the hydrophobic core (residues Cys-11, Ile-16, Thr-13, Tyr-24, and His-27).

The α-helices of C32H and USH-F9 are superimposable up to residue Gln-29, after which the effect of the Cys → His substitution becomes apparent (Fig. 6A). Although the residues immediately preceding the final zinc binding ligand exist in an extended conformation in USH-F9 ($\phi_0 = -150^\circ$ and $\phi_1 = -109^\circ$), the same residues adopt a conformation that is closer to α-helical in C32H ($\phi_0 = -104^\circ$ and $\phi_1 = -75^\circ$). Residues Cys-32 and Lys-33 in USH-F9 also adopt an extended conformation and do not form any regular secondary structure. In C32H, residues His-32 and Lys-33 both appear to form another half-turn of α-helix, although the turn is not strictly defined as α-helix. This change in conformation probably arises from the differences in the size of the cysteine and histidine side chains. The distance from C$^\text{C}$ to the zinc binding atom is 2.8 and 4.5 Å, respectively, in the two residues. This difference clearly requires different packing to maintain the zinc binding atom of this ligand in the same position in space, which preserves the overall fold of the finger. Despite the changes in packing, however, the H-bonding patterns observed in this region are unchanged; both Tyr-31 H$^\text{N}$ and Cys-32 H$^\text{N}$ make H-bonds with His-27 O.

The impact of the histidine substitution is immediately apparent when examining the positions of the aromatics on the loop, Phe-30 and Tyr-31. Although Phe-30 undergoes a subtle change of conformation and still occupies almost the same region of space, Tyr-31 is shifted well away from its original position (Fig. 6A). The loop preceding His-32 is less pronounced, and the positions of the C$^\text{C}$ atoms of Phe-30 and Tyr-31 now superimpose well with the corresponding residues of classical CCHH ZnFs.

**Interaction of C32H with Pan-NF**—A $^1$H, $^{15}$N HSQC titration was carried out to investigate the ability of C32H to bind Pan-NF. After the addition of 5 molar eq of unlabeled Pan-NF to $^{15}$N-labeled C32H, a number of signals had either decreased in intensity or disappeared completely. These were the same signals that changed in the wild-type USH-F9 titration, indicating that the basic mode of interaction was unchanged. Despite this, HSQC spectra of USH-F9 and C32H, each containing 1 molar eq of Pan-NF, indicated that the C32H-Pan-NF interaction was substantially weaker. Peaks that disappeared rapidly in the titration of USH-F9 were still clearly visible in the titrations of C32H (Fig. 7).

**Protein Binding Versus DNA Binding in Classical Zinc Fingers**—Finally, we sought to determine whether classical-type zinc fingers that contact other proteins (including the CCHC fingers of FOG proteins) have diverged significantly from DNA binding classical zinc fingers. A sequence alignment of 64 zinc finger domains that are known to interact with DNA together with the protein binding zinc fingers of Roaz (41), FOGR-1 (15, 17), Ikaros (42), and YY1 (43) was created using the Pileup program from the Wisconsin sequence analysis package (Supplemental Fig. 1). The program performs comparisons between each possible pair of sequences and creates a final alignment in which the most closely related sequences are nearest to each other in the alignment, whereas the more distantly related ones are separated. This alignment suggests that the sequences of many of the protein binding domains are distinct overall from DNA binding zinc fingers. In part, this presumably reflects the fact that particular residues important for DNA recognition (for example, the basic residue in position 12 that contacts the sugar-phosphate backbone) are not conserved in the protein-binding fingers.

