Structure, Vol. 11, 1393–1402, November, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.str.2003.09.015

The Zinc Finger-Associated Domain of the *Drosophila* Transcription Factor Grauzone Is a Novel Zinc-Coordinating Protein-Protein Interaction Module

Ralf Jauch,[1] Gleb P. Bourenkov,[3] Ho-Ryun Chung,[1] Henning Urlaub,[2] Ulrich Reidt,[2] Herbert Jäckle,[1,*] and Markus C. Wahl[2]

[1]Max-Planck Institut für biophysikalische Chemie
Abteilung Molekulare Entwicklungsbiologie
[2]Abteilung Zelluläre Biochemie/
Röntgenkristallographie
Am Faßberg 11
D-37077 Göttingen
[3]MPG-ASMB c/o DESY
Arbeitsgruppe Proteindynamik
Notkestraße 85
D-22603 Hamburg
Germany

Summary

About one-third of the more than 300 C2H2 zinc finger proteins of *Drosophila* contain a conserved sequence motif, the zinc finger-associated domain (ZAD). Genes that encode ZAD proteins are specific for and expanded in the genomes of insects. Only three ZAD-encoding gene functions are established, and the role of ZAD is unknown. Here we present the crystal structure of the ZAD of Grauzone (ZAD_{Grau}), a *Drosophila* transcription factor that specifically controls the maternal Cdc20-like APC subunit Cortex. ZAD forms an atypical treble-clef-like zinc-coordinating fold. Head-to-tail arrangement of two ZAD_{Grau} molecules in the crystals suggests dimer formation, an observation supported by crosslinking and dynamic light scattering. The results indicate that ZAD provides a novel protein-protein interaction module that characterizes a large family of insect transcription factors.

Introduction

A large number of protein domains, collectively referred to as zinc fingers, bind zinc ions through various combinations of four cysteine and/or histidine residues (Berg and Shi, 1996; Klug and Schwabe, 1995). The coordination of a zinc ion allows such modules to adopt compact three-dimensional structures which are otherwise too small to maintain extensive hydrophobic cores (Grishin, 2001). The prototype zinc finger proteins contain a 30 amino acid C2H2 motif (Miller et al., 1985; Pavletich and Pabo, 1991). In addition to this motif, a number of variants were found which are distinguished by zinc-coordinating residues, the secondary structural elements contributing the zinc ligands and/or topology (Krishna et al., 2003; Laity et al., 2001). While a large number of zinc fingers function as DNA binding domains (Pavletich and Pabo, 1991; Rosenberg et al., 1986; Wolfe et al., 2000), it becomes increasingly clear that they can also mediate interactions of the protein with RNA, other proteins and small molecules of other kinds (Krishna et al., 2003; Mackay and Crossley, 1998; McCarty et al., 2003). Furthermore, zinc-coordinating folds may confer specific enzymatic activities on proteins (Lorick et al., 1999).

The C2H2 zinc finger motif characterizes one of the most abundant eukaryotic protein families (Lander et al., 2001). C2H2 zinc finger proteins frequently contain additional protein modules in their N termini (Collins et al., 2001) such as the mammalian Krüppel-associated box (KRAB) or the insect zinc finger associated domain (ZAD) (Lespinet et al., 2002). Both KRAB- and ZAD-containing C2H2 zinc finger proteins are characterized by lineage-specific expansions in the respective genomes (Chung et al., 2002; Lespinet et al., 2002; Looman et al., 2002). The KRAB domain acts as a transcriptional repressor module (Margolin et al., 1994; Peng et al., 2000), whereas the function of ZAD is unknown.

In silico studies have recently identified more than 90 ZAD proteins in *Drosophila*, accounting for ~28% of the total C2H2 zinc finger proteins of this organism (Chung et al., 2002). This family of proteins is characterized by a conserved constellation of four cysteines within the ZAD, the chromosomal clustering of the corresponding genes, and the lack of homologs in noninsect genomes (Chung et al., 2002). Although the identification of expressed sequence tags ensured that the majority of ZAD-encoding genes of the *Drosophila* genome are transcribed, only four annotated members of the gene family have been examined in some detail. Two of them, *Serendipity delta* (*Sry*-δ) and *grauzone* (*grau*), were shown to encode transcriptional activators of the egg-polarity gene *bicoid* (Payre et al., 1994) and the gene *cortex* which encodes a Cdc20-like APC subunit, respectively (Chen et al., 2000). The third factor, the Dorsal-interacting protein 1 (DIP1), was identified in a yeast two-hybrid screen as an interaction partner of the dorsoventral polarity transcription factor Dorsal (Bhaskar et al., 2000). Finally, *Zeste-white-5* (*Zw5*) was reported to confer enhancer-blocking activity by association with the boundary element *scs* and may thereby participate in chromatin structuring by providing an insulating activity (Blanton et al., 2003; Gaszner et al., 1999). *grau*, *Zw5*, and *Sry*-δ have also been characterized by mutant analysis showing that they carry vital functions (Chen et al., 2000; Crozatier et al., 1992; Gaszner et al., 1999). In two cases, single amino acid replacements within the ZAD, such as cysteine 7 by tyrosine in Sry-δ (Crozatier et al., 1992) and arginine 4 by glycine in Zw5 (Gaszner et al., 1999), were the cause of lethality, implying that conserved amino acids within the ZAD are essential for its function.

Here we present the structure and biochemical features of the ZAD of Grauzone (ZAD_{Grau}). Grauzone is a 570 amino acid long transcriptional regulator characterized by the N-terminal ZAD and an array of eight C-terminal C2H2 zinc finger domains which mediate the binding of Grauzone to a promotor element of its target gene *cortex* (Chen et al., 2000). In genetic terms, *grau*

*Correspondence: hjaeckl@gwdg.de

is necessary for the proper transcriptional activation of the gene *cortex*, coding for a Cdc20-like APC subunit (Chu et al., 2001). Maternal loss-of-function mutations of *grau* cause an irregular growth arrest in meiosis II, whereas homozygous *grau* mutant embryos, which have received maternal *grau* activity, develop into normal adults. The *grau* mutant phenotype suggests that *grau* is exclusively required during oogenesis and *cortex* is its only target gene (Chen et al., 2000). Our results show that ZAD_{Grau} exhibits a C4 zinc-coordinating fold with a novel treble-clef-like structure. The crystallographic data and additional biochemical evidence indicate that ZAD_{Grau} is a protein interaction module with the capability to form homodimers and suggest that ZAD provides a taxon-specific means for the assembly of C2H2 transcription factor complexes.

