Articles

Structure of a Designed Dimeric Zinc Finger Protein Bound to DNA[†,‡]

Scot A. Wolfe,*[,§] Robert A. Grant, and Carl O. Pabo[||]

Department of Biology and the Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received May 19, 2003; Revised Manuscript Received September 22, 2003

ABSTRACT: Proteins that employ dimerization domains to bind cooperatively to DNA have a number of potential advantages over monomers with regards to gene regulation. Using a combination of structure-based design and phage display, a dimeric Cys_2His_2 zinc finger protein has been created that binds cooperatively to DNA via an attached leucine zipper dimerization domain. This chimera, derived from components of Zif268 and GCN4, displayed excellent DNA-binding specificity, and we now report the 1.5 Å resolution cocrystal structure of the Zif268-GCN4 homodimer bound to DNA. This structure shows how phage display has annealed the DNA binding and dimerization domains into a single functional unit. Moreover, this chimera provides a potential platform for the creation heterodimeric zinc finger proteins that can regulate a desired target gene through cooperative DNA recognition.

Cys_2His_2 zinc fingers are the motif of choice for the design and selection of DNA-binding proteins with novel sequence specificity (*1*−*3*). Methods for creating zinc finger proteins with novel DNA-binding specificity are well-established (*4*−*6*), and when fused to activation or repression domains, these designer transcription factors can regulate target genes in vivo (*7*−*10*). The zinc finger DNA-binding domains for these designer transcription factors have been assembled both 1) by tandem covalent linkage of a series of zinc fingers (*11*−*14*) and 2) by inclusion of dimerization domains (*15*−*17*). Either of these methods of assembly can be used to increase the size of the effective target site, which may result in further improvements in DNA-binding specificity (*12*, *14*, *17*).

In this paper, we continue our studies of a dimeric zinc finger system that was developed by attaching leucine zippers at the C-terminus of a two-finger DNA-binding domain(*17*). Dimerization systems of the type presented here have a number of very attractive features. The most fundamental potential advantage is that dimerization domains also allow the cooperative assembly of complexes on the DNA. When dimerization occurs in conjunction with DNA binding, this narrows the concentration range over which DNA binding occurs, creating a sharper transition between the fully off and fully on states as a function of protein concentration, thereby providing tighter control over gene regulation (*18*). If dimerization only occurs after DNA binding, this form of association also will allow each monomer to search the DNA independently for its binding site. This should significantly reduce the equilibration rate of the dimer when compared to the same number of covalently linked fingers (*19*, *20*). Proteins with very high affinity for their target sites, such as zinc finger proteins that contain six or more covalently linked fingers, may also bind noncognate sites for appreciable lengths of time (kinetic trapping (*18*)). Given the vast number of nonspecific sites in the genome, relatively rapid equilibration may be an important consideration when attempting to create zinc finger proteins that recognize unique sequences in the human genome (*18*).

We have created a leucine zipper-based dimerization system for zinc fingers using a combination of computer modeling and phage display (*17*). In the computer modeling phase, we used the principles of structure-based design (*15*) to plan how we could fuse fingers 2 and 3 of Zif268 (which we refer to as Zif23) to the leucine zipper of GCN4 (Figure 1). The amino acid sequence that joins the zinc fingers and leucine zipper was subsequently optimized by phage display. This dimeric zinc finger protein, Zif23-GCN4, was selected to recognize a binding site that contains a two base pair overlap between the half-sites normally bound by each monomer (*17*). By overlapping the binding sites of neighboring monomers, a higher density of DNA-binding residues per base pair should be present and thus may allow a higher degree of sequence discrimination than is possible for standard covalently linked zinc fingers. Indeed, the sequence specificity of the four finger Zif23-GCN4 homodimer (as measured by the ratio of the nonspecific and specific dissociation constants) is better than the three fingers of Zif268 while recognizing a site of similar length (*17*).

[†] S.A.W. was supported as a Special Fellow of the Leukemia and Lymphoma Society; C.O.P. and R.A.G. were supported by the Howard Hughes Medical Institute.

[‡] The Zif23-GCN4 structure has been deposited in the Protein Data Bank (accession code 1LLM).

* Corresponding author. E-mail: scot.wolfe@umassmed.edu.

[§] Current address: Program in Gene Function and Expression, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605.

[||] Current address: 257 Throckmorton Ave., Mill Valley, CA 94941.

