BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2009,** *48,* 4431–4439 4431
DOI:10.1021/bi9001997

The Tandem Zinc-Finger Region of Human ZHX Adopts a Novel C2H2 Zinc Finger Structure with a C-Terminal Extension[†,‡]

Hans Wienk,[§] Ivonne Lammers,[§] Anna Hotze,[§] Jin Wu,[§] Rainer W. Wechselberger,[§] Ray Owens,[∥] David K. Stammers,[∥] David Stuart,[∥] Robert Kaptein,[§] and Gert E. Folkers[§,*]

[§]*Bijvoet Center for Biomolecular Research, NMR Spectroscopy Research Group, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, and* [∥]*Oxford Protein Production Facility and Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Headington, Oxford OX3 7BN, United Kingdom*

Received February 6, 2009; Revised Manuscript Received April 2, 2009

ABSTRACT: Binding of the nuclear factor-Y complex (NF-Y) to the inverted CCAAT-box interferes with transcription activation through nucleosome reorganization. The three homologous proteins forming the zinc-fingers and homeoboxes (ZHX) family interact with the activation domain of NF-Ya to repress transcription. Each ZHX-protein contains two generic C2H2 zinc-fingers (ZNF1 and ZNF2) followed by five homeodomains. Although the proteins have been related to the occurrence of certain cancers, the function and structure of the individual ZHX domains are still unknown. Here, we determined the structure of the tandem zinc-finger region of human ZHX1. Folding and secondary structure predictions combined with expression screening revealed that the C-terminal extension (E) to ZNF2 could form a single domain with the two hZHX1 zinc-fingers. We therefore decided to determine the solution structure of the zinc-fingers followed by this extension. We show that both zinc-fingers adopt canonical $\beta\beta\alpha$-folds in which a zinc ion is coordinated by two cysteine and two histidine residues. The C-terminal extension to ZNF2 forms two β-strands to make a β-sheet with the β-strands of this zinc-finger. The ZNF1 and ZNF2-E domains do not show evident contacts and their mutual orientation seems variable. The high degree of sequence conservation among ZHX family members permitted us to prepare homology models for ZHX2 and ZHX3, revealing distinct surface characteristics for each family member. Implications of these structural features for ZHX-functioning in transcription regulation are discussed.

Although all cells contain identical chromosomes, cell types of multicellular organisms differ in structure and function. Differentiation is caused by the regulated transcription of distinct parts of the genome resulting in the production of different sets of proteins during cell growth.

The CCAAT-box is a common element in eukaryotic promoters that is regularly found as a single copy located 60−100 base pairs 5′ from the transcription initiation site (*1*) either in its forward or in its inverted form. The latter (5′-ATTGG-3′) is known as the Y-box (*2*) and is recognized by the ubiquitous transcription activator nuclear factor-Y (NF-Y).[1] Active NF-Y is a heterotrimer consisting of NF-Ya, NF-Yb, and NF-Yc subunits. NF-Yb and NF-Yc form a tight dimer before NF-Ya associates and interactions with the nucleosome structure take place. Here, high-affinity and sequence-specific interactions with the Y-box are mediated by NF-Ya (*3*), while the subunits NF-Yb and NF-Yc, structurally resembling the core histones H2B and H2A, interact with the core histones H3 and H4 in the nucleosome (*4*). The simultaneous interaction of NF-Y with the Y-box and the histone H3−H4 dimer causes disruption of the nucleosome by dissociating the DNA from the histone core toward the transcription site. This renders the gene accessible, thereby permitting transcription.

NF-Y is known to play an important role in the expression of developmental, tissue-specific, and cell-cycle regulated genes, and is down-regulated by interactions with transcription repressors (*3*). These repressors either hamper the formation of the NF-Y complex or influence the binding of the NF-Y complex to the Y-box or to histones H3 and H4. One group of NF-Y interacting repressors is formed by the protein family zinc-fingers and homeoboxes (ZHX). This family, found in bony vertebrates and clawed frogs, belongs to the ZNF-class of the homeobox superfamily (*5*, *6*) and consists of three highly similar members that can form homo- or heterodimers with distinct functional properties (*5*−*9*). Previously, the ZHX-proteins have been implicated in podocyte disease since deregulated ZHX expression leads to the change in expression profile for various genes involved in podocyte differentiation and focal sclerosis (*10*, *11*). Also, abnormal expression levels of hZHX2 were shown to be related to multiple myeloma, a plasma cell cancer that manifests

[†]This work was supported by the Center for Biomedical Genetics, the EU Structural Proteomics in Europe (SPINE) program (QLG2-CT-2002-00988), and SPINE2-complexes (Contract number 031220). This work was carried out under the auspices of European Large Scale Facility Utrecht (EU-NMR, contract # RII3-026145).

[‡]Backbone and side-chain assignments are deposited in the BioMagResBank (BMRB-entry 7034). Coordinates of the NMR structure ensemble are deposited in the Protein Data Bank (PDB-entry 2GHF).

*To whom correspondence should be addressed. Phone: +31 30 253 9930. Fax: +31 30 253 7623. E-mail: g.e.folkers@uu.nl.

[1]Abbreviations: E, C-terminal extension to the zinc-fingers of hZHX1; NF-Y, nuclear factor-Y; NMR, nuclear magnetic resonance; ZHX, zinc-fingers and homeoboxes; ZNF, zinc-finger.

Published on Web 4/6/2009

itself in the bone (*12*). In line with this, gene expression profiling experiments show regulated expression of ZHX2 during B-cell development (*13*).

Despite their clinical importance, the mode of action of the ZHX-proteins is poorly understood. Amino acid sequence analysis revealed the presence of several conserved domains (Figure 1A). All ZHX-proteins contain two C2H2 zinc-fingers (ZNF) and five homeodomains (HD) (*15*), which are common motifs among transcription factors. Additionally, there is an arginine-rich domain (RR) and a C-terminal acidic amino acid