**DISCUSSION**

We have shown here that the molecular details of the interaction between the N-terminal zinc finger of GATA family proteins and CCHC-type zinc fingers from FOG family proteins are highly conserved between different fingers and across different phyla. Thus, the predominantly hydrophobic surface on finger 9 from D. melanogaster U-shaped that contacts the N-terminal ZnF on its in vivo partner, Pan-Nier, is the same as that identified previously in a U-shaped finger 1-GATA-1 NF complex (16). This presumably reflects the conservation of key contact residues in the CCHC fingers (Fig. 1) as well as con-
servation among the N-fingers of GATA family proteins. Interestingly, two of the contact residues (Phe-30 and Tyr-31) are located on the loop preceding the final zinc-ligating residue. Thus, the conformation induced by the His → Cys substitution in this subclass of classical zinc fingers may be required to position the two aromatic side chains in such a way as to allow specific contacts with GATA NFs. These two aromatics are largely conserved across FOG family CCHC ZnFs that are capable of interacting with the NF, and it has been noted previously (44) that protein-protein interfaces exhibit a higher concentration of aromatic residues than protein surfaces as a whole. These residues, however, are not conserved in other classical-type zinc finger domains that are known to mediate protein-protein interactions (such as certain fingers from Roaz, Ikaros, and YY1). In fact, there appears to be substantially more sequence variation among protein binding classical fingers than among DNA binding fingers. This probably reflects the diversity of protein partners that are contacted by these zinc fingers compared with the relative homogeneity of a DNA substrate. In this context, it is also interesting to note that the sequences of many DNA binding zinc fingers are conserved not only on the surface involved in DNA recognition but also on the other surfaces. This may indicate that such domains have additional recognition surfaces; the third finger of erythroid Kruppel-like factor falls into this category because it mediates interactions both with DNA and with GATA-1 (11).

A combination of calorimetric and NMR methods has been used to demonstrate that these interactions are relatively weak, with affinities in the range $10^4$–$10^5$ M$^{-1}$, depending on the domains involved. These affinities are substantially lower than the majority of well characterized biological interactions, and it has become increasingly clear in recent years that weak interactions can play very important roles in biology. The GATA-FOG interactions are also obviously specific; the C-terminal zinc finger from GATA-1 is unable to bind FOG proteins (15), and some FOG CCHC fingers cannot interact with GATA proteins (17). Recent work has demonstrated that FOG-GATA interactions are essential for normal development (45), and it is interesting to speculate as to how weak interactions such as these are formed between proteins that are probably at low concentration in the nuclear milieu. It is possible that high local concentrations of transcription factors exist as a result of interactions that are capable of interacting with the NF, and it has been suggested that such interactions are finely balanced and not tolerant to mutation. Conversely, one can see how novel protein-protein interactions can evolve as a result of small changes in domains that may have originally had a predominant role in DNA binding.

Acknowledgment—We thank Dr. Bill Bulb for expert maintenance of the DRX600 NMR spectrometer at the University of Oxford for valuable advice.