10.1021/bi034830b CCC: $25.00
Published on Web 10/30/2003

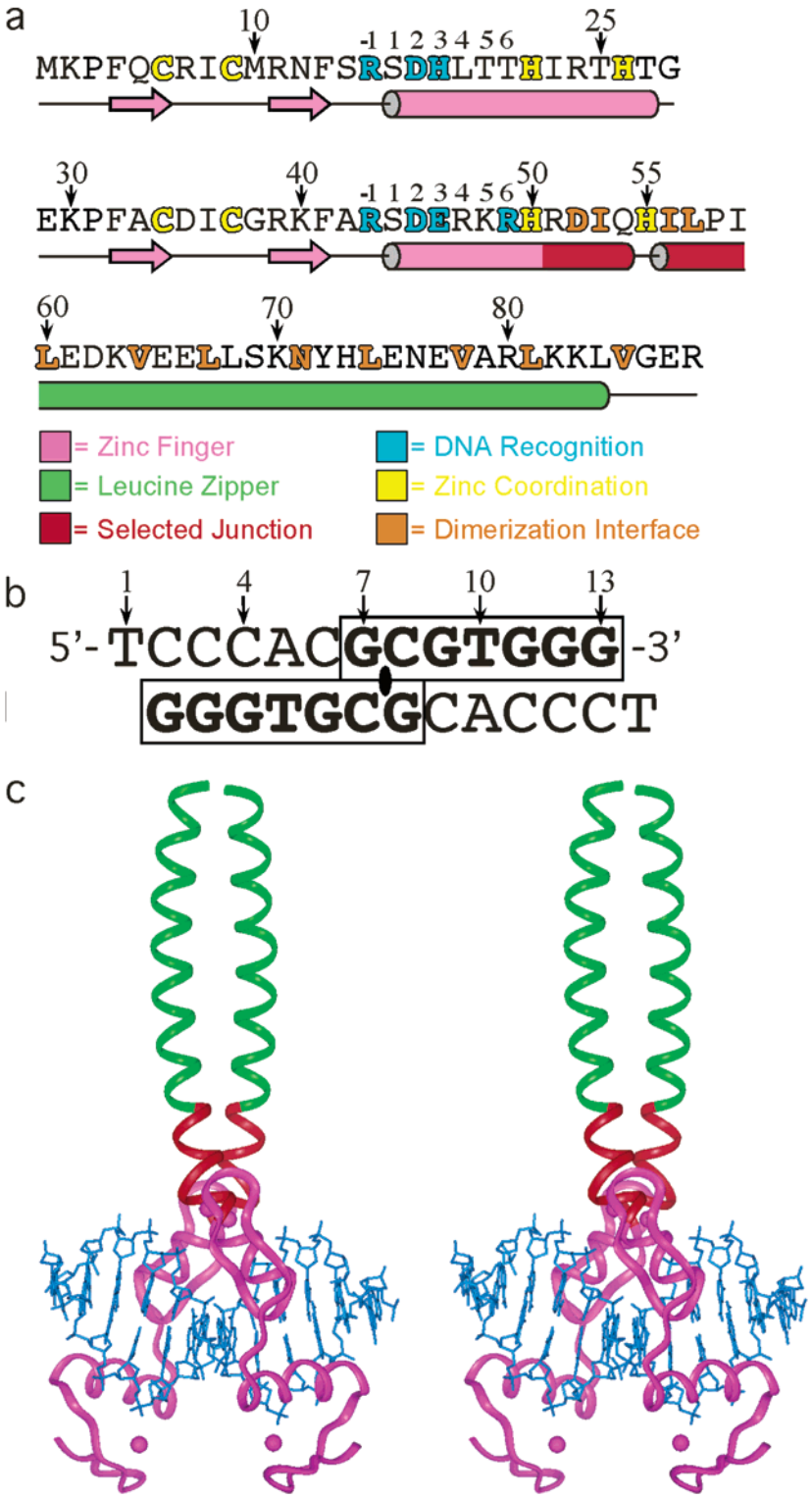

FIGURE 1: Structure of the Zif23-GCN4 homodimer bound to DNA. (a) Sequence of the Zif23-GCN4 protein. The α-helical regions are denoted by a cylinder (below the amino acid sequence), and *β*-strands are indicated by filled arrows. Each monomer contains fingers 2 and 3 of Zif268 (magenta) fused to the leucine zipper of GCN4 (green) through a junction sequence selected by phage display (red) (*17*). Amino acids involved in DNA recognition, zinc coordination, or dimerization are (respectively) blue, yellow, or orange. The numbering scheme (−1, 1, 2, 3, 5, and 6) for residues in the recognition helix of each zinc finger is indicated above its amino acid sequence, with −1 indicating the residue immediately preceding the helix. (b) Sequence of the self-complementary oligonucleotide used for crystallization. The bold, boxed sequence indicates the recognition site on the purine-rich strand for fingers 2 and 3 of Zif268. The Zif23-GCN4 dimer was selected to recognize a palidromic site containing this sequence that is overlapped by two base pairs at the 5′ end. (c) Stereoview of the Zif23-GCN4 homodimer−DNA complex. The ribbon diagram of the protein (residues 4−85) is color-coded as described in panel a, and the DNA is shown in blue (bases 2−13). The Zif23-GCN4 homodimer grasps the DNA with all four fingers (i.e., using both fingers of both monomers). The selected junction between the fingers and the leucine zipper creates a pseudo-continuous α-helix that runs from the start of the α-helix in the final finger all the way through the leucine zipper. The helical axis has a slight kink in the selected region due to a proline that was invariant in the set of sequences obtained by the phage display.

Table 1

Data Collection and MIR Phasing[a]					
data set	native-2	native-1	I-dC #1	I-dU #2	$CdCl_2$
resolution range (Å)	20−1.5	20−1.85	20−2.5	20−3.00	20−2.5
measured reflections	1187929	481498	112166	62959	85678
unique reflections	83074	43197	18890	10952	18509
completeness, %	98.9 (97.9)	95.3 (83.3)	100 (100)	97.8 (96.9)	99.7 (99.2)
R_{sym},[b] %	9.5	8.3	10.4	12.4	11.8
reflections with $I/\sigma_I > 2$, %	89.5 (71.2)	86.9 (61.5)	82.7 (65.7)	83.5 (67.4)	80.3 (61.3)
R_{iso},[c] %			20.0	29.8	13.0
phasing power[d] (acentrics/centrics)			0.98/1.27	0.72/1.00	1.15/1.37
figure of merit, MIR	0.54				

Refinement	
resolution range, Å	20−1.5
reflections, $F > 2\sigma(F)$	77251
no. of non-H atoms	1607
R_{cryst},[e] %	21.6
R_{free},[f] %	23.4
no. of water molecules	312
average *B*-value, Å^2	25.1
Δ*B*-values, bonded atoms (Å^2)	1.6
RMSD from ideal bond lengths, Å (protein/DNA)	0.015
RMSD from ideal bond angles, deg (protein/DNA)	1.640

[a] Values in parentheses are for the highest resolution shell. [b] $R_{sym} = \Sigma(|I - \langle I\rangle|)/\Sigma(I)$. [c] $R_{iso} = \Sigma(|I_{nat} - I_{der}|)/\Sigma(I_{nat})$. [d] Phasing power $= \Sigma(|F_H|)/\Sigma(||F_{PH}| - |F_P + F_H||)$. [e] $R_{cryst} = (\Sigma_{(h,k,l)}||F_o| - |F_c||/\Sigma_{(h,k,l)}|F_o|)$. [f] R_{free} was calculated for the 10% of reflections omitted from the refinement.