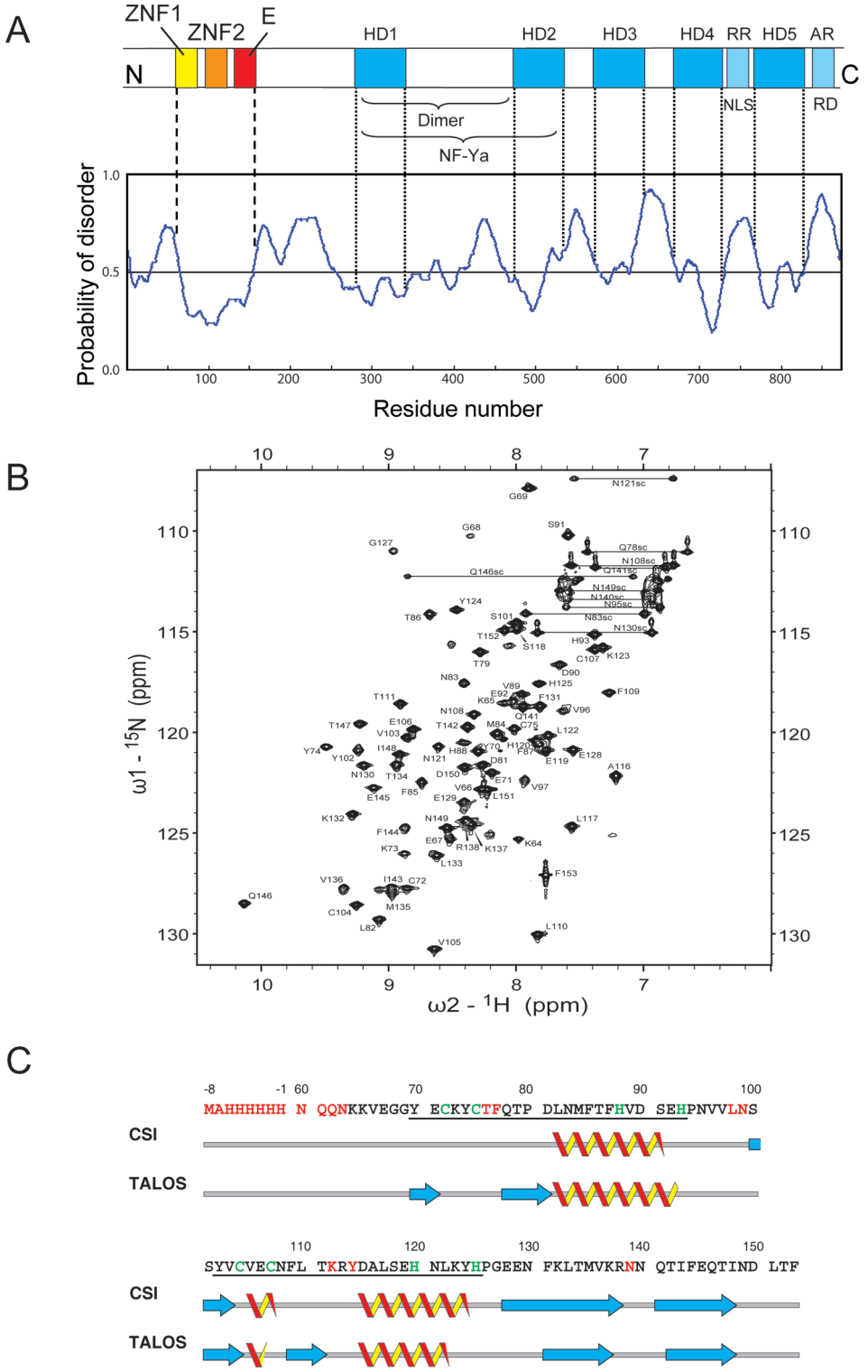

FIGURE 1: (A) Domain organization of ZHX-proteins. Indicated are two zinc-fingers (ZNF1 and ZNF2), the C-terminal extension to ZNF2 (E), five homeodomains (HD1−HD5), an arginine-rich (RR) nuclear localization signal (NLS), and an acidic region (AR) that functions as a repressor domain (RD). The regions providing dimerization and interactions with NF-Ya are indicated. Below, the probability of disorder determined using the RONN program (*14*) is presented as a function of hZHX1 residue number. (B) Assigned [^{15}N;^{1}H]-HSQC spectrum of the ZHX1-ZNF domain. (C) Amino acid sequence of the His-tagged ZHX1 ZNF-region N60−F153 with the secondary structure elements determined from chemical shift values by CSI and TALOS. Underlined are the two zinc-finger regions ZNF1 and ZNF2. In green, zinc-coordinating residues are shown; in red, amino acids for which no backbone amide assignment was obtained are indicated. Helices are colored red/yellow and β-strands blue.

rich (AR) region. Although it is known that homeodomains and C2H2 zinc-finger domains are often involved in DNA binding, so far this has not been established for any of the ZHX-domains. Instead, using the yeast two-hybrid approach, intact ZHX1 was shown to interact with proteins. Among others these include (i) UAP56 (BAT1), which is important for mRNA splicing and export, and is involved in immunopathological disorders (*16*), (ii) casein kinase 2 α1 polypeptide (*17*), a component of the *Wnt* signaling pathway involved in the regulation of important cellular processes such as transcription, and (iii) the activating transcription factor interacting protein ATF-IP and the corepressor BS69 (*8*, *18*). The latter has recently also been linked to chromatin remodeling (*19*).

The NF-Ya interaction domain of ZHX has been localized to the first two homeodomains (*5*, *8*) (Figure 1A). The ZHX-dimerization domain overlaps with this NF-Ya interaction domain and comprises HD1 and the linker to HD2 (*6*). For ZHX1, the RR-region between HD4 and HD5 was proposed to be a nuclear localization signal (NLS (*6*)), while ZHX3 contains a nuclear localization signal in the N-terminal region including the first zinc-finger (ZNF1) (*8*). The C-terminal AR-domain is also called the repression domain (RD) because it is involved in the transcriptional repressor activity of ZHX (*6*). Finally, recently ZHX1 was found to interact with the BS69 MynD domain using a PxLxP motif located N-terminal to the zinc-finger region (*18*).

To understand the ZHX-dependent NF-Y repression and how different domains contribute to quench transcription, we started to analyze ZHX in a divide-and-conquer approach. As part of this, the three-dimensional solution structure of the ZHX1 zinc-finger region was determined by NMR spectroscopy. Because the hZHX1 zinc-finger region has a reduced probability of disorder extending C-terminally from the two zinc-fingers (Figure 1A) as determined using the RONN disorder prediction tool (*14*), we reasoned that this region (E) might contain structural features that could participate in the overall fold of the ZHX1 zinc-finger region. Here, we show that ZNF1 and ZNF2 obtain their consensus ββα-folds. Remarkably, the ZHX1 zinc-finger region together with the C-terminal extension shows an unprecedented zinc-finger feature in that the extension to ZNF2 forms a four-stranded sheet with the β-strands of ZNF2. The obtained zinc-finger structure and the homology models for the ZHX family members derived from the ZHX1 structure suggest a similar function for all three family members. However, on the basis of sequence conservation and surface properties we assume that the three proteins possess distinct substrate specificity.

MATERIALS AND METHODS

Sample Preparation. On the basis of disorder predictions ((*14*) Figure 1A), we identified the domain boundaries for the zinc-finger region of ZHX1 N60−F153. The corresponding fragment (i.e., including the C-terminal extension to the second zinc-finger) was PCR amplified and cloned in the expression vector pDEST-14 (Invitrogen), which also contained an N-terminal His_6-tag (residues −8 to −1) to facilitate protein isolation (Bird, L. E., and co-workers, unpublished results). The resulting plasmid, encoding a polypeptide of 102 residues (M_w = 12.1 kDa; pI = 6.33), was transformed to *Escherichia coli* BL21 derivative Rosetta (Novagen) and cultured at 37 °C in 500 mL of modified M9 medium containing 25 μM $ZnCl_2$, 50 μg/mL ampicillin, and 35 μg/mL chloramphenicol essentially as described before (*20*). For isotope labeling, $^{15}NH_4Cl$ (1 gr/L) and [^{13}C]-glucose (3 gr/L) were used. When cultures reached OD_{600} of 0.6−0.7, isopropyl-1-thio-β-D-galactopyranoside was added to a final concentration of 500 μM to induce protein production. At $OD_{600} \approx 2$, the bacteria were harvested and resuspended in 25 mL of 50 mM sodium phosphate at pH 8.0 containing 300 mM NaCl, 10 mM imidazole, 0.2% (v/v) Triton X-100, 1 mM 2-mercaptoethanol, 1 mM phenyl-methyl-sulfonyl fluoride, 25 μM $ZnCl_2$, and EDTA-free protease inhibitor mixture for His-tag proteins (Sigma). After freeze−thawing, mild sonication on ice, and centrifugation, the cell lysate was filtered and purified by nickel affinity chromatography (*20*). Fractions containing the target protein were concentrated by ultrafiltration using a 3-kDa cutoff membrane (Amicon, Millipore) and purified to homogeneity by gel filtration (Superdex G75 column, GE Healthcare). After buffer exchange and concentration, 500 μL NMR samples contained approximately 0.3 mM hZHX1-ZNF in 50 mM NaP_i at pH 8.0, 150 mM NaCl, 40 μM $ZnCl_2$, and 10% D_2O. EDTA-free protease inhibitor cocktail (Boehringer) and 0.05% (w/v) NaN_3 were added for stabilization.