Results and Discussion

Sequence alignments of a total of 91 *Drosophila* C2H2 proteins and corresponding proteins of a variety of other insects such as *Anopheles gambiae* revealed a conserved N-terminal sequence motif that contains two invariant pairs of cysteines (Figure 1A) (for details see Chung et al., 2002). Due to the specific association with C2H2 family members of zinc finger proteins, this motif was termed zinc finger-associated domain (ZAD). In order to characterize this protein motif, we studied the ZAD of the transcription factor Grauzone (ZAD_{Grau}). The ZAD_{Grau} open reading frame (amino acid residues 2–90; Figure 1A) was fused to Glutathione S-transferase (GST) and the recombinant fusion protein was produced in bacteria (see Experimental Procedures). After an initial capture step, the ZAD_{Grau} portion of the fusion protein was liberated by protease digestion, purified by column chromatography, and crystallized (see Experimental Procedures). SDS-PAGE analysis of dissolved crystals revealed that they contain only the ZAD_{Grau} protein (data not shown). The fact that ZAD_{Grau} crystallizes provides first evidence that ZADs encompass an independently folding protein module.

ZAD Is a Zinc Binding Protein Module

Based on the four conserved cysteine residues, ZADs have been suggested to comprise metal binding modules (Chung et al., 2002; Lander et al., 2001; Lespinet et al., 2002). Because metal association is expected to contribute significantly to the stability of the fold, heat denaturation experiments were performed in the presence and absence of EDTA. In the absence of EDTA, ZAD_{Grau} was stable for at least 15 min at 60°C, whereas in the presence of EDTA (5 mM) a considerable portion of the protein precipitated when heated above 50°C for 15 min (Figure 1B). In these experiments, care was taken to exclude divalent metal ions from the purification procedure. The result therefore suggests that ZAD_{Grau} contains an endogenous metal center that confers stability on its structure. In order to test this proposal, we performed an X-ray fluorescence scan. With this approach, the characteristic emission lines for zinc were observed (Figure 1C). This result and the finding that the anomalous diffraction from the zinc center could be exploited to solve the structure with a native crystal (see below) demonstrate that ZAD_{Grau} includes a zinc ion.

Structure Solution and Quality of the Model

The primary sequence of ZAD_{Grau} is shown in Figure 1A. The crystal structure of ZAD_{Grau} was solved de novo by a two-wavelength multiple anomalous dispersion experiment around the K-edge of an endogenous Zn^{2+} ion. A segmented poly-alanine model could be manually fitted to the experimental electron density map and allowed the subsequent incorporation of side chains. The sequence assignment was guided by well-defined aromatic residues and by the two pairs conserved cysteines (C4, C7, C53, and C56; see Figures 1A and 2A), which coordinated the Zn^{2+} ion (Figure 1D). Three regions in flexible loops (V11-C17, D22-E27, and E41-E47; see Figures 1A and 2A) initially displayed weak densities and were fitted during the subsequent rounds of refinement and manual model building. These loop regions, which are made up of nonconserved residues, lack substantial intra- and interchain contacts. Thus, they gained above-average temperature factors during the refinement. The main chain electron density of the molecule remained fragmented around position S12-A14 and D22 in the final $2F_o$-F_c map, consistent with high internal flexibility and lack of crystal packing contacts in these regions. All other side chains of the refined model, except those of some hydrophilic surface residues, were entirely covered by the final $2F_o$-F_c map. No electron density developed for the C-terminal 9 residues (positions 82–90; Figure 1A) of ZAD_{Grau} and the 5 N-terminal vector-derived residues that are contained in the recombinant protein. The final model therefore encompasses the residues D2–S81 of ZAD_{Grau} that could be traced unambiguously. Presumably because of the considerable fraction of flexible residues at the termini, which could not be accounted for in the final model, the refinement converged with an R factor/R_{free} factor of 24.1%/26.7%. Convergence at these numbers is consistent with a rather high B factor extracted from a Wilson plot (52.6 $Å^2$), which is comparable to the averaged B factor of the final structure (52.9 $Å^2$).

During all refinement steps, 5% of the reflections were set aside to monitor the R_{free} factor (Table 1). Of the final model, 91.9% of the residues resided in the preferred regions of the Ramachandran plot, 6.8% in the additionally allowed areas. Only the single residue K29, which was well defined in both the experimental and the final model-derived maps assumed an unconventional ϕ/ψ conformation even after manual interference. The mean positional error of the model was estimated at 0.15 Å (Luzzati, 1952).

Structural Properties of ZAD_{Grau}

The structured portion of ZAD_{Grau} resembles the letter "b." Its approximate dimensions are 60 × 30 × 30 Å (Figures 2A and 2B). The N-terminal body of the "b" comprises a globular fold structured around a zinc ion. The C-terminal stem is formed by a long α helix (α2, positions 54–80) that contains almost one-third of all residues of the domain. Residues 82–90 are not included in the ZAD consensus sequence (Chung et al., 2002)