As intended, the leucine zipper dimerization system allows the cooperative binding of the chimeric protein to the DNA. It also has the advantage of creating a modular system: heterodimers that recognize asymmetric binding sites can be created by substituting the GCN4 leucine zipper with the c-Fos and c-Jun leucine zippers and by substituting the Zif268 fingers with zinc fingers that have a different DNA-binding specificity (*17*).

On the basis of these attributes, the Zif23-GCN4 dimerization system would appear to provide a promising platform for the creation of zinc finger proteins that regulate single genes in vivo. To understand the molecular details of DNA recognition in this system, we have crystallized the Zif23-GCN4 homodimer bound to DNA and solved the structure of the complex at a resolution of 1.5 Å. The structure validates our original design concept, defines the role that the junction sequences (selected by phage display) play in DNA recognition, and suggests additional refinements that could improve the dimerization system.

EXPERIMENTAL PROCEDURES

Crystallization. The Zif23-GCN4 monomer contains a Met-Lys dipeptide, followed by fingers 2 and 3 of Zif268 (residues 34−81 of the zinc finger domain) (*21*), then the selected linker (RDIQHILPI), and finally residues 253−281 of GCN4 (*22*). This construct is clone #2 from the Zif23-GCN4 proteins previously selected to recognize a two base pair overlap between the zinc finger recognition sequences (*17*). The protein was overexpressed and purified as previously described (*23*). The oligonucleotide used to form the self-complementary 13-mer duplex contains the palindromic-binding site for fingers 2 and 3 of Zif268 (bold) with a two base pair overlap at the 5′ end of the recognition sequence (5′-TCCCACGCGTGGG-3′). This oligonucleotide was synthesized and purified as previously described (*24*). When annealed, the duplex has a 5′ thymine overhang at each end. All crystallization steps were carried out in an anaerobic chamber ([O_2] ∼1 ppm). Prior to complex formation, the protein was refolded at a concentration of 1.9 mM using 1.5 equiv/finger $CoCl_2$ in 80 mM Bis-Tris propane (pH 8.0). The protein−DNA complex was then formed by mixing 12.5 μL of refolded protein with 237.5 μL of duplex DNA (0.056 mM) in 10 mM Bis-Tris propane (pH 8.0) with 1 mM $MgCl_2$. This sample was allowed to equilibrate overnight and was subsequently centrifuged to remove any precipitate. Crystals were grown using hanging drop vapor diffusion: 3.2 μL of complex was allowed to equilibrate over a well solution containing approximately 600 mM NH_4OAc. Crystals (long rods) appeared after about 24 h.

Data Collection. Each crystal was cryo-protected by slowly increasing the concentration of glycerol in the drop to approximately 20% over a matter of minutes. The crystal was then looped directly out of the drop, flash frozen in liquid nitrogen, and then mounted in a liquid nitrogen cooled nitrogen stream (−143 °C). Crystals belonged to the space-group $P3_121$ with unit cell dimensions $a = b = 87.4$ Å and $c = 118.7$ Å. Diffraction data were typically collected using a rotating anode X-ray generator (Rigaku RU-200) and an R-axis IV image plate system (Data sets: Native-1, I-dC#1, I-dU#2, and $CdCl_2$; Table 1); an additional 1.5 Å resolution native data set was also collected at the National Synchrotron Light Source on Beamline X4A. Data were indexed, integrated, and scaled with the *HKL* suite (*25*) (Table 1).

MIRAS Phasing, Model Building, and Refinement. The asymmetric unit contains the homodimer of Zif23-GCN4 bound to one DNA duplex. The phase problem was solved by MIRAS using heavy-atom derivatives generated (1) by substituting 5-iodo-deoxycytosine (I-dC) for cytosine at position 4, (2) by substituting 5-iodo-deoxyuracil (I-dU) for thymine at position 10 in the DNA (derivatives #1 and #2,

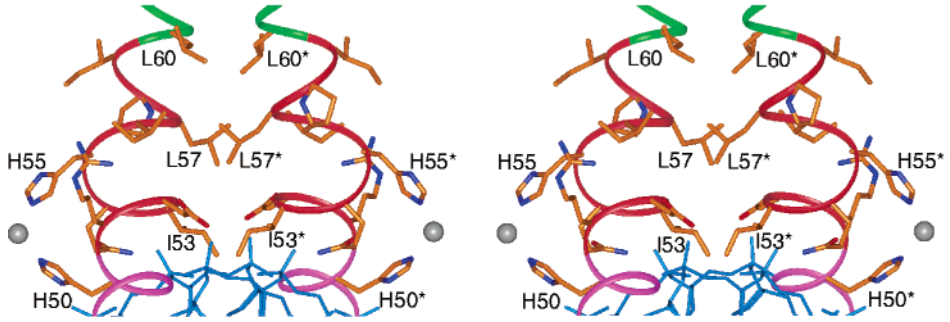

FIGURE 2: Stereoview of the structure of the selected junction. The ribbon depicting the protein backbone is colored as indicated in Figure 1a. DNA is blue. Asterisks denote amino acids in the complementary monomer.

respectively), and (3) by folding the zinc finger protein with $CdCl_2$ instead of $CoCl_2$. Heavy atom coordinates were determined from visual inspection of the isomorphous difference and anomalous difference Patterson maps, and initial phases were calculated and refined using MLPHARE (*26*). Using these phases, the resulting experimental electron density map was of sufficient quality to build an initial model using O (*27*). This model was iteratively refined and rebuilt to generate a high quality 1.85 Å resolution model using a native data set collected on the Rigaku generator (Native-1).