NMR Experiments and Structure Calculations. NMR experiments were conducted at 293 K on Bruker Avance 700-MHz and 900-MHz spectrometers equipped with triple resonance probes, processed using NMRPipe (*21*), and analyzed with SPARKY (*22*). Resonance assignments were obtained from 700-MHz [^{15}N;^{1}H]-HSQC, [^{13}C;^{1}H]-HSQC, CBCA(CO)NH, HNCACB, HN(CA)HA, HNCO, CC(CO)NH, H(C)CH-TOCSY, and (H)CCH-COSY spectra. On the basis of these assignments, the programs CSI (*23*) and TALOS (*24*) were used to analyze secondary structure elements, which also resulted in 92 and 100 dihedral angle restraints, respectively. Distance restraints were obtained from 900-MHz 2D [^{1}H;^{1}H]-NOESY and 3D NOESY-[^{15}N;^{1}H]-HSQC experiments. On the basis of existing zinc-finger structures, unambiguous restraints were implemented to enforce zinc-coordination by cysteine and histidine residues: (i) upper limits for Cys-Sγ−Zn^{2+} and His-Nε2−Zn^{2+} distances were kept at 2.6 and 2.3 Å, respectively; (ii) tetrahedral zinc-coordination was imposed using distances between 3.9 and 3.5 Å, between 3.7 and 3.3 Å, and between 3.3 and 2.9 Å, for Cys-Sγ−Cys-Sγ, Cys-Sγ−His-Nε2, and His-Nε2−His-Nε2, respectively.

hZHX1-ZNF structure calculations were performed in three steps without any manual assignment of NOESY signals. (i) An initial fold was obtained with CYANA 2.1 (*25*, *26*) using eight iterations of automatic assignment and structure calculations using the aforementioned restraints and 10,000 time-steps of simulated annealing. Proton and nitrogen assignment tolerances were kept low, at 0.010 and 0.175 ppm, respectively. In this first step, only the 70% most intense NOESY signals were used. During each iteration, 100 structures were calculated, of which the 20 with lowest energy were used in the next step. (ii) The ensemble resulting from step (i) was used as input for a second CYANA run in which all NOESY signals were used. Still, a relatively low percentage of the signals was used to obtain the final structure, mainly due to the low tolerances used for the automatic assignment procedure and because peak-picking levels were low enough to enclose noise. A total of approximately 4500 automatically assigned signals yielded over 2000 NOEs (Supporting Information). (iii) Water refinement was performed after transforming the final CYANA-restraints to CNS format (*27*) and a single iteration of structure calculation applying Cartesian dynamics. For this and the structure validation with PROCHECK (*28*) and WHATCHECK (*29*), the existing RECOORD

protocols (*30*) were adapted to account for the presence of zinc (II) ions and unprotonated cysteine-Sγ atoms.

Homology models for the hZHX2 and hZHX3 zinc-finger regions were calculated with MODELER-8v2 (*31*). Protein structures were visualized by MOLMOL-2k (*32*) or PyMOL (*33*).

RESULTS AND DISCUSSION

Zinc-fingers come in several types that are named after the number of cysteines and histidines that coordinate a zinc(II) ion, e.g., C2H2 (CCHH), CCHC, and CCCC (for overviews, see, e.g., ref (*34*−*36*)). The C2H2 zinc-fingers are most common. They are about 30 residues long and their primary sequences share the pattern x_2−C−$x_{2,4}$−C−x_{12}−H−x_{3-5}−H, where x can be any amino acid. The spacing between the two cysteines and between the two histidines is somewhat variable; the spacing between the second cysteine and the first histidine is invariably 12 residues. Often, the C2H2 zinc-finger sequence starts with a tyrosine, while phenylalanine is frequently found two or four residues after the second cysteine. Three residues before the first zinc-coordinating histidine phenylalanine, tyrosine or leucine are often found.

C2H2 zinc-fingers fold into a $\beta\beta\alpha$-motif, with a two-stranded antiparallel β-sheet pinned to a C-terminal α-helix using the zinc ion, which gets coordinated tetrahedrally by two cysteines near the loop between the β-strands and two histidines located at the C-terminus of the α-helix. Additional stabilization of the fold is provided by van der Waals contacts between the β-sheet and the α-helix. The conserved aromatic amino acids between the second cysteine and first histidine are embedded in the hydrophobic interior of the zinc-finger. Most of these general C2H2 properties are recognized twice in the three ZHX proteins (Figure 2A). Overall, the zinc-finger regions of ZHX are largely invariable, showing significantly higher conservation than the surrounding regions. Human ZHX1 and ZHX2 share 61% identity and 90% homology, while hZHX1-ZNF and hZHX3-ZNF are somewhat less related with 37% identity and 74% homology.

Assignment and Secondary Structure of hZHX1-ZNF. Standard NMR experiments led to the assignment of nitrogens, carbons, and 78% of all protons of isolated hZHX1-ZNF. Apart from the N-terminal His_6-tag and the first residues of the ZHX1-ZNF region (M-8−N63), which might have undergone efficient NH exchange due to the relatively high pH, most of the protein backbone amide groups were assigned (Figure 1), and for the region K64−F153, the overall proton assignment was 87%. Probably due to line-broadening caused by internal dynamics, the segments T76−Q78, L98−N99, K112−D115, and N139 were incompletely assigned. In addition, several protons of longer side