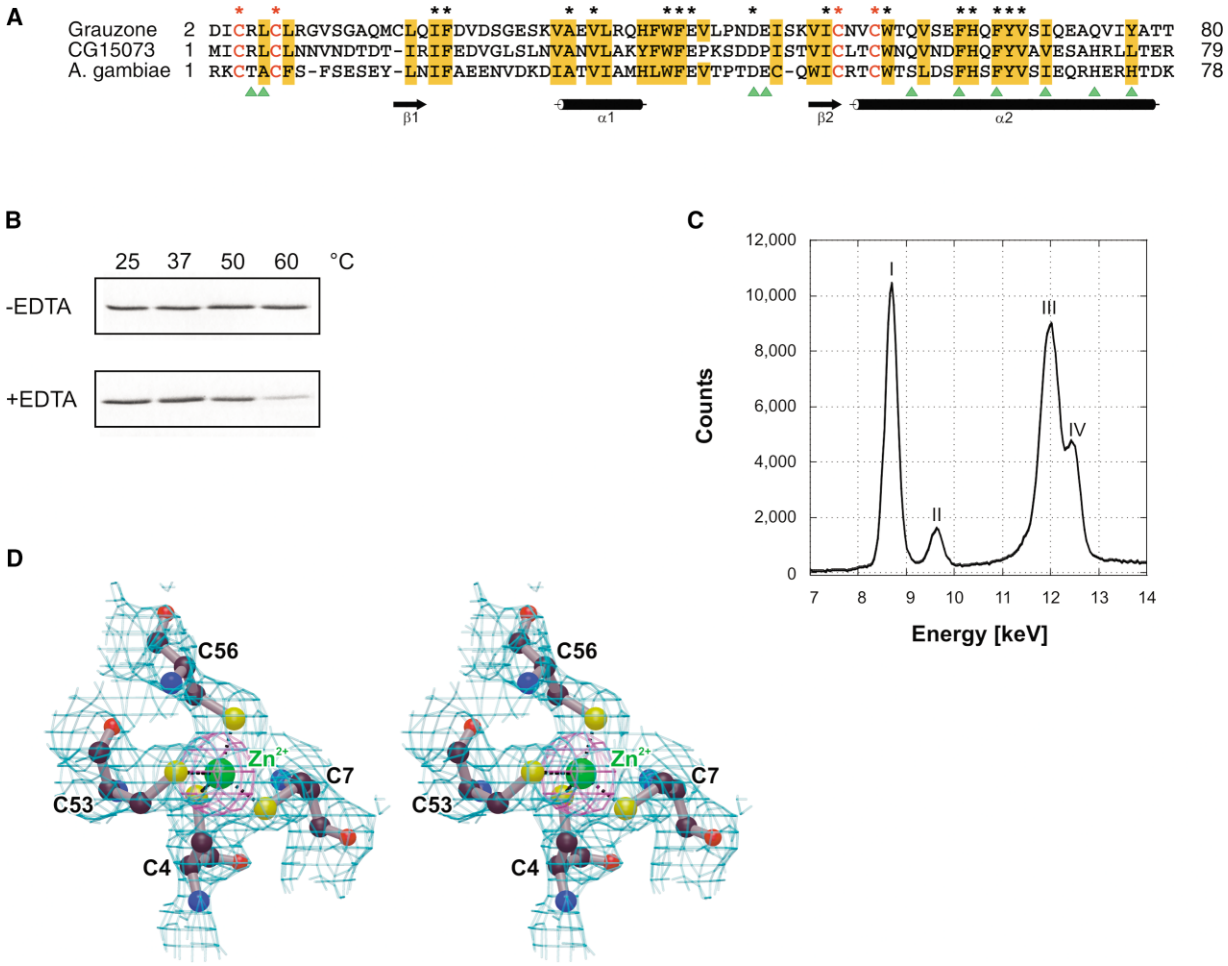

Figure 1. ZAD as a Zinc-Coordinating Fold

(A) Alignment of the Grauzone ZAD with its closest paralog, CG15073, in the *Drosophila melanogaster* genome and with a putative *Anopheles gambiae* homolog (ZAD only fragment, see Experimental Procedures). Asterisks mark 100% conservation in all three ZADs; red characters, invariant cysteine pairs; yellow boxes, conserved hydrophobic or aromatic amino acid residues.

(B) Portion of ZAD_{Grau} remaining soluble after incubation at the indicated temperatures in the absence (−EDTA) and in the presence of 5 mM EDTA (+EDTA). Reduction of the soluble fraction is clearly visible when ZAD_{Grau} is heated above 50°C in presence of the chelator.

(C) X-ray emission spectrum of a native ZAD_{Grau} crystal irradiated with X-radiation of $\lambda = 1.0$ Å. I, ZnKα line; II, ZnKβ line; III, Compton scattering; IV, elastic scattering.

(D) Electron density maps around the zinc center in ZAD_{Grau}. Cyan, experimental MAD electron density map (1 σ); magenta, anomalous difference Fourier map (10 σ), generated with the anomalous differences at the peak wavelength and the phases obtained after solvent flattening. The zinc-coordinating cysteine residues are labeled and shown in ball-and-stick. If not mentioned otherwise, all structural figures were prepared with Bobscript (http://www.strubi.ox.ac.uk/bobscript/) and rendered with Raster3D (Merritt and Bacon, 1997).

and are disordered in the present structure. Presumably, they provide a flexible linker to the remainder of the molecule, i.e., an acidic region followed by an array of eight C2H2 zinc fingers in the case of Grauzone (Chen et al., 2000).

The arrangement of the different ZAD structural elements is summarized in Figures 2A–2C and detailed in the Figure 2 legend. The fold of ZAD_{Grau} appears to be critically dependent on the zinc coordination, a result that is consistent with the EDTA-dependent precipitation of ZAD_{Grau} shown above. The two pairs of coordinating cysteines are approximately 50 residues apart. Zinc coordination, therefore, links the β2-α2 transition region at the center of the molecule with the N terminus of the domain likely to solidify the structure. In this view, the position of the zinc ion allows the fold to constrain the slanting angle of the C-terminal helix, α2, with respect to the long axis of the β sheet (Figures 2B–2C). The relative orientation of this long helix and the β sheet is further defined by helix α1, which rests with one of its surfaces on the end of the β sheet distal to the zinc ion and the adjacent loop regions. With a neighboring surface, helix α1 is in contact with the N-terminal and central portions of helix α2. Both the α1 sheet and the α1-α2 associations are based on extensive hydrophobic contacts forming a considerable hydrophobic core (α1: V29, V33, L34, H37, and F38; β1/β2: L18, I20, and I52; α2: W57, V60, F63, H64, and Y67; Figures 1A and 2A–2C). In addition, the conserved H37 residue (atom NE2) of helix α1 hydrogen bonds at the edge of this core to the