Model refinement was carried out using simulated annealing, positional refinement, and isotropic *B*-factor refinement protocols in X-PLOR (*28*), using a 2 σ cutoff and a bulk solvent correction. As the refinement progressed with the original Native data set (Native-1), waters were added. However, the conformation of Arg49 in this data set was unclear, which led us to collect a higher resolution data set on the same crystal at the synchrotron. This new data set was merged with the original Native-1 data set to create a master data set (Native-2; $R_{merge} = 4.4\%$ to 1.85 Å). Using this master data set, the multiple conformations for Arg49 in each monomer were added, and the structure was subsequently refined to 1.5 Å.

To assess the quality of the model, simulated annealing omit maps were created for residues in the recognition helices and in the junction region that had been selected via phage display. Procheck analysis of the final model indicates that 93.6% of the residues lie in the core region of the Ramachandran plot, with the remaining residue conformations corresponding to allowed regions.

Analysis of the Zif23-GCN4 Structure. Superpositions of the Zif23-GCN4 complex with itself (complementary monomers are not symmetry related) or with Zif268 (*21*) or GCN4 (*22*) were performed in Insight98 (Molecular Simulations). Statistics for the superposition between Zif23-GCN4 and Zif268 were calculated for each monomer using the zincs, backbone atoms for residues 4−48 of Zif23-GCN4 and for residues 35−79 of Zif268, and the 5′-GCGTGGG-3′ base pairs in each structure. This comparison gives an RMSD of 0.34 Å for one monomer and 0.41 Å RMSD for the other. Conformational statistics for the DNA were calculated using 3DNA (*29*).

RESULTS

Overall Structure of the Chimeric Zinc Finger−DNA Complex. The objective of the phage display selections was to optimize the stability and function of the Zif23-GCN4 chimera by creating a complementary fusion between the C-terminal end of the zinc finger α-helix (finger 3 of Zif268) and the N-terminal end of the leucine zipper. The structure of the Zif23-GCN4 dimer shows that this objective was achieved (Figure 1). The selected junction, which consists of residues 51−54 and 56−59 (with His55 of the original zinc finger helix retained to ensure proper zinc coordination), fuses the recognition helix of the final zinc finger and the leucine zipper into a single pseudo-continuous helix with a slight kink at Pro58. As expected, the structures of individual zinc fingers and of the leucine zipper are essentially identical to the corresponding parts of the parent proteins (*21*, *22*). For example, superposition of the Zif23-GCN4 and GCN4-(*22*) backbones in the leucine zipper region (residues 60−81 and 253−274 respectively) gives an RMSD of 0.56 Å.

The Zif23-GCN4 homodimer binds the palindromic DNA sequence using both zinc fingers from each monomer (Figure 1). The GCN4-derived leucine zipper forms a coiled−coil structure that extends out essentially perpendicular to the axis of the DNA double helix. Although Zif23-GCN4 binds DNA as a homodimer, the structure does not display precise 2-fold symmetry at the dimer interface. The average RMSD for superposition of opposite monomers is 1.40 Å. This asymmetry results from a ∼5° angle between the superhelix axis of the leucine zipper and the pseudo 2-fold axis of the dimer defined by the zinc fingers bound to the palidromic site.

The observed asymmetry is most pronounced within the selected junction at Leu57: this residue has different side chain conformations in the two monomers (Leu57: $\chi1 = -179°$, $\chi2 = 55°$; Leu57*: $\chi1 = -70°$, $\chi2 = 172°$; Figure 2). In the dimer, the leucine zipper is also rotated by ∼30° relative to the DNA about the C_2 axis of the coiled−coil interface as compared to the zipper in the parent GCN4 complex (*30*). This change in orientation should reduce the possibility that zinc finger-leucine zipper proteins will form functional heterodimers with members of the bzip family of proteins since the formation of such a heterodimer will either strain the dimer interface or position one of the DNA-binding domains incorrectly.

The DNA conformation in the Zif23-GCN4 structure is B-form, and like the DNA in the Zif268 structure, it has an enlarged major groove (*31*). However, the DNA in the Zif23-GCN4 structure is not as significantly underwound as in the Zif268 structure (overall helical twist for Zif268 11.3 bp/turn, Zif23-GCN4 10.7 bp/turn, and B-DNA 10.0 bp/turn) (*32*). A less underwound duplex (10.7 bp/turn) has also been observed for a dimeric complex in which two monomers of fingers 1 and 2 of Zif268 dimerize through a selected

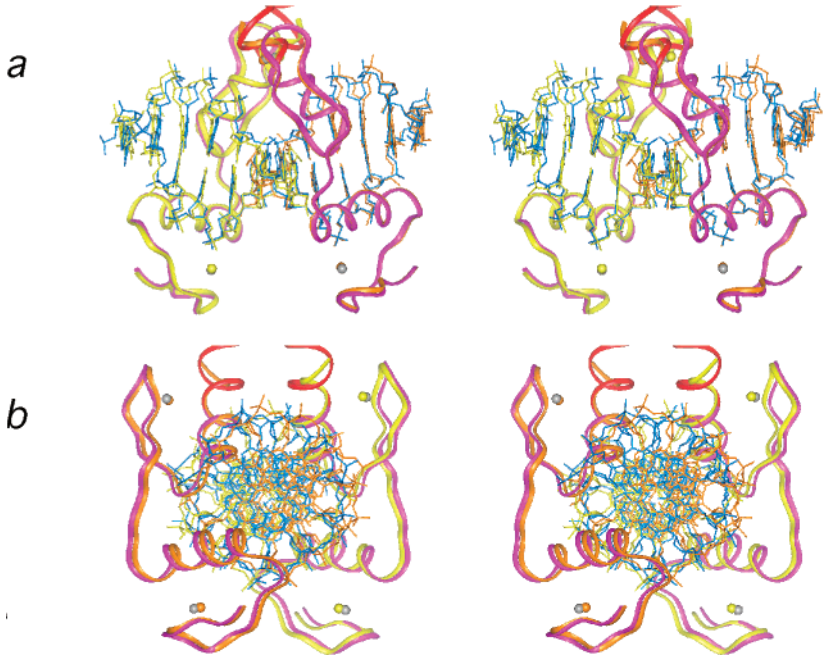

FIGURE 3: Zinc fingers from Zif23-GCN4 and Zif268 dock in essentially identical manners in the major groove. (a and b) Stereoview(s) from two different angles of the superposition of Zif268 (*21*) with the zinc fingers and DNA from each monomer of Zif23-GCN4. One Zif23-GCN4 monomer superimposes with Zif268 slightly better (average RMSD = 0.34 Å) than the other (average RMSD = 0.41 Å). The backbone for each protein is displayed as a ribbon trace, the bound zincs are shown as spheres, and the DNA in each complex is displayed as a stick figure. The Zif23-GCN4 structure is colored as indicated in Figure 1a. The entire Zif268 complex is either orange or yellow.