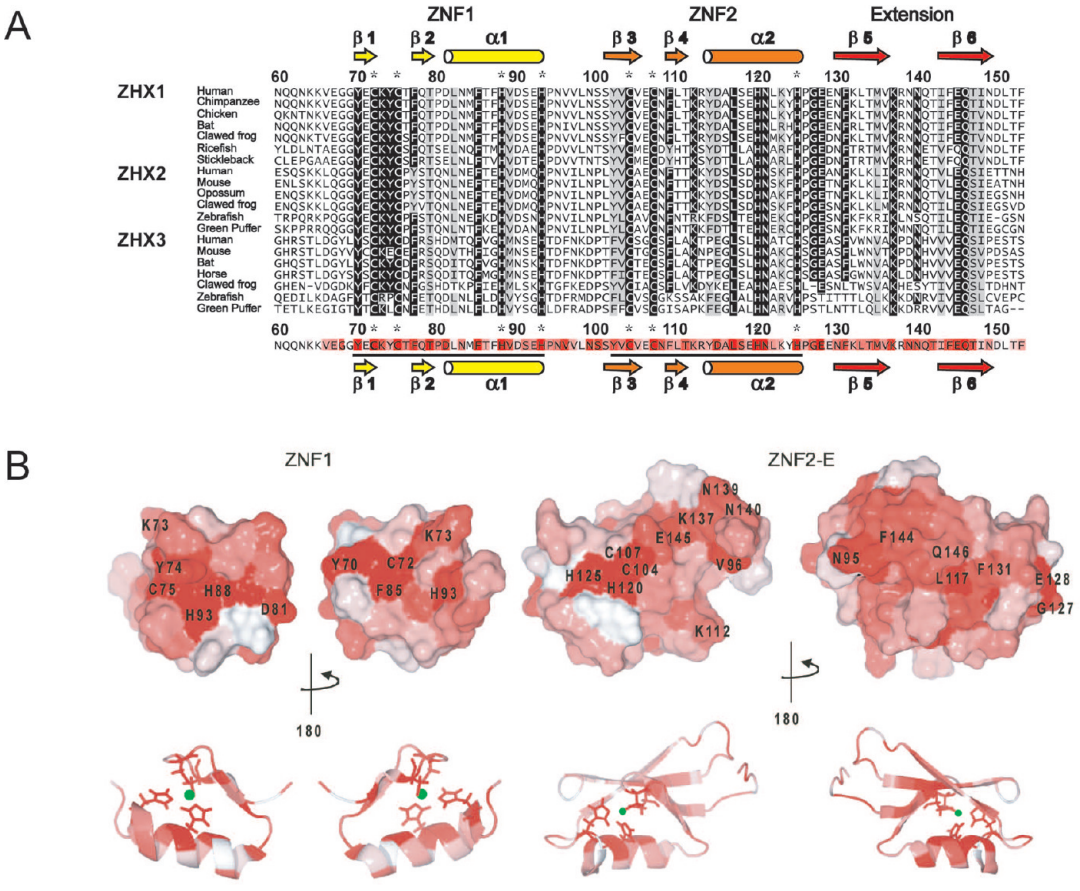

FIGURE 2: (A) Amino acid sequence alignment and secondary structure elements of ZHX-ZNF regions. Amino acid numbering is given for hZHX1 according to UniProtKB/SwissProt-entry Q9UKY1. A representative multiple sequence alignment of the most distantly related ZHX zinc-finger domains obtained from the SMART database (*37*) using ClustalW is shown. Sequence similarity is presented by box-shading using the Boxshade server. Using the rate4site algorithm implemented in the Consurf server (*38*), the conservation score is calculated for all ZHX sequences available taking the phylogenetic relationships between the sequences into account. Most conserved residues are colored in red, least conserved residues in white. Secondary structure elements are indicated above and below the multiple sequence alignment; the zinc-fingers ZNF1 and ZNF2 are underlined. (B) Top panel: surface representation of the sequence conservation for ZNF1 and ZNF2-E as shown in A. The most conserved (partially) surface exposed residues are annotated. Lower panel: ribbon representation of the top panel. The zinc ion is colored green; the zinc-coordinating residues are presented as sticks.

chains as well as the side chains of some phenylalanine and tyrosine residues could not be unambiguously assigned.

Analysis of the collected chemical shift values revealed that both ZHX1 regions that were predicted to be zinc-fingers indeed display the archetypal $\beta\beta\alpha$ secondary structure elements (Figure 1C). This confirmed the isolated protein as a tandem C2H2 zinc-finger in which two zinc(II) ions are coordinated by C72, C75, H88, and H93 (ZNF1) and C104, C107, H120, and H125 (ZNF2). Interestingly, the C-terminal extension to ZNF2 forms two β-strands, which is in agreement with the local order that was predicted for this region (Figure 1A).

Three-Dimensional Structure of hZHX1-ZNF. The final calculated and validated NMR ensemble is presented in Figure 3A (statistics and quality parameters are summarized in the Supporting Information). Because the N-terminal 12 residues including the His-tag (M-8−N63) were unassigned and completely unstructured, only residues K64−F153 are displayed. In agreement with the chemical shift analysis (Figure 1C), two α-helices and six β-strands are present. Both zinc-fingers are folded as typical C2H2 zinc-fingers where the N-terminal β-strands form a small two-stranded antiparallel sheet, which is bound to the C-terminal helix (Figure 3B). The pairwise backbone rmsd of 1.6 Å (MSDfold) indicates that the overall folds of both zinc-fingers are highly similar. Also, within the ensemble, both zinc(II) ions and the coordinating residues are well defined. Each zinc-finger fold is stabilized by internal hydrophobic interactions. For ZNF1, besides the four zinc-coordinating residues, these involve G69, Y70, Y74, T79, L82, F85, T86, and V89. For ZNF2, a hydrophobic core is built up from the residues coordinating zinc complemented by Y102, E106, F109,