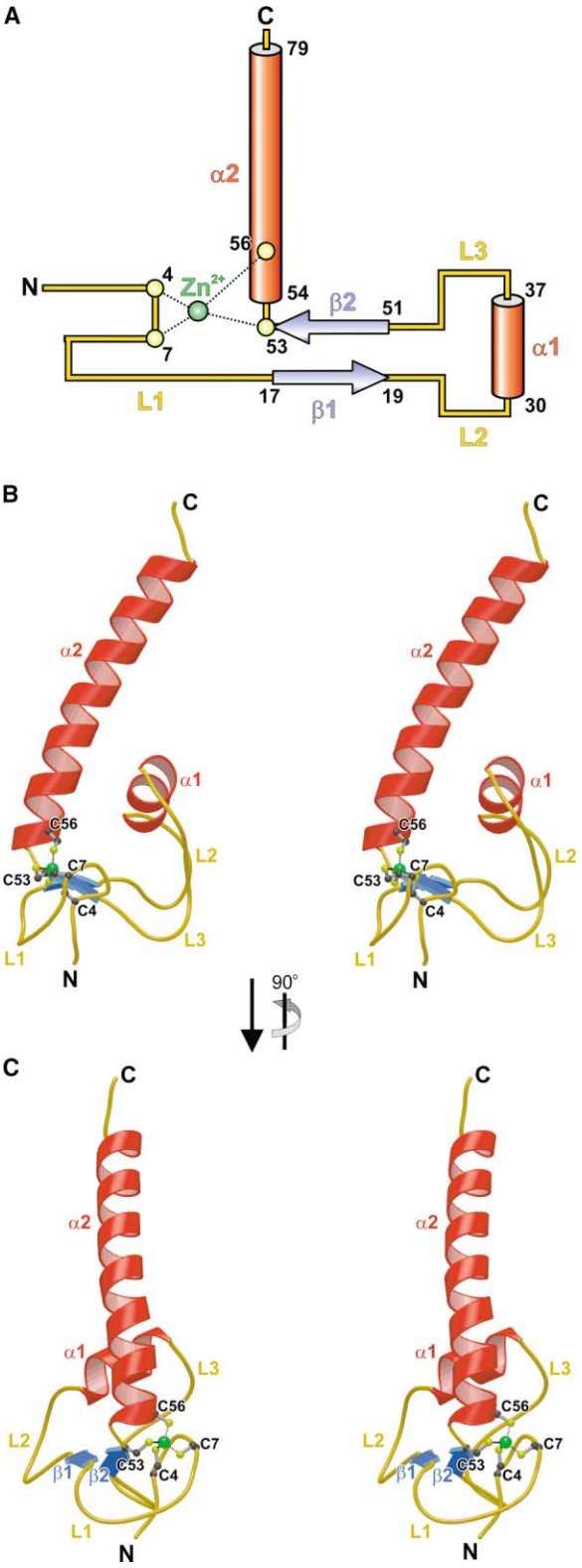

Figure 2. ZAD Structure and Topology

(A and B) Two orthogonal stereo ribbon plots of the ZAD_{Grau} structure. α helices, red; β strands, blue; loops, gold. In the side view of (A), the structure resembles the letter "b." N- and C termini, secondary structure elements, and residues of the zinc center are labeled. The Zn^{2+} ion (green) and its coordinating cysteine side chains are shown in ball-and-stick.

conserved H64 (atom ND1) of helix α2. In conclusion, the L2/α1/L3 region between the two central β strands and the zinc ion provide platforms which orient the C-terminal helix relative to the globular portion of the ZAD_{Grau} molecule.

To further support the overall structure of ZAD_{Grau} obtained by X-ray crystallography (Figures 2A–2D), we took circular dichroism spectra of the molecule in solution (data not shown). The calculated secondary structure content (45%–50% helix, 5%–10% strands) obtained from these recordings compare favorably to those seen in the crystal structure (37% helix, 6% strands). The conservation of length, the predicted secondary structure, the arrangement of the annotated secondary structure elements and the critical amino acid residues of ZADs (Chung et al., 2002) (Figure 1A) suggest that the present ZAD_{Grau} crystal structure provides a prototype for ZAD folding.

Structure Comparison

Next we compared the ZAD fold with known structures by manual inspection, with special emphasis on reported zinc finger fold groups (Krishna et al., 2003). We found that ZAD exhibits a structural relation to treble clef zinc finger (TCZF) domains. Figure 3A shows a comparison of ZAD with the structure of the C-terminal C4 TCZF of chicken GATA-1 (PDB ID code: 2gat) (Tjandra et al., 1997), revealing a similar topology of the secondary structure elements. The treble clef element has been identified among the members of seven different SCOP (http://scop.berkeley.edu/) fold groups including RING fingers, LIM domains, ribosomal proteins L24E and S14, and nuclear receptor-like fingers (Grishin, 2001; Krishna et al., 2003). Thus, despite the lack of significant sequence similarity between ZADs and TCZFs, which prevented recognition of their relation in the past, ZAD can be regarded as a novel, atypical TCZF that exerts several distinct features. Unlike the canonical zinc knuckle of

(C) Topological diagram of the ZAD_{Grau} structure identifying the origin of the four zinc-coordinating cysteines and the borders of the secondary structure elements (numbers).

(D) Alignment of 10 ZAD sequences. Sequence numbering corresponds to that of ZAD_{Grau}. Highly conserved residues (Chung et al., 2002) are shown on a golden background. Green triangles above the alignment indicate residues involved in dimer formation (see also Figure 4). Secondary structure elements as seen in the ZAD_{Grau} structure are indicated below the alignment. Note that the N terminus of ZAD_{Grau} is folded into a hairpin motif (D2-V11) and contains two conserved zinc-coordinating cysteine residues (C4 and C7; Figure1A). Replacement of C7 by a tyrosine residue has been described as a mutation of Sry-δ (Crozatier et al., 1992), implying that C7 is essential for ZAD function directly involved in zinc coordination. The N-terminal hairpin is followed by a loop (L1: S12-M16) which extends into a two-stranded antiparallel β sheet (β1: C17-Q19; β2: V51-C53) that is interspersed by a long loop-helix-loop insertion (L2: I20-K29; α1: V30-H37; L3: F38-K50) and followed by a C-terminal helix, α2. This helix is directly linked to the second β strand with a slanting angle of ~80° between the helix axis and the average long axis of the β sheet. Helix α1 of the L2-α1-L3 motif is positioned perpendicular to both the long axis of the sheet and the axis of helix α2, wedging between the two motifs. Conserved C53 and C56 residues reside in the C-terminal part of strand β2 and the first turn of helix α2, respectively, and complete the zinc coordination sphere.