N-terminal peptide extension (*33*). This difference may represent a fundamental improvement in DNA recognition for the dimeric zinc finger complex when compared with a covalently linked monomer. In principle, the dimer interface has been annealed through selection to optimize the register of its four fingers with the DNA duplex. The more significantly unwound DNA in the Zif268 structure may result from a slight difference in the helical register of the fingers and the DNA that might become more exaggerated (and hence more unfavorable) as more canonically linked fingers are added.

Structure and Function of the Selected Junction. The structure of the selected junction region bridging the final zinc finger and the leucine zipper reveals the advantages of using phage display to optimize this region instead of simply employing a flexible linker. The amino acids at positions 51−59 (RDIQHILPI) in this region are representative of the consensus sequence for the pool of clones obtained in the selections (*17*). Ile 53, Gln 54, and Pro 58 were absolutely conserved in the selected pool. Leucine or isoleucine were the only residues found at position 57, and arginine and aspartate were highly conserved (respectively) at positions 51 and 52 (seven out of eight clones). A hydrophobic residue was also preferred at position 56.

The selected junction sequence contributes both to dimerization and to DNA recognition. This region is well-ordered (average *B*-value = 17.0 ± 2.0 Å^2) and α-helical, except for the kink introduced by Pro58 (Figure 2). In addition to Pro58, other highly conserved residues from the selection (*17*) play obvious roles in defining the structure of the junction. A hydrophobic core is formed via the interaction of Ile 53 and Leu 57 with identical residues from the other monomer, via the interaction of parts of Asp 52 with Ile 53* in the neighboring monomer, and via the interaction of Ile 56 with Leu 57* in the neighboring monomer. (Complementary interactions exist for these residues in the other monomer). Gln 54 makes a phosphate contact. Arg 51 and Asp 52 interact with Asp 35 and Lys 48 (respectively) in the final finger of the same monomer.

The selected junction appears to be relatively rigid due to the hydrophobic core formed between neighboring monomers by Ile53, Ile56, Leu57, and parts of Asp52 (Figure 2). These interactions extend the dimerization interface beyond the final histidine involved in zinc coordination and into the α-helix of the C-terminal zinc finger in each monomer. The resulting interface fixes the relative register of neighboring monomers, thus ensuring a preference to bind a composite sequence that contains a two base pair overlap between the half-sites recognized by each monomer.

Base-Specific DNA Recognition. The protein−DNA interactions observed in the Zif23-GCN4 complex are almost identical to those observed between the DNA and the corresponding fingers in the cocrystal structure of the parent protein, Zif268 (*21*). The average RMS deviation for superposition of the protein backbone, DNA bases, and zincs from each monomer of Zif23-GCN4 with fingers 2 and 3 of Zif268 is only 0.38 Å (Figure 3). Thus, the zinc fingers within the context of the Zif23-GCN4 dimer are properly positioned for base-specific DNA recognition.

The base-specific interactions found in the Zif268 structure are preserved in the Zif23-GCN4 structure except in the region where the two DNA half-sites overlap (Figure 4a). For example, the base-specific contacts from residues at positions −1, 2, and 3 in the α-helix of finger 3 are identical in both structures. However, in the Zif23-GCN4 structure Arg49 in each monomer (position 6 in the α-helix of finger 3) has multiple conformations, and none are identical to the corresponding arginine in the Zif268 structure (Figure 4b). In one conformation, Arg49 contacts the O^6 position of both G7 and G7′. (This conformation cannot be simultaneously