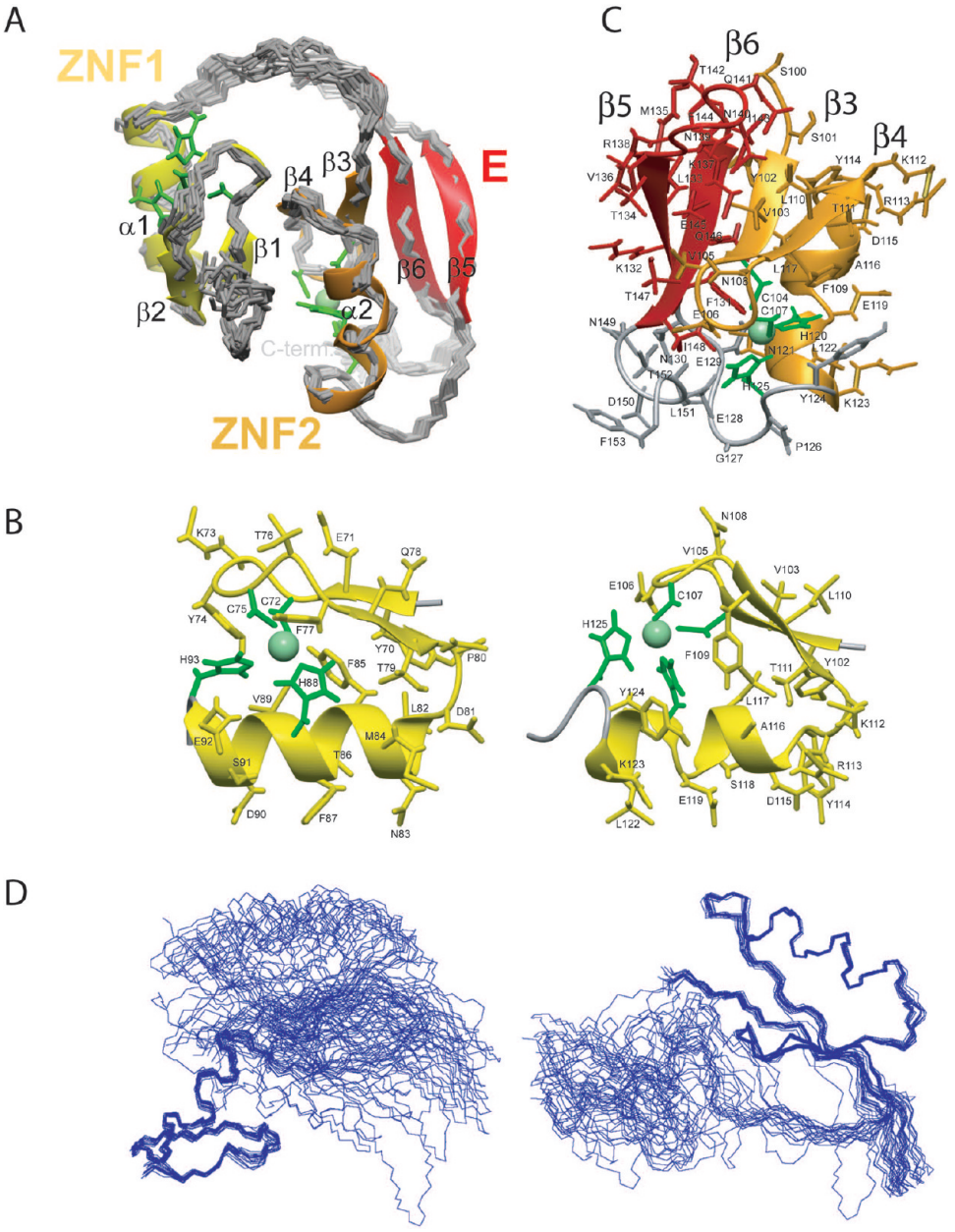

FIGURE 3: (A) Backbone representation of the final hZHX1 ensemble. For the lowest energy structure, secondary structure elements are shown and numbered. The regions ZNF1, ZNF2, and extension E are colored yellow, orange, and red, respectively; zinc ions and zinc-coordinating cysteine and histidine residues are green. (B) Manually aligned ZNF1 (left) and ZNF2 (right); zinc and zinc-coordinating residues are colored green. (C) Amino acid organization of the ZNF2-E domain. ZNF2 is colored orange, extension E red. Zinc and zinc-coordinating residues are colored green. (D) Lack of structure convergence in early CYANA runs. Left panel: fit on ZNF1-backbone. Right panel: fit on ZNF2-backbone.

T111, A116, L117, and Y124. In both zinc-fingers, the van der Waals contacts are clustered predominantly around conserved residues, i.e., Y70, Y74, and F85 in ZNF1, and F109 and L117 in ZNF2 (Figures 2A and 3B).

ZNF2-E Motif. C-terminal to ZNF2, the β-strands β5 and β6 are found antiparallel to each other. The extension unambiguously and exclusively contacts ZNF2 by forming a four-stranded mixed parallel−antiparallel β-sheet together with β3 and β4 (Figure 3C). Several hydrogen bonds are present that associate the parallel strands β3 and β6, which is supported by $H\alpha-H^N$ and H^N-H^N contacts between ZNF2 residues S101−E106 and extension residues I143−I148. The four-stranded β-sheet is twisted, mainly due to contacts between the loop-regions Q141−I143 and S101, between F144 and S100, and between K137 and V103, V105 and N108. Furthermore, the core residues F131 and I148 contribute to the curvature of the β-sheet, possibly enforced by aromatic ring-stacking of Y114, Y102, and F144.

Significant amino acid sequence homology for the C-terminal region beyond ZNF2 is absent in PSIBLAST-searches (*39*, *40*) using either ZHX1 ZNF1-ZNF2-E residues 64−153 or the C-terminal extension alone (residues 126−153). In addition, PROSITE-searches (*41*) using the absolutely conserved ZHX-residues (Figure 2A) did not yield zinc-finger proteins with a C-terminal extension. Finally, no structural homologues were identified with the MSDfold- and NCBI VAST-servers (*42*, *43*) using the structure of the hZHX1 ZNF2-E domain (Y102−D150). Hence, the observed hZHX1 ZNF2-E organization appears to be unique among zinc-finger proteins.

Besides the zinc-coordinating residues, several other residues are strictly conserved in ZHX proteins from different species (Figure 2A). Strikingly, most of the conserved residues that do not participate in the folding of the individual zinc-fingers contribute to the overall organization of the ZNF2-E domain (Figure 3C). First, G127 and E128 form a linker between ZNF2 and the extension that could be important to ensure the curvature of the four-stranded β-sheet by allowing hydrogen bond formation between the glutamate side chain and those of N130 and N149. Second, the side chains of N121, F131, and Q146 are mostly buried within the ZNF2-E domain and contact each other to force the C-terminal extension to approach ZNF2 helix α2. Third, the absolutely conserved extension residues E145 and Q146 connect β-strand β6 to β3 as well as β5 by making contacts with the side chains of V103 and V105 and with F131−K137, respectively. In most of the structures of the ensemble, a salt bridge is formed between E145-$O^{\varepsilon 1}$ and K137-N^{ζ}, as well as between E145-$O^{\varepsilon 2}$ and R138-$N^{\eta 1}$. In addition, some structures show a hydrogen bond between the N140 and K137 side chains. The fact that many interactions between ZNF2 and the C-terminal extension are mediated by absolutely conserved ZHX-residues strongly indicates that the ZNF2-E fold is a crucial element for ZHX functioning.

Mutual hZHX1 Zinc-Finger Orientation. In the literature, only a few structures are described in which tandem zinc-fingers are organized in a well-defined arrangement. For the double zinc-finger from human enhancer protein MBP-1 ((*44*) PDB-entry 1BBO), a head-to-tail arrangement is stabilized by some hydrophobic interactions and a single hydrogen bond. The two zinc-fingers of the Zap1 zinc-responsive domain of *Saccharomyces cerevisiae* ((*45*) PDB-entry 1ZW8) form a more compact solution structure. They contact each other through a network of hydrophobic interactions including direct α-helix contacts, and there are also specific connections between tryptophans that are found between the zinc-coordinating cysteine residues in both zinc-fingers. A similar zinc-finger organization had been described earlier for zinc-fingers 1 and 2 of the sequence-specific DNA-binding GLI protein ((*46*) PDB-entry 1ZAA). Usually, however, tandem zinc-fingers tend to organize as beads-on-a-string in which no direct contacts are formed between different zinc-finger domains (*34*−*36*).

For hZHX1, in most structure calculation runs, the complete ZNF1-ZNF2-E region converged into a single globular organization where β1 and β4 form a unique β-cross element, with the strands neither parallel nor antiparallel but with an angle of 50°−55°. However, structure calculation runs that differed only in the seed of the random number generator showed significant variations in the mutual zinc-finger orientation. In addition, preliminary calculations showed structural divergence originating from the linker connecting the two zinc-fingers (e.g., Figure 3D). This led us to believe that the compactness of the ensemble shown in Figure 3A is artificial. Detailed spectral analysis showed that the only unambiguous long-range contacts between ZNF1 and the rest of the protein connect the Y74 and the finger linker-residue V96 side chains. The direct contacts that do occur between the two zinc-fingers after the final automatic assignment are without exception ambiguous. Further supporting an ill-defined ZNF1−ZNF2 contact surface, a third of the signals expected from the final ensemble to be involved in mutual zinc-finger contacts do not show up at all in the NOESY spectra.