Table 1. Crystallographic Data

Dataset	Remote	Peak
Data Collection		
Spacegroup	$P4_12_12$	
Unit cell lengths (Å)		
a	48.7	
b	48.7	
c	82.1	
Wavelength (Å)	1.0500	1.2828
Resolution (Å)	50.0–2.0	50.0–2.5
Unique Reflections	6907	6089
Redundancy	4.6	2.6
Completeness (%)	97.0 (99.7)	96.3 (98.5)
I/_(I)	36.7 (6.3)	28.2 (14.2)
R_{sym}[a] (%)	4.2 (44.7)	3.1 (8.2)
Phasing		
Resolution (Å)	20.0–2.1	
Heavy atom sites	1	
R_{Cullis}[b]		
Centrics	0.50	
Acentrics	0.61	
Phasing power[c]		
Centrics	1.67	
Acentrics	1.97	
FOM[d]		
Before DM	0.42	
After DM	0.78	
Refinement		
Resolution (Å)	15.0–2.0	
Number of reflections	6590	
Number of nonhydrogen atoms		
Protein	644	
Zn^{2+} ions	1	
Water oxygens	56	
R_{work}[e] (%)	24.1 (29.6)	
R_{free}[e] (%)	26.7 (33.7)	
Rmsd from ideal geometry		
Bond lengths (Å)	0.009	
Bond angles (°)	1.20	
Average B factors (Å^2)		
Protein	50.6	
Zn^{2+} ion	37.2	
Water oxygens	80.3	
Wilson B factor	52.6	
Rmsd B factors (Å^2)		
Main chain bonds	1.9	
Main chain angles	3.3	
Side chain bonds	3.5	
Side chain angles	5.2	
Ramachandran analysis (%)		
Preferred	91.9	
Additionally allowed	6.8	
Disallowed	1.4	
Mean residual error (Å)	0.15	

Data for the last 0.05 Å in parentheses. DM, density modification (solvent flattening); rmsd, root-mean-square deviation.

[a] $R_{sym}(I) = (\Sigma_{hkl}\Sigma_i[|I_i(hkl) - \langle I(hkl)\rangle|]/\Sigma_{hkl}\Sigma_i[I_i(hkl)]$; $I_i(hkl)$ – intensity of the ith measurement of hkl; $\langle I(hkl)\rangle$ – average value of hkl for all i measurements.

[b] $R_{Cullis} = \Sigma_{hkl}[||F_{PH} \pm F_P| - |F_{H,calc}||]/\Sigma_{hkl}[|F_{PH} \pm F_P|]$; $(F_{PH} + F_P)$ if signs are opposite, $(F_{PH} - F_P)$ if equal.

[c] Phasing power = $(\Sigma_n[|F_h|^2]/\Sigma_n[|E|^2])^{1/2}$; $\Sigma_n|E^2|$ = lack of closure error = $\Sigma_n[|F_{PH}|(obs) - |F_{PH}|(calc)]^2$

[d] FOM = figure of merit = $[|F(hkl)_{best}|]/|F(hkl)|$; $F(hkl)_{best} = \Sigma_\alpha[P(\alpha)F_{hkl}(\alpha)]/\Sigma_\alpha[P(\alpha)]$.

[e] $R_{work} = \Sigma_{hkl}[||F_{obs}| - k|F_{calc}||]/\Sigma_{hkl}[|F_{obs}|]$; $R_{free} = \Sigma_{hkl\subset T}[||F_{obs}| - k|F_{calc}||]/\Sigma_{hkl\subset T}[|F_{obs}|]$; $hkl\subset T$ – test set

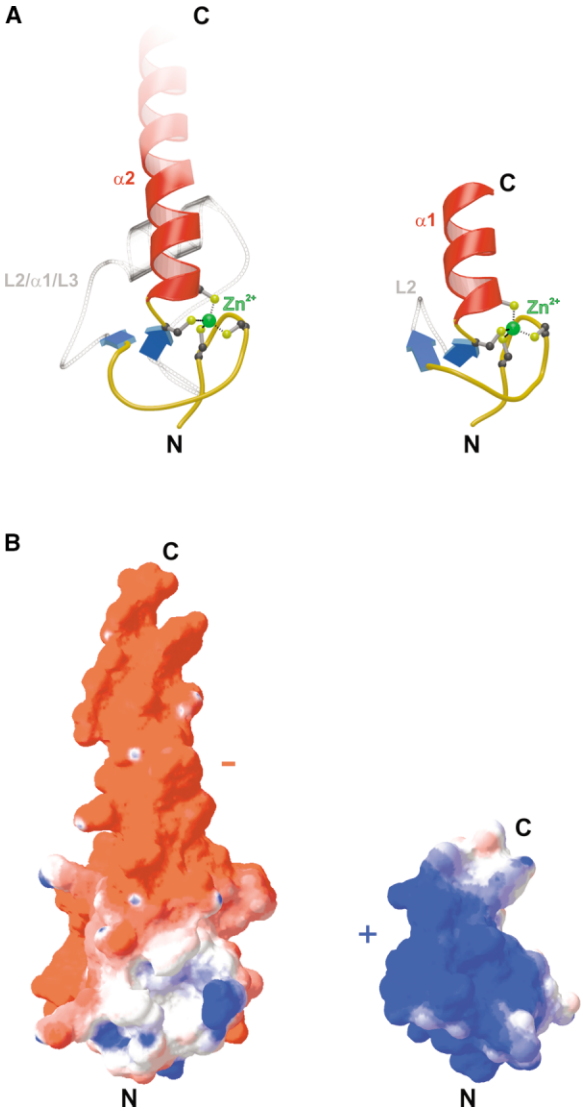

Figure 3. Structural Comparison

Ribbon (A) and electrostatic surface representation (B) of ZAD_{Grau} (left panels) and the C-terminal TCZF from chicken GATA-1 (right panels). Color coding for the ribbon plot is as in Figure 2. In the surface plot, red indicates negative potential, blue positive potential. The orientation is the same as in Figure 2B. This and all other surface images were prepared with SwissPDBViewer (Guex and Peitsch, 1997).

the TCZFs (Grishin, 2001), the ZAD N-terminal hairpin lacks the CPXCG consensus and an extended β conformation at both ends. Instead, the residue following the first cysteine in ZAD is a conserved arginine, and the loop is stabilized by only one backbone-to-backbone hydrogen bond. Furthermore, in a typical TCZF the two central β strands are connected by a terse loop, whereas in ZAD, the corresponding loop is expanded into the L2/α1/L3 motif. In addition, helix α2 of ZAD comprises six turns and thus is considerably longer than the corresponding helices of TCZFs, which typically contain a maximum of four turns only (Grishin, 2001). As detailed above, insertion of helix α1 via the L2/α1/L3 motif may