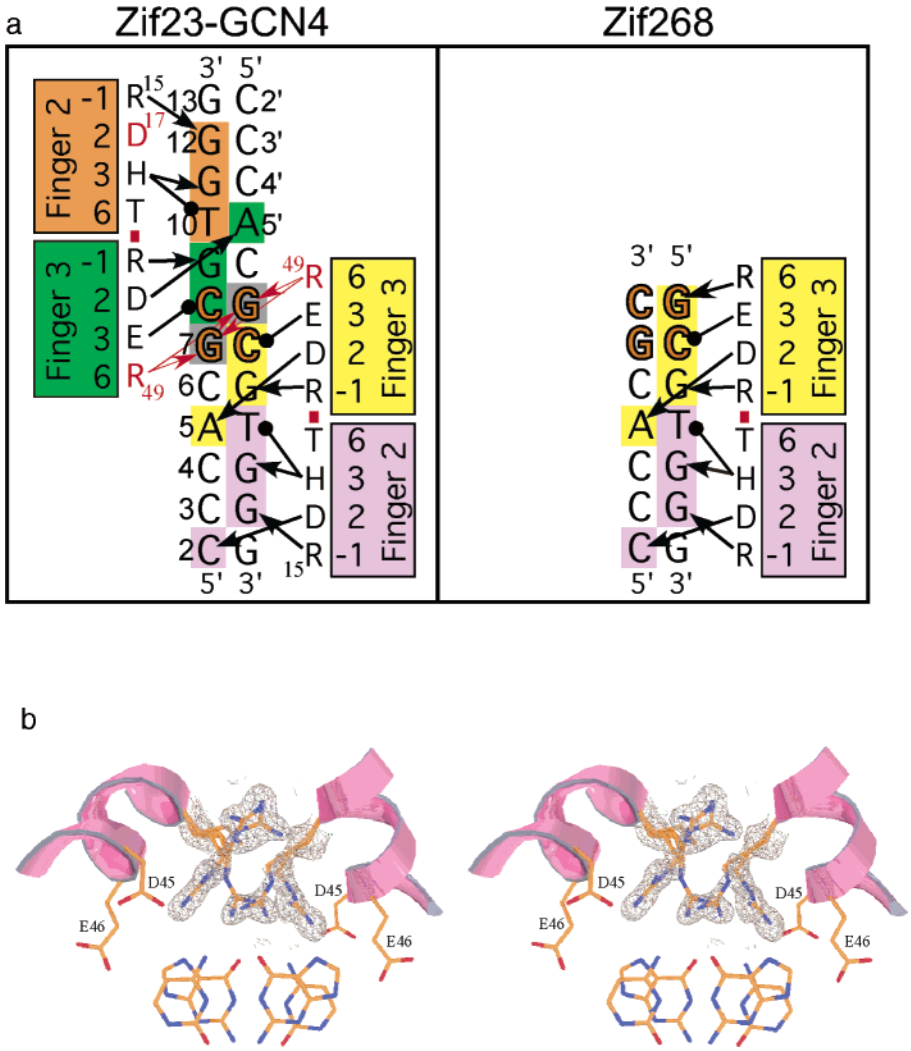

FIGURE 4: DNA recognition by Zif23-GCN4 is nearly identical to DNA recognition by Zif268. (a) Diagram comparing the base contacts made by Zif23-GCN4 (left panel) and fingers 2 and 3 of Zif268 (*21*) (right panel). DNA recognition helices of the zinc fingers are represented as colored rectangles. Positions on the recognition helix typically involved in DNA recognition are numbered, and the amino acid at each position is indicated using the one letter abbreviation. Hydrophobic interactions are indicated by black dots, and hydrogen bonds are indicated by arrows. The DNA region contacted by each finger is color-coded to emphasize the size (and overlapping arrangement) of the binding sites. Amino acids that differ in their method of DNA recognition between the two structures (Asp 17 in one monomer and Arg 49 in both halves of the complex) are red in the Zif23-GCN4 panel. Arg 49 (position 6 in finger 3) has multiple conformations in the Zif23-GCN4 structure. One conformation interacts simultaneously with the O^6 positions of G7 and G7′. This interaction cannot be made concurrently by both monomers because of steric constraints. Red bars between residues at positions 6 and −1 on neighboring fingers indicate the presence of interfinger interactions (*37*). (b) Stereoview of the DNA contacts observed within the two base pair overlap between the binding sites for finger 3 of each monomer. Multiple conformations of Arg 49 are apparent in the Zif23-GCN4 structure. Each monomer has two shared conformations for Arg 49. One conformer interacts with the O^6 positions of G7 and G7′, and the other conformer interacts with Asp 45. Two phosphate contacting conformations are also apparent for Arg 49 from one monomer. Each cytosine (C8/C8′) in the overlap region is contacted by Glu 46 (position 3) from the neighboring monomer as is observed in the Zif268 structure. Electron density for Arg49 is shown from a simulated-annealing omit map (with the all of the Arg 49 conformers from each monomer omitted from the calculation) contoured at 0.75 σ.

assumed by both monomers because it would lead to collisions in a symmetric interface.) In the second conformation, Arg49 interacts with Asp45 (position 2, finger 3) and does not appear to make any productive base-specific contacts. In one monomer, there is also an additional set of conformations for the Arg49 side chain that interact with the phosphate backbone.

It is not clear what precludes Arg49 from establishing the classic bidentate arginine−guanine interaction with O^6 and N^7 of guanine observed in the Zif268 structure. On the basis of a superposition of the structures, it appears that Arg49 in Zif23-GCN4 could assume the same conformation as the corresponding arginine in the Zif268 structure. The distance between the position 6 α carbon and the guanine in both structures is essentially identical, and there do not appear to be any steric clashes between neighboring monomers that would interfere with these contacts.

The only other difference between base-specific contacts in the Zif23-GCN4 and those in the Zif268 structures occurs at Asp17. In one monomer, the interactions of Asp17 (position 2 in finger 2) and Arg 15 (position −1) with the DNA are exactly like those observed in the Zif268 structure. However, in the other monomer, Asp 17 does not interact with the DNA or make the typical bidentate interaction with Arg 15. Crystal packing interactions with a symmetry-related dimer appear to be responsible for the altered conformation

of Asp 17 of this finger. Arg 15 does contact G12 at the O^6 and N^7 positions, but the guanidinium group is rotated 180° from its standard conformation around $\chi 3$.

DISCUSSION

Role of the Selected Junction in DNA Recognition by Zinc Finger−Leucine Zipper Chimeras. The selected junction creates a highly organized interface that simultaneously orients the DNA-binding domains for recognition and the hydrophobic face of the attached leucine zippers for dimerization. Both the length and the composition of the selected junction are critical for its function. Addition or subtraction of a single residue from the end of the junction would place the leucine zipper dimerization interface out of phase with the zinc finger DNA-binding domains for the coordinated recognition of its composite sequence. The selected junction also has one unanticipated property: it extends the dimerization interface into the recognition helix of the final zinc finger. This interaction was not designed into the system, but these residues are highly conserved among the phage-selected proteins and clearly are important in the organization of the Zif23-GCN4 dimer for DNA recognition.