Surface Properties of ZHX-Proteins. To compare the surface properties of the three ZHX-family members, homology models were created for the zinc-finger regions of hZHX2 and hZHX3, using the amino acid sequence alignments (Figure 2A) and the presented lowest energy hZHX1-ZNF structure (Figure 3A). Clearly, besides the amino acid similarity the three family members show conservation in their overall surface charge distribution (Figure 4). Also, we observed that while most of the absolutely conserved residues are buried to ensure optimal folding of the ZNF1 and ZNF2-E domains, several of them are (also) solvent-exposed (Figure 2B). Combined, these surface properties may give clues about the mode of action of the zinc-finger region in ZHX functioning.

Overall, ZNF1 has a neutral to negatively charged surface potential, with the exception of a strictly conserved positively charged residue corresponding to K73 of ZHX1. Sometimes, the residue corresponding to T87 is replaced by a lysine, ZHX3 has occasionally positively charged residues at positions that correspond to ZHX1 Q78, N83, or E92, but these substitutions are not conserved. K73 is part of a conserved region that also includes the aromatic core residues Y70, Y74, and F85 and maybe some of the zinc-coordinating residues (Figures 2B and 4). Possibly this conserved lysine-73, in combination with the flanking hydrophobic surface, is important for physiological binding events. The distribution of the negative charges in ZNF1 is strikingly similar between hZHX1 and hZHX3, with four sites corresponding to hZHX1 E71, D81, D90, and E92. In hZHX2, the negative charge is distributed differently since only two of these four residues are preserved (E71 and D90), while other acidic residues cluster in a region that in most cases is uncharged in ZHX1 and ZHX3. These charged residues show distinct conservations between ZHX2 and ZHX3. While ZHX1 residues E71 and D90 are preserved in ZHX2 but not in ZHX3, the reverse is true for D81 and partly for E92. It is conceivable that the distinct charge distribution contributes to

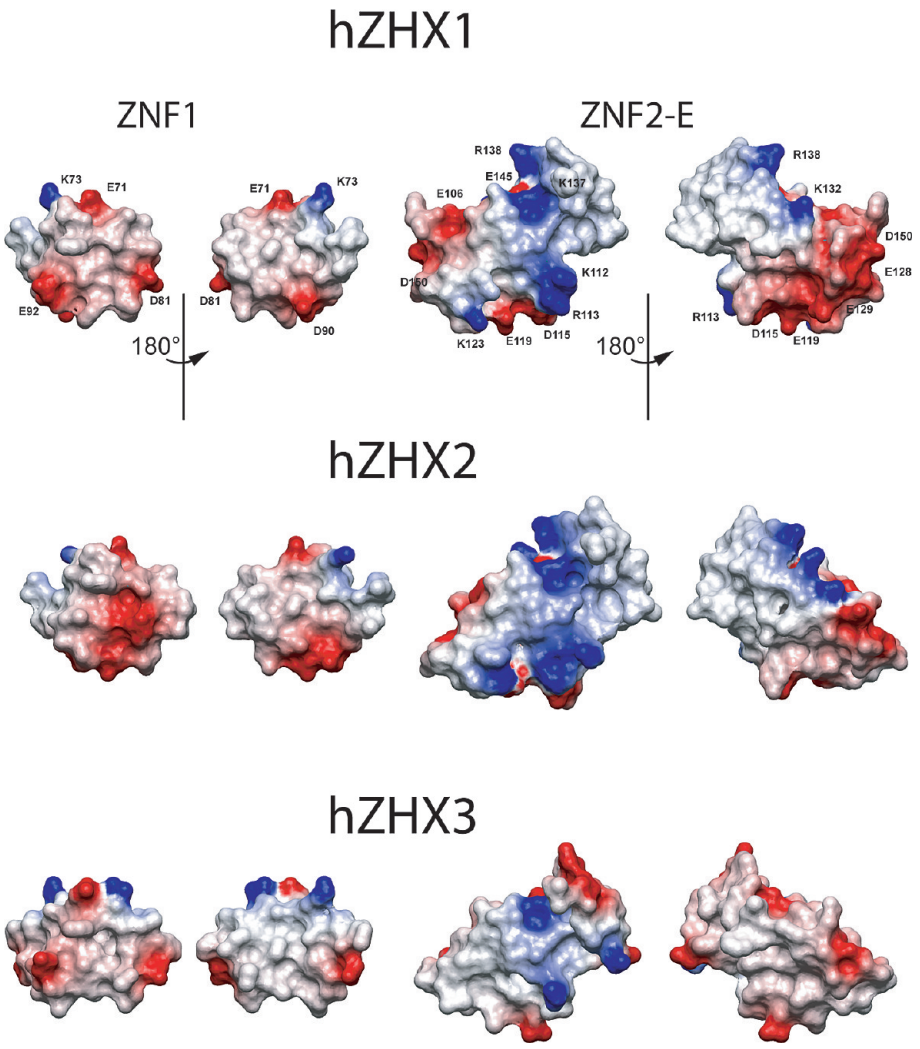

FIGURE 4: Solvent accessible surfaces of ZNF1 (left) and ZNF2-E (right) for the hZHX1 NMR structure (top), and the hZHX2 (middle) and hZHX3 (bottom) hZHX1-derived homology models. Surfaces are colored according to their electrostatic potential; negatively charged regions are red, positively charged regions blue. Exposed charged residues are indicated. Orientations are as described in Figure 2B.

the differential specificity in substrate recognition by the ZHX protein.

On the ZNF2-E module, three conserved surface accessible patches are found (Figure 2B). The first is formed by G127 and E128 and the second by L117, F131, and Q146, maybe reinforced by N121. All of these residues are involved in the folding of the C-terminal extension on ZNF2, which could explain their conservation. However, in all three ZHX-proteins, these residues are also part of a neutral-to-negatively charged surface region, which could be functionally relevant. This extended negatively charged cluster in hZHX1 is less pronounced in hZHX2 and almost completely absent in hZHX3 (Figure 4), which could be due to differences in interaction specificity. On the opposite side of the ZNF2-E module, conserved residues are also clustered. These residues do include not only the zinc-coordinating C107, H120, and H125 but also the K137, N139 (not in ZHX3 though), N140, and E145. The surface charge distribution of this side of the ZHX1 ZNF2-E domain is rather conserved among ZHX2, while ZHX3 seems to contain less positively charged residues on this side of the ZNF2-E module. Most of the conserved residues that do not coordinate zinc concentrate around a cleft that is about 8 Å wide and is lined by the positively charged residues R138, K137, K112, and R113 in ZHX1. In the presented ensemble, this region of the ZNF2-E domain faces the conserved region of ZNF1. This may point toward functional interactions between the two structural domains, where the interfinger linker permits conformational freedom between the two zinc-fingers and can alter the accessibility and/or width of this cleft to regulate biological functions.