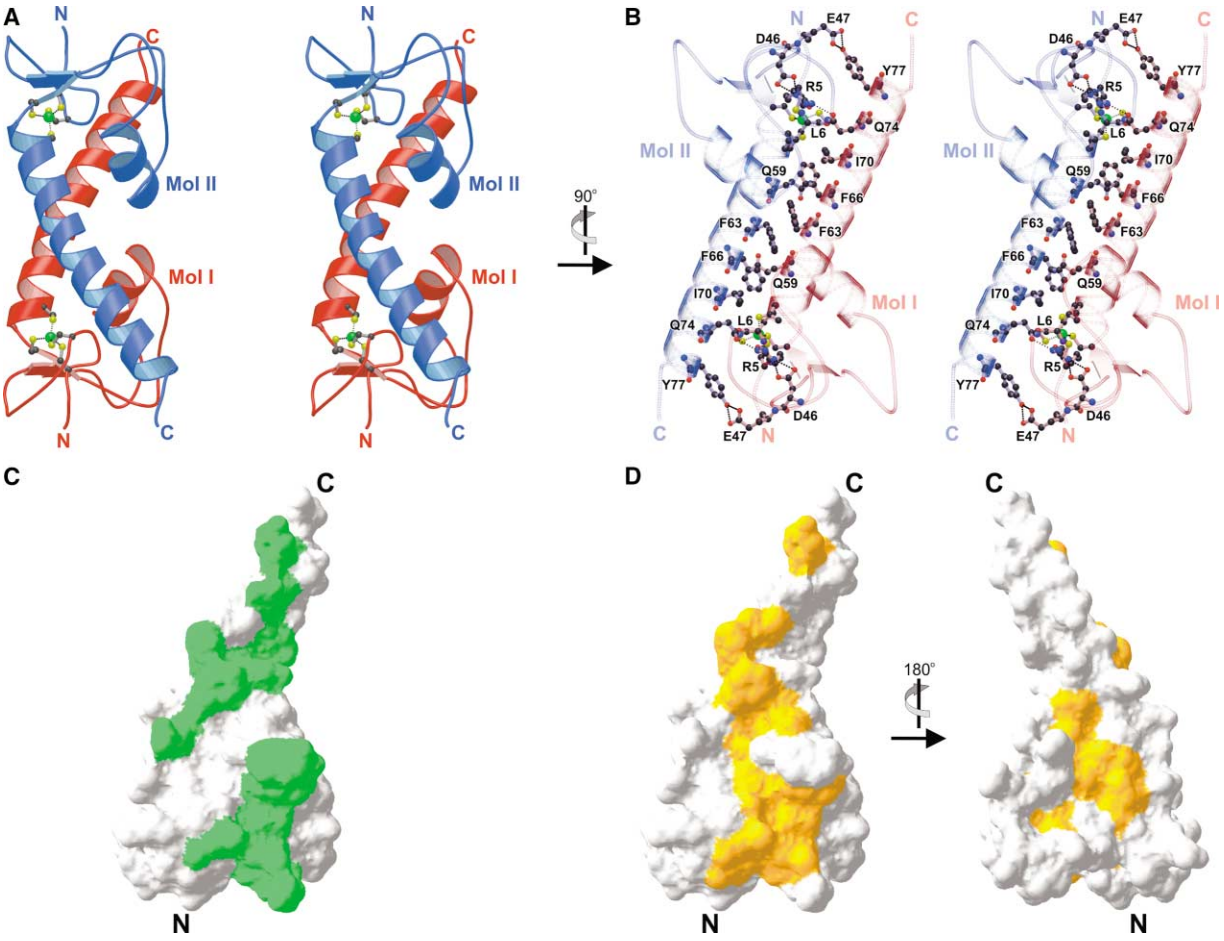

Figure 4. The ZAD_{Grau} Dimer

(A) Stereo ribbon plot of a ZAD_{Grau} dimer as seen in the crystal. The two subunits (Mol I and Mol II) are colored red and blue; the two zinc centers are depicted in ball-and-stick. N- and C termini are labeled. The orientation of the rear subunit is the same as in Figure 2A.

(B) Stereo representation of residue interactions in the dimer interface, 90° from the view in (A). The two ZAD_{Grau} subunits are colored as in (A) but are rendered semitransparent to clearly reveal the interacting residues in ball-and-stick. All residues and the molecular termini are labeled.

(C) Mapping of the contact residues within the dimer (green) onto the surface of a ZAD_{Grau} monomer (gray). The orientation of the molecule is the same as that of the red subunit in (A).

(D) Mapping in orange of conserved residues on the surface of ZAD_{Grau} in two diametric views. The left orientation is the same as in (C), indicating that the largest conserved surface patch and the dimer interface largely coincide.

be a prerequisite for extending the length of helix α2, without loosing its defined orientation relative to the β sheet. Finally, helix α2 displays a conserved pattern of hydrophobic residues in ZAD (Figure 1D; see also below) which is absent from other TCZFs. ZAD can therefore be classified as a distinct subgroup of TCZF domain structures.

Crystal Structure Suggests ZAD_{Grau} Homodimers

The evolutionary restriction of ZADs to certain C2H2 zinc finger genes in insects, the chromosomal clustering of the majority of these ZAD-containing genes and the unique folding characteristics of the domain next to the DNA binding domain of transcriptional regulators such as Grauzone strongly argue that the ZAD is associated with a specific and distinct biological function. However, the mere similarity to the fold group of the TCZFs does not allow any conclusions about the specific role of ZAD, because treble clef motifs embody functions as diverse as binding to nucleic acids, proteins as well as small ligands and some may even exert enzymatic activity (Grishin, 2001).