A small degree of asymmetry is observed in the Zif23-GCN4 structure, but its significance is unclear. The ~5° tilt of the leucine zipper relative to the DNA double helix is most likely due to the effect of asymmetric crystal packing forces on the leucine zipper. The most distinct evidence of this asymmetry is found in the difference in orientation of Leu57 in each monomer within the junction. This type of junction asymmetry between a leucine zipper dimerization domain and the attached DNA-binding domain has also been observed for the heterodimer of c-Fos and c-Jun (AP1) bound to DNA in the presence and absence of NFAT1 (*34*, *35*). In the absence of NFAT1, two distinct AP1 complexes bound to DNA are observed in the asymmetric unit. These complexes differ in the degree of bending in the fork (junction) between the basic region and the leucine zipper (*34*). In one complex, the leucine zipper is nearly perpendicular to the DNA, whereas the other complex is bent by ~10°. The distortion in the fork region is even more pronounced in the AP1-NFAT1-DNA structure where the coiled−coil of AP1 bends dramatically away from the perpendicular axis to interact with NFAT1(*35*). The structure of the AP1 coiled−coil is essentially identical in the presence and absence of NFAT1 with the deformation localized to the fork. The junction in Zif23-GCN4 appears to be a natural breakpoint much like the fork in the basic leucine zipper proteins. The structure of the coiled−coil is defined by the complementary knobs-into-holes packing at the hydrophobic surface, and the orientation and register of the DNA-binding domains is templated by multiple interactions with the DNA. These complexes should be relatively rigid. Indeed, comparison of these structural elements with the structures of GCN4 (*22*) and Zif268 (*21*) show that they are essentially unchanged in Zif23-GCN4. Consequently, any torque that is applied to the leucine zipper by crystal packing forces is likely to be dissipated within junction between these domains.

The structure of the selected junction in this Zif23-GCN4 chimera also provides insight into the potential roles of junction residues in other selected Zif23-GCN4 chimeras (*17*). In our previous study, junction sequences for the Zif23-GCN4 chimera were simultaneously, but separately, optimized for the recognition of zinc finger target sequences that overlap by zero, one, or two base paris. Zif23-GCN4 chimeras that recognize a one base pair overlap displayed good DNA-binding specificity and a tight consensus sequence (RFNVHxKK) of fixed length. Using the current Zif23-GCN4 structure as a template, we can predict the role of the junction residues in the RFNVHxKK consensus. The key difference in DNA recognition between these two chimeras is the reduction in overlap between the binding sites of neighboring monomers by one base pair. This difference results in a rotation of one monomer relative to the other by ~33° around the helical axis and a displacement of ~3.3 Å along the helical axis. This rotation and displacement increases the gap between the junction of neighboring monomer backbones by ~6 Å (between Cα of Ile53 from opposite monomers) and places the hydrophobic surfaces of the junction and the coiled−coil domains out of alignment. The differences between the sequences of the selected junction for the one base pair and two base pair overlapped sites must compensate for these positional and rotational alterations to form a functional interface. The consensus sequence for the junction selected to recognize the one base pair gap between the zinc finger target sequences (RFNVHxKK) contains one fewer amino acid than the junction selected for recognition of the two base pair gap (RDIQHILPI in this structure). This reduction, along with the absence of the kink introduced by Pro58, could rotate the hydrophobic surfaces of the coiled−coils back into the proper register. The substitution of Asp52 with Phe, coupled with a slight rotation of the zinc finger helix in the major groove (or a slight change in the DNA conformation), could potentially fill the larger gap between monomers generated by the reduced overlap between neighboring subsites. This would position the Ile53Asn substitution to make the phosphate contact observed from Gln54 in the current structure (Figure 2). Val54 in the consensus sequence would be expected to pack into the cleft formed by His50 and His54, as a β-branched amino acid is commonly found prior to the final histidine involved in zinc coordination in monomeric zinc finger proteins. Finally, the two conserved lysines at the end of the RFNVHxKK consensus sequence would be positioned to make phosphate contacts.

Enhancing the DNA-Binding Specificity of the Zif23-GCN4 Chimera. The Zif23-GCN4 structure provides a basis for engineering additional improvements into the DNA-binding specificity of this dimer. The Zif23-GCN4 dimer already binds with a higher degree of sequence specificity than Zif268, while recognizing a site of similar length (*17*). This has been accomplished by overlapping four fingers in a binding site that is normally recognized by three fingers in a canonical zinc finger protein. Within the region where the DNA half-sites overlap, Zif23-GCN4 has a higher density of potential DNA recognition residues per base pair than Zif268 (Figure 4a). Using Zif23-GCN4 as a scaffold, it should be possible to select amino acids within the overlapped region that make full use of its potential for enhanced DNA-binding specificity. The prospects for selecting dimers with exquisite specificity appear even more promising in the context of a heterodimer, where complementary, but not necessarily identical, residues could be used for DNA recognition at the dimer interface. We are currently exploring

this possibility in constructs where the GCN4 leucine zipper has been replaced with the c-Fos and c-Jun leucine zippers (*17*). The selection of a heterodimeric junction sequence could also be used to assist in the formation of obligate heterodimeric complexes.

The Zif23-GCN4 dimer also displays a level of discrimination similar to Zif268 for a single base pair substitution at the center of its binding site (*17*, *36*). Because Arg49 seems underutilized with regards to sequence-specific DNA recognition, there appears to be room for additional improvement in sequence-specific DNA recognition. The Zif23-GCN4 dimer may even provide a system for examining the limits of single base pair discrimination for DNA recognition in the major groove. Exquisite specificity for a single base pair change could be useful for discriminating between two alleles of a gene that differ by a single nucleotide polymorphism or a single base pair mutation.

Implications for Structure-Based Design of Chimeric Proteins. The Zif23-GCN4 structure demonstrates the advantages of our two-step approach for creating proteins with novel functions. Structure-based design provides a framework for assembling motifs with a desired function. This is followed by selection (via phage display in these studies) to optimize the function of the new chimeric protein. The importance of selecting the junction sequence, as opposed to using a flexible linker, is evident in the dramatic improvement in the affinity (250-fold) and DNA-binding specificity (38-fold improvement in the specificity ratio) observed for the Zif23-GCN4 chimera with the selected junction as compared with the original chimera that had contained a designed linker (*17*).