Toward hZHX1-ZNF Functionality. Although up to now no experimental information is available about the role of the zinc-finger region in the repression of transcription by ZHX proteins, current results provide a structural framework that may be related to potential functions. Zinc-fingers have been found to contact different biomolecules. While some bind RNA or protein or merely fulfill the role of the spacer module (*47*), mostly the canonical C2H2 zinc-finger binds double-stranded DNA in the major groove (*48*, *49*) using four key residues at the solvent-exposed side of the α-helix (*34*). The presence of improbable key residues (D81, N83, M84, and F87), with varying conservation (Figures 2A and 3B) and in combination with its overall negatively charged surface potential (Figure 4), makes it highly unlikely that hZHX1-ZNF1 binds DNA. ZNF2, however, shows an overall positive surface charge, which appears conserved among all ZHX-proteins. In addition, it has solvent accessible amino acids that could be used as key residues for DNA binding, namely, K112, D115, and A116. On the basis of results from extensive DNA binding studies using mutant C2H2 zing-finger proteins, with some fantasy this domain could recognize the

DNA-sequences 5′-TGG-3′ or 5′-TGT-3′ (*34*). Interestingly, 5′-TGG-3′ is also present in the Y-box, which may well be merely a coincidence but could also indicate a DNA-binding function for ZHX in the repression of NF-Ya mediated transcription.

While the individual ZNF1 and ZNF2 fold as typical C2H2 zinc-fingers, the conservation of many stabilizing residues implies functional relevance of the extension in the ZNF2-E domain. Functionally important C-terminal extensions to zinc-fingers are not unprecedented. For instance, the second C4 zinc-finger of the nuclear hormone receptor RXR contains a C-terminal extension with a helical conformation in solution (*50*). RXR can bind to DNA as a homodimer or as a heterodimer with other family members, and this C-terminal region has been shown to participate in both protein−protein and protein−DNA contacts (*50*) and to contribute to binding specificity and polarity (*51*). Also for C2H2 zinc-fingers, C-terminal extensions have been reported. The human spliceosomal protein U1C harbors a zinc-finger with a disrupted helix that is followed by another helix (*52*). In addition, like ZNF593, a negative modulator for the DNA-binding activity of the Oct-2 transcription factor (*53*) the dsRNA-binding protein ZFa shows a relatively long C-terminal helix-kink-helix region in its second zinc-finger (*54*). Although these C-terminal extensions share an overall positive charge that could be important for their binding to RNA (*52*, *54*) or DNA (*53*), as far as we know their involvement in the interactions with nucleotides has not been firmly established yet. Furthermore, an extra N-terminal β-strand was found in the Swi5 (*55*) and tramtrack (*56*) C2H2 zinc-finger domain structures. Although the functional relevance for the latter extension remains illusive, it was shown for the Swi5 domain that the presence of the additional N-terminal strand increases DNA binding affinity (*55*). Although the structure of these extensions is completely different from ZNF2-E, these findings underscore the possible importance of conserved extensions for zinc-finger function.

Structure similarity searches using ProFunc (*57*) indicate that the ZNF2-E region, with a single helix facing a four-stranded parallel/antiparallel β-sheet, forms a β-grasp fold that resembles the overall organization of ubiquitin. In addition, a DALI structure homology search (*58*) revealed a single significant hit with low structure homology: the immunoglobulin light chain-binding domain of protein L ((*59*), PDB-entry 2PTL). The fact that both do not contain zinc-fingers again suggests different functionality for the elongated second zinc-finger of the ZHX proteins. Nevertheless, two arguments point toward a protein-binding function of the hZHX1 zinc-finger region. First, ZHX has frequently been found in protein-containing complexes using the yeast two-hybrid approach (*8*, *16*, *17*), while its DNA binding has not been described so far. Second, the overall surface charge distribution of ZNF1 does not allow interaction with nucleotides.

Concluding Remarks. The current study yields some fascinating structural properties for the zinc-finger region in ZHX. First, the extension C-terminal to ZNF2 combines into a single structural subunit with this zinc-finger. Second, the overall folding and surface properties of ZHX1, ZHX2, and ZHX3 are highly similar, suggesting that they fulfill similar biological roles. The differences in surface charge distribution suggests that the three ZHX-proteins have different substrate-binding specificities. Finally, although it cannot be excluded that ZNF2-E interacts directly with the Y-box DNA, we assume that it is more likely that both ZHX zinc-finger domains interact with protein. One candidate of interest is BS69, for which it was recently shown that its MynD domain interacts with ZHX1 residues located N-terminally near the zinc-finger region (*17*). This sequence forms a PxLxP motif, which is not present in ZHX2 and ZHX3, agreeing with the BS69 interaction specificity. Possibly, in ZHX2 and ZHX3, at the same region preceding the zinc-finger region, motifs exist that are recognized by other transcriptional corepressors. It could be that the ZHX1 zinc-finger region immediately following this PxLxP domain contributes to functional interactions with BS69, which, like the NF-Y complex, has been implicated in chromatin remodeling. Of course, future biochemical and structural research is imperative to get information on the binding partners and the exact function of the zinc-finger region in ZHX family members. Eventually, this should provide insight into the mode of action of ZHX-dimers in the repression of genomic activity by interfering with NF-Y functioning.

SUPPORTING INFORMATION AVAILABLE

Statistics for the solution structure calculation of the NMR ensemble of the ZHX1 zinc-finger region. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol. 212*, 563–578.
2. Mantovani, R. (1998) A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res. 26*, 1135–1143.
3. Mantovani, R. (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene 239*, 15–27.
4. Baxevanis, A. D., Arents, G., Moudrianakis, E. N., and Landsman, D. (1995) A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res. 23*, 2685–2691.
5. Kawata, H., Yamada, K., Shou, Z., Mizutani, T., Yazawa, T., Yoshino, M., Sekiguchi, T., Kajitani, T., and Miyamoto, K. (2003) Zinc-fingers and homeoboxes (ZHX) 2, a novel member of the ZHX family, functions as a transcriptional repressor. *Biochem. J. 373*, 747–757.
6. Yamada, K., Kawata, H., Matsuura, K., Shou, Z., Hirano, S., Mizutani, T., Yazawa, T., Yoshino, M., Sekiguchi, T., Kajitani, T., and Miyamoto, K. (2002) Functional analysis and the molecular dissection of zinc-fingers and homeoboxes 1 (ZHX1). *Biochem. Biophys. Res. Commun. 297*, 368–374.
7. Hirano, S., Yamada, K., Kawata, H., Shou, Z., Mizutani, T., Yazawa, T., Kajitani, T., Sekiguchi, T., Yoshino, M., Shigematsu, Y., Mayumi, M., and Miyamoto, K. (2002) Rat zinc-fingers and homeoboxes 1 (ZHX1), a nuclear factor-YA-interacting nuclear protein, forms a homodimer. *Gene 290*, 107–114.
8. Yamada, K., Kawata, H., Shou, Z., Hirano, S., Mizutani, T., Yazawa, T., Sekiguchi, T., Yoshino, M., Kajitani, T., and Miyamoto, K. (2003) Analysis of zinc-fingers and homeoboxes (ZHX)-1-interacting proteins: molecular cloning and characterization of a member of the ZHX family, ZHX3. *Biochem. J. 373*, 167–178.
9. Kawata, H., Yamada, K., Shou, Z., Mizutani, T., and Miyamoto, K. (2003) The mouse zinc-fingers and homeoboxes (ZHX) family; ZHX2 forms a heterodimer with ZHX3. *Gene 323*, 133–140.
10. Liu, G., Clement, L. C., Kanwar, Y. S., Avila-Casado, C., and Chugh, S. S. (2006) ZHX proteins regulate podocyte gene expression during the development of nephrotic syndrome. *J. Biol. Chem. 281*, 39681–39692.
11. Chugh, S. S. (2007) Transcriptional regulation of podocyte disease. *Transl. Res. 149*, 237–242.
12. Harousseau, J.-L., Shaughnessy, J.Jr., and Richardson, P. (2004) Multiple myeloma. *Hematology ASH Educational Book* 237–256.
13. Hystad, M. E., Myklebust, J. H., Bø, T. H., Sivertsen, E. A., Rian, E., Forfang, L., Munthe, E., Rosenwald, A., Chiorazzi, M., Jonassen, I., Staudt, L. M., and Smeland, E. B. (2007) Characterization of early stages of human B cell development by gene expression profiling. *J. Immunol. 179*, 3662–3671.
14. Yang, Z. R., Thomson, R., McNeil, P., and Esnouf, R. M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics 21*, 3369–3376.