However, the crystal structure of ZAD_{Grau} clearly supports a model, in which ZAD represents a protein-protein interaction module involved in homodimerization. Figure 4 shows that in the crystal two ZAD_{Grau} molecules are associated through a 2-fold axis in an isologous head-to-tail fashion. As revealed by the protein-protein interaction server (http://www.biochem.ucl.ac.uk/bsm/PP/server/), this contact buries $\sim$1000 Å^2 of accessible surface area (ΔASA). This value seems on the lower side when compared to known homodimers (Jones and Thornton, 1996). However, the interaction surface of ZAD_{Grau} covers a total of close to 20% of the entire surface area. More importantly, a large number of amino acid residues, which are strongly conserved among ZAD

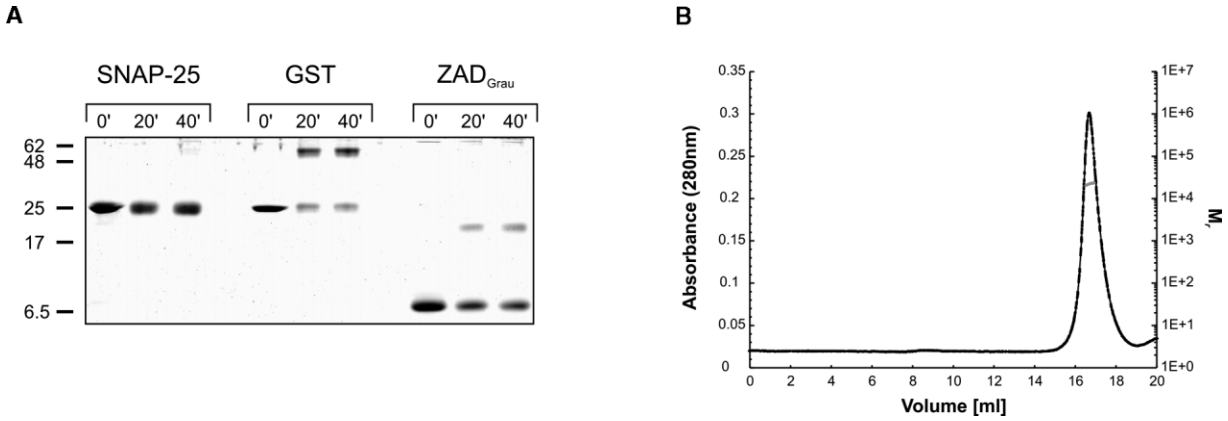

Figure 5. Detection of ZAD_{Grau} Dimers in Solution

(A) Results from chemical crosslinking with glutaraldehyde displayed on a denaturing polyacrylamide gel. Crosslinking took place for the times indicated above the lanes. SNAP-25 served as a negative control, GST as a positive control. Molecular weight markers are given on the left (kDa).

(B) Results from a gel filtration chromatography run combined with multiangle-laser-light scattering analysis of the emerging peak. Emergence of protein from the column was detected by absorbance at 280 nm (black trace, left *y* axis), readout from the light scattering of the single, symmetrical peak is depicted as a gray line (right *y* axis).

family members (for details see Chung et al., 2002), are responsible for the contacts between the two subunits (Figure 4B). When the conserved residues are mapped on the surface of ZAD_{Grau}, it becomes obvious that the largest conserved surface patch closely coincides with the presumed dimer interface (Figures 4C and 4D). In particular, hydrophobic residues of the long C-terminal helix α2 (F63, F66, I70, and Y77; Figure 1A) build up major parts of the contact interface (Figure 4B). As a consequence, 72.5% of the amino acid residues of the dimer interface are nonpolar. The presumed dimer interface is thus designed very differently from the remainder of the surface, which is lined with polar residues and exhibits a highly negative electrostatic potential (Figure 3B). In the region where the tip of helix α2 from one molecule contacts the globular portion of the other subunit, some intermolecular hydrogen bonds are observed as well (Q74-R5; Y77-E47; Figure 4B).

Because of the involvement of its hydrophobic side chains, the presumed homodimerization mode provides a straightforward explanation for the amphipathic design and the unusual length of helix α2 (Figure 2A). Since, in addition, both helix α2 and the globular portion of ZAD are involved in this dimerization mode, the importance of restraining their relative orientations by the inserted L2/α2/L3 module and the zinc coordination becomes obvious. Further supporting its significance, the present dimerization may explain the lethal phenotype observed with the conserved R4 (corresponding to R5 in Grauzone) mutated to glycine in Zw5 (Gaszner et al., 1999). Its side chain is positioned by ionic interactions with D46 to engage in a hydrogen bond with Q74 of the neighboring molecule (Figure 4). We note, however, that there are a number of other contacts between the two subunits that could stabilize subunit interactions. It is therefore possible that the mutation simply perturbs the globular structure of the N terminus which might preclude dimerization. In any case, the analysis of the crystal packing strongly suggests a functional homodimerization of ZAD_{Grau}.

Besides the above symmetrical contact between two ZAD_{Grau} molecules, other crystal contacts were observed. Some of these are not functional because of an unreasonably small ΔASA. Other contacts did not involve conserved residues. One rather intimate alternative association takes place through the crystallographic 4-fold screw axis. However, this symmetry element gives rise to heterologous contacts, leaving the bonding potentials of ZAD unsaturated. The latter interaction mode could therefore lead to the formation of larger oligomers. We therefore examined the arrangement of ZAD in solution.

ZAD_{Grau} Forms Dimers in Solution

The quaternary structure of ZAD_{Grau} in solution was investigated by two approaches. First, we performed chemical crosslinking experiments with glutaraldehyde (Figure 5A). In contrast to SNAP-25, which was reported not to self-interact (Fasshauer et al., 1999), and in parallel with the dimerizing GST, ZAD_{Grau} could be efficiently crosslinked to the dimer state, but no higher oligomers were observed. Secondly, multiangle-laser-light-scattering following size exclusion chromatography yielded strong evidence for homdimerization in solution (Figure 5B). ZAD_{Grau} eluted as a single symmetrical peak from various size exclusion columns with different optimal separation regimes. Comparison of the elution times with those of reference proteins was consistent with ZAD_{Grau} dimers (data not shown). The scattering signal at 632.8 nm across the elution peak indicated a molecular weight of 21.11 kDa, which matched the theoretical dimer mass (21.14 kDa) almost perfectly (Figure 5B). This high congruence and the lack of a monomer or a trimer signal, strongly suggested a monodisperse dimer solution.

Collectively, our results provide strong evidence for ZAD_{Grau} homodimerization under near-physiological salt conditions. They leave the 2-fold symmetrical association of ZAD_{Grau} in the crystal (Figure 4) as the only dimerization mode, which is consistent with all observations.

Because ZAD_{Grau} is an independently folding unit, dimerization is expected to prevail within the context of the full-length Grauzone transcription factor. By extension of this finding, a general function of ZADs could be to provide dimerization modules that mediate homodimer and/or heterodimer formation among closely related members of the ZAD transcription factor family. In support of this observation, homodimerization was also reported for the ZAD-containing transcription factor Sry-δ (Payre et al., 1997; Ruez et al.,1998). It is noteworthy that Sry-δ comprises a deviated version of ZAD family members and could thus not be considered as prototype for ZAD function (Chung et al., 2002).