The creation of novel proteins through the combination of design and selection is obviously not limited to proteins containing zinc fingers or those involved in DNA recognition. This approach could be used to create hybrid proteins in any case where the structures of both parent proteins are known and in which a selection protocol can be devised. New data from the rapidly expanding field of bioinformatics and growth in the number of known structures (including those from homology-based modeling) will surely provide fuel for our approach. Moreover, improved selection methods (*6*) are being developed that will reduce the amount of time required to optimize any given design. Creative use of these new data and technologies, in combination with structure-based design and selection, should facilitate more efficient strategies for engineering novel hybrid proteins tailored to perform specific tasks.

CONCLUSIONS

The structure of this dimeric protein reveals that the zinc fingers in the context of the engineered dimer dock in the major groove of the DNA in a manner that is essentially identical to a canonical zinc finger protein. This implies that zinc fingers in this chimera, like naturally occurring zinc fingers, should be able to recognize a wide variety of DNA sequences. This system also has the advantages of cooperative DNA recognition. Therefore, this zinc finger dimerization system may represent an especially powerful method for artificially regulating a target gene. The success of these experiments illustrates how a combination of structure-based design and selection can produce a novel protein that juxtaposes structural features not previously observed in natural systems. Using this general approach, it should be possible to create a broad range of chimeric proteins for applications in biochemical research or in the treatment of disease.

ACKNOWLEDGMENT

We would like to thank E. Ramm, B. Wang, J. K. Joung, J. Miller, E. Peisach, and C. Ogata for their helpful advice and assistance. Research was carried out in part at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the U.S. Department of Energy, Division of Material Sciences and Division of Chemical Sciences.

REFERENCES

1. Wolfe, S. A., Nekludova, L., and Pabo, C. O. (2000) *Annu. Rev. Biophys. Biomol. Struct. 29*, 183−212.
2. Pabo, C. O., Peisach, E., and Grant, R. A. (2001) *Annu. Rev. Biochem. 70*, 313−40.
3. Beerli, R. R., and Barbas, C. F., III (2002) *Nat. Biotechnol. 20*, 135−41.
4. Greisman, H. A., and Pabo, C. O. (1997) *Science 275*, 657−61.
5. Segal, D. J., Dreier, B., Beerli, R. R., and Barbas, C. F., III (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 2758−63.
6. Joung, J. K., Ramm, E. I., and Pabo, C. O. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 7382−7.
7. Kim, J. S., and Pabo, C. O. (1997) *J. Biol. Chem. 272*, 29795−800.
8. Beerli, R. R., Segal, D. J., Dreier, B., and Barbas, C. F., III (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 14628−33.
9. Zhang, L., Spratt, S. K., Liu, Q., Johnstone, B., Qi, H., Raschke, E. E., Jamieson, A. C., Rebar, E. J., Wolffe, A. P., and Case, C. C. (2000) *J. Biol. Chem. 275*, 33850−60.
10. Rebar, E. J., Huang, Y., Hickey, R., Nath, A. K., Meoli, D., Nath, S., Chen, B., Xu, L., Liang, Y., Jamieson, A. C., Zhang, L., Spratt, S. K., Case, C. C., Wolffe, A., and Giordano, F. J. (2002) *Nat. Med. 8*, 1427−32.
11. Liu, Q., Segal, D. J., Ghiara, J. B., and Barbas, C. F., III (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 5525−30.
12. Kim, J. S., and Pabo, C. O. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 2812−7.
13. Kamiuchi, T., Abe, E., Imanishi, M., Kaji, T., Nagaoka, M., and Sugiura, Y. (1998) *Biochemistry 37*, 13827−34.
14. Moore, M., Klug, A., and Choo, Y. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 1437−41.
15. Pomerantz, J. L., Wolfe, S. A., and Pabo, C. O. (1998) *Biochemistry 37*, 965−70.
16. Wang, B., and Pabo, C. (1999) *Proc. Natl. Acad. Sci. 96*, 9568−73.
17. Wolfe, S. A., Ramm, E. I., and Pabo, C. O. (2000) *Structure 8*, 739−50.
18. Ptashne, M. (1992) *A genetic switch*, 2nd ed., Blackwell Scientific Publications and Cell Press, Cambridge, MA.
19. Kohler, J. J., and Schepartz, A. (2001) *Bioorg. Med. Chem. 9*, 2435−43.
20. Kohler, J. J., Metallo, S. J., Schneider, T. L., and Schepartz, A. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 11735−9.
21. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure 4*, 1171−80.
22. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell 71*, 1223−37.
23. Elrod-Erickson, M., Benson, T. E., and Pabo, C. O. (1998) *Structure 6*, 451−64.
24. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell 77*, 21−32.
25. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−26.
26. CCP4 (1994) *Acta Crystallogr., Sect. D 50*, 760−3.
27. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−9.

28. Brünger, A. T. (1992) *X-PLOR Version 3.1: A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.
29. Lu, X. J., and Olson, W. K. (2003) *Nucleic Acids Res. 31*, 5108−21.
30. Ellenberger, T. E., Brandl, C. S., Struhl, K., and Harrison, S. C. (1992) *Cell 71*, 1223−37.
31. Nekludova, L., and Pabo, C. O. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 6948−52.
32. Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C., and Berman, H. M. (2001) *J. Mol. Biol. 313*, 229−37.
33. Wang, B. S., Grant, R. A., and Pabo, C. O. (2001) *Nat. Struct. Biol. 8*, 589−93.
34. Glover, J. N. M., and Harrison, S. C. (1995) *Nature 373*, 257−61.
35. Chen, L., Glover, J. N., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) *Nature 392*, 42−8.
36. Elrod-Erickson, M., and Pabo, C. O. (1999) *J. Biol. Chem. 274*, 19281−5.
37. Wolfe, S. A., Grant, R. A., Elrod-Erickson, M., and Pabo, C. O. (2001) *Structure 10*, 1−15.

BI034830B