15. Barthelemy, I., Carramolino, L., Gutiérrez, J., Barbero, J. L., Márquez, G., and Angel Zaballos, A. (1996) *Zhx-1*: a novel mouse homeodomain protein containing two zinc-fingers and five homeodomains. *Biochem. Biophys. Res. Commun. 224*, 870–876.
16. Lehner, B., Semple, J. I., Brown, S. E., Counsell, D., Campbell, D., and Sanderson, C. M. (2004) Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics 83*, 153–167.
17. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell 122*, 957–968.
18. Ogata-Kawata, H., Yamada, K., Uesaka-Yoshino, M., Kagawa, N., and Miyamoto, K. (2007) BS69, a corepressor interacting with ZHX1, is a bifunctional transcription factor. *Front. Biosci. 12*, 1911–1926.
19. Velasco, G., Grkovic, S., and Ansieau, S. (2006) New insights into BS69 functions. *J. Biol. Chem. 281*, 16546–16550.
20. Folkers, G. E., van Buuren, B. N., and Kaptein, R. (2004) Expression screening, protein purification and NMR analysis of human protein domains for structural genomics. *J. Struct. Funct. Genomics 5*, 119–131.
21. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.
22. Goddard, T. D., and Kneller, D. G., *SPARKY 3*, University of California, San Francisco, CA.
23. Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR 4*, 171–180.
24. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR 13*, 289–302.
25. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol. 273*, 283–298.
26. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol. 319*, 209–227.
27. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D 54*, 905–921.
28. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283–291.
29. Hooft, R. W. W., Vriend, G., Sander, C., and Abola, E. E. (1996) Errors in protein structures. *Nature (London) 381*, 272–272.
30. Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A. E. M., Nabuurs, S. B., Güntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R., and Bonvin, A. M. J. J. (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins 59*, 662–672.
31. Marti-Renom, M. A., Stuart, A., Fiser, A., Sánchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct. 29*, 291–325.
32. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics 14*, 51−55; 29−32.
33. DeLano, W. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA.
34. Wolfe, S. A., Nekludova, L., and Pabo, C. O. (1999) DNA recognition by Cys_2His_2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct. 3*, 183–212.
35. Pabo, C. O., Peisach, E., and Grant, R. A. (2001) Design and selection of novel Cys_2His_2 zinc finger proteins. *Annu. Rev. Biochem. 70*, 313–340.
36. Folkers, G. E., Hanzawa, H., and Boelens, R. (2001) Zinc Finger Proteins, in *Handbook on Metalloproteins* (Bertini, I., Sigel, A., and Sigel, H., Eds.) pp 961−1000, M. Dekker, New York.
37. Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J., and Bork, P. (2006) SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res. 34*, D257–D260.
38. Mayrose, I., Graur, D., Ben-Tal, N., and Pupko, T. (2004) Comparison of site-specific rate-inference methods: Bayesian methods are superior. *Mol. Biol. Evol. 21*, 1781–1791.
39. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389–3402.
40. Schäffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V., and Altschul, S. F. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res. 29*, 2994–3005.
41. Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M., and Sigrist, C. J. A. (2006) The PROSITE database. *Nucleic Acids Res. 34*, D227–D230.
42. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D 60*, 2256–2268.
43. Gibrat, J.-F., Madej, T., and Bryant, S. H. (1996) Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol. 6*, 377–385.
44. Omichinski, J. G., Clore, G. M., Robien, M., Sakaguchi, K., Appella, E., and Gronenborn, A. M. (1992) High-resolution solution structure of the double Cys2His2 zinc fingers from the human enhancer binding protein MBP-1. *Biochemistry 31*, 3907–3917.
45. Wang, Z., Feng, L. S., Matskevich, V., Venkataraman, K., Parasuram, P., and Laity, J. H. (2006) Solution structure of a Zap1 zinc-responsive domain provides insights into metalloregulatory Transcriptional Repression in *Saccharomyces cerevisiae*. *J. Mol. Biol. 357*, 1167–1183.
46. Pavletich, N. P., and Pabo, C. O. (1993) Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science 261*, 1701–1707.
47. Nolte, R. T., Conlin, R. M., Harrison, S. C., and Brown, R. S. (1998) Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl. Acad. Sci. U.S. A. 95*, 2938–2943.
48. Pavletich, N. P., and Pabo, C. O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science 252*, 809–817.
49. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) Zif268 protein-DNA complex refined at 1.6Å: a model system for understanding zinc finger-DNA interactions. *Structure 4*, 1171–1180.
50. Lee, M. S., Kliewer, S. A., Provencal, J., Wright, P. E., and Evans, R. M. (1993) Structure of the retinoid X receptor alpha DNA binding domain: a helix required for homodimeric DNA binding. *Science 260*, 1117–1121.
51. Zechel, C., Shen, X. Q., Chen, J. Y., Chen, Z. P., Chambon, P., and Gronemeyer, H. (1994) The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats. *EMBO J. 13*, 1425–1433.
52. Muto, Y., Pomeranz Krummel, D., Oubridge, C., Hernandez, H., Robinson, C. V., Neuhaus, D., and Nagai, K. (2004) The structure and biochemical properties of the human spliceosomal protein U1C. *J. Mol. Biol. 341*, 185–198.
53. Hayes, P. L., Lytle, B. L., Volkman, B. F., and Peterson, F. C. (2008) The solution structure of ZNF593 from *Homo sapiens* reveals a zinc finger in a predominantly unstructured protein. *Protein Sci. 17*, 571–576.
54. Möller, H. M., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. (2005) Solution structure of the N-terminal zinc fingers of the *Xenopus laevis* double-stranded RNA-binding protein ZFa. *J. Mol. Biol. 351*, 718–730.
55. Dutnall, R. N., Neuhaus, D., and Rhodes, D. (1996) The solution structure of the first zinc finger domain of SWI5: a novel structural extension to a common fold. *Structure 4*, 599–611.
56. Fairall, L., Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993) The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature (London) 366*, 483–7.
57. Laskowski, R. A., Watson, J. D., and Thornton, J. M. (2005) ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res. 33*(Web Server issue), W89–W93.
58. Holm, L., and Sander, C. (1998) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res. 26*, 316–319.
59. Wikström, M., Drakenberg, T., Forsén, S., Sjöbring, U.and Björck, L. (1994) Three-dimensional solution structure of an immunoglobulin light chain-binding domain of Protein L. Comparison with the IgG-binding domains of Protein G. *Biochemistry 33*, 14011–1417.