REFERENCES

1. Miller, J., McLachlan, A. D., and Klug, A. (1985) EMBO J. 4, 1609–1614
2. Mackay, J. P., and Crossley, M. (1998) Trends Biochem. Sci. 23, 1–4
3. Tsang, A. P., Fujitani, Y., Hon, D. B., and Orkin, S. H. (1998) Genes Dev. 12, 1176–1186
4. Ramain, P., Heitzler, P., Haenlin, M., and Simpson, P. (1993) Development 119, 1277–1291
5. Haenlin, M., Cubadda, Y., Blendea, F., Heitzler, P., Lutz, Y., Simpson, P., and Ramain, P. (1997) Genes Dev. 11, 3083–3085
6. Fossett, N., Zhang, Q., Gajewski, K., Choi, C. Y., Kim, Y., and Schulz, R. A. (2000) Proc. Natl Acad. Sci. U. S. A. 97, 7348–7353
7. Weiss, M. J., and Orkin, S. H. (1995) Exp. Hematol. 23, 99–107
8. Newton, A., Mackay, J., and Crossley, M. (2001) J. Biol. Chem. 276, 35794–35801
9. Blobel, G. A., Nakajima, T., Eckner, R., Montminy, M., and Orkin, S. H. (1998) Proc. Natl Acad. Sci. U. S. A. 95, 2061–2066
10. Merika, M., and Orkin, S. (1995) Mol. Cell. Biol. 15, 2437–2447
11. Veis, M., and Orkin, S. H. (1995) Mol. Cell. Biol. 15, 2437–2447
12. Tevosian, S. G., Deconinck, A. E., Cantor, A. B., Rieff, H. I., Fujitani, Y., Corfias, G., and Orkin, S. H. (1989) Proc. Natl Acad. Sci. U. S. A. 86, 850–855
13. Holmes, M., Turner, J., Fox, A., Chishold, O., Crossley, M., and Chong, B. (1999) J. Biol. Chem. 274, 23491–23498
14. Svensson, E. C., Tufts, R. L., Polk, C. E., and Leiden, J. M. (1999) Proc. Natl Acad. Sci. U. S. A. 96, 856–860
15. Tsang, A. P., Viivader, J. E., Turner, C. A., Fujitani, Y., Yu, C., Weiss, M. J., Crossley, M., and Orkin, S. H. (1997) Cell 90, 109–119
16. Liew, C. K., Kowalski, K., Fox, A. H., Newton, A., Sharpe, B. K., and Mackay, J. P. (2000) Structure (London) 8, 1157–1166
17. Fox, A. H., Liew, C., Holmes, M., Kowalski, K., Mackay, J., and Crossley, M. (1999) EMBO J. 18, 2812–2822
18. Matthews, J. M., Kowalski, K., Liew, C. K., Sharpe, B. K., Fox, A. H., Crossley, M., and Mackay, J. P. (2000) Eur. J. Biochem. 267, 1030–1038
19. Freson, K., Devriendt, K., Matthijs, G., Van Hoof, A., De Vos, R., Thys, C., Minner, K., Hoylaerts, M. F., Vernynsen, J., and Van Geet, C. (2001) Blood 98, 85–92
20. Freson, K., Matthijs, G., Thys, C., Marien, P., Hoylaerts, M. F., Vernynsen, J., and Van Geet, C. (2002) Hum. Mol. Genet. 11, 147–152
21. Mualley, M. G., Newton, A. L., Gandhi, M. J., Crossley, M., and Drachman, J. G. (2001) Blood 98, 2681–2688
22. Nichols, K. E., Cristino, J. P., Ponce, M., White, J. G., Orkin, S. H., Maris, J. M., and Weiss, M. J. (2001) J. Nat. Genet. 24, 266–270
23. Kowalski, K., Czolij, R., King, G. F., Crossley, M., and Mackay, J. P. (1999) J. Biol. Chem. 274, 12539–12544
Biophysical Characterization of the GATA-FOG Interaction

35729

36. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
37. Koradi, R., Billeter, M., and Wuthrich, K. (1996) J. Mol. Graph. 14, 51–55
38. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) J. Biomol. NMR 8, 477–486
39. Fox, A. H., Kowalski, K., King, G. F., Mackay, J. P., and Crossley, M. (1998) J. Biol. Chem. 273, 33595–33603
40. Kim, S., Cullis, D. N., Peig, L. A., and Baleja, J. D. (2001) Biochemistry 40, 6776–6785
41. Tsai, R. Y., and Reed, R. R. (1998) Mol. Cell. Biol. 18, 6447–6456
42. Sun, L., Liu, A., and Georgopoulos, K. (1996) EMBO J. 15, 5358–5369
43. Lee, J. S., Galvin, K. M., Shi, Y., and Riggs, K. J. (1993) Proc. Natl Acad. Sci. U. S. A. 90, 6145–6149
44. Lo Conte, L., Chothia, C., and Janin, J. (1999) J. Mol. Biol. 285, 2177–2198
45. Crispino, J. D., Lodish, M. B., Mackay, J. P., and Orkin, S. H. (1999) Mol. Cell 3, 219–228
46. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) Nucleic Acids Res. 25, 4876–4882