ZAD_{Grau} Lacks DNA Binding Features

The majority of zinc finger proteins were shown to be nucleic acid binding proteins (Berg and Shi, 1996; Laity et al., 2001; Pavletich and Pabo, 1991; Wolfe et al., 2000). Because of this amply documented function, we asked whether the ZAD_{Grau} structure has features to support nucleic acid binding as well. In order to obtain first hints, we compared the ZAD with the structure of the C-terminal treble clef module of the chicken erythroid transcription factor GATA-1 (Tjandra et al., 1997), a representative of sequence-specific DNA binding TCZFs. Figure 3B shows that the two molecules display opposite electrostatic surface potentials. Whereas the TCZF of GATA-1 exhibits an almost continuously electropositive surface potential that facilitates a close association with the negatively charged sugar-phosphate backbone of the DNA, the ZAD_{Grau} is almost completely wrapped into negative potential. This observation strongly argues against DNA binding features of ZAD.

Superimposition of one subunit of a ZAD_{Grau} dimer on the structure of DNA-bound GATA-1 TCZF (Tjandra et al., 1997) revealed that the second ZAD_{Grau} module faces the minor groove of the DNA and would thus collide with the DNA backbone (data not shown). Furthermore, Grauzone mutants, which lack the ZAD-containing region of the protein, had unaltered DNA binding properties in vitro, whereas the part of Grauzone protein that contains the ZAD in the absence of C2H2 zinc fingers had no DNA binding activity (Chen et al., 2000), supporting the argument that ZAD_{Grau} has no DNA binding features. Since the key positions and features of the amino acid residues are conserved among ZADs (for an alignment, see Chung et al., 2002), it appears unlikely that any of them carries DNA binding properties. We therefore propose that ZADs, in general, are protein interaction modules involved in homodimerization and/or the formation of protein complexes. To further support this conclusion, we also compared the ZAD_{Grau} dimer with the DNA-bound dimer of other members of the TCZF fold group, the glucocorticoid receptors (Luisi et al., 1991; Grishin, 2001). In these receptor molecules, the DNA binding helix equivalent to ZAD helix α2 is employed as a recognition element, which is positioned in the DNA major groove. Two DNA binding domains interact through their C-terminal helical extension, which have no equivalent in ZADs. These C-terminal extensions yield a spacing between the two recognition helices from neighboring subunits, which is large enough to position them both into neighboring turns of the DNA major groove. Therefore, the dimerization mode of ZADs is fundamentally different from the interaction mode seen in the DNA binding domains of glucocorticoid receptors, consistent with these molecules serving different functions.

Experimental Procedures

Sample Preparation and EDTA-Dependent Precipitation

The DNA coding for Grauzone ZAD residues 2–90 was PCR amplified, cloned into the Not1/EcoR1 sites of the pGEXT-3 vector, and recombinantly expressed as a GST-fusion protein in a BL21(DE3) *E. coli* strain. Cells were grown to an OD_{600} 0.7–1.0 at 30°C, shifted to 20°C, and induced with 0.7 mM IPTG overnight. Cells were harvested by centrifugation and resuspended in phosphate-buffered saline containing Complete protease inhibitor tabs (Roche) and traces of lysozyme. Cells were lysed by sonication and the insoluble fraction was removed by centrifugation. GST-ZAD_{Grau} was captured using Glutathione (GSH)-Sepharose 4B (Amersham) beads in disposable columns following the manufacturer's suggestions. Samples were eluted with 50 mM Tris-HCl, pH 8.0, 50 mM GSH. Thrombin was directly added to the eluate to 1 U/100 μg protein and incubated overnight at room temperature. The protein identity was confirmed by peptide finger printing and subsequently purified with a 1 ml Ressource-Q 15 μm anion exchange column applying a linear 20 ml salt gradient (20 mM Tris-HCl, pH 8.0, 0–1 M NaCl), a second GSH column and a Superdex 25/60 75 pg gel filtration (GF) column (GF buffer: 10 mM Tris, pH 7.5, 100 mM NaCl, 5 mM DTT). Samples were concentrated to 8 mg/ml via 20 ml 5000 Da molecular weight cutoff concentrators (Vivascience). For EDTA-dependent precipitation studies, 10 μM ZAD_{Grau} samples in GF buffer either lacking EDTA or containing 5 mM EDTA were incubated for 15 min at room temperature, 37°C, 50°C, and 60°C. Samples were pelleted and the supernatant was analyzed by SDS-PAGE.

Crystallization and Data Collection

Crystallization screens were carried in the sitting drop vapor diffusion format on 24-well Cryschem plates (Hampton Research, Laguna Niguel). 1–4 μl of an 8 mg/ml ZAD_{Grau} solution in GF buffer were combined with 1 μl reservoir buffer and equilibrated against a 400 μl reservoir. Crystals grew at 22°C with 0.2 M ammonium acetate, 0.1 M sodium citrate, pH 5.6, and 30% polyethylene glycol 4000 in 3–10 days.

After transfer into Paratone-N (Hampton Research) and removal of residual mother liquor, crystals could be shock frozen in a liquid nitrogen stream. Diffraction data were collected at 100 K on the HASYLAB beamline BW6 (DESY, Hamburg, Germany; http://www-hasylab.desy.de/facility/experimental_stations/stations/BW6.htm). X-ray fluorescence spectra of native crystals clearly indicated zinc emission lines (Figure 1B). A single such crystal yielded complete anomalous data sets at the f″-maximum of the K-edge (λ = 1.2828 Å) and at a remote wavelength (λ = 1.0500 Å). The data were recorded on a Mar-Research (Norderstedt, Germany) CCD detector and processed with the HKL package (Otwinowski and Minor, 1997) (see also Table 1).

Structure Solution, Model Building, and Refinement

If not mentioned otherwise, programs from the CCP4 collection (CCP4, 1994) were employed for the structure solution. The location of a single Zn^{2+} ion could be deduced from anomalous difference Patterson maps at the f″-maximum (program RSPS) or by direct methods (SHELX, http://shelx.uni-ac.gwdg.de/SHELX/) and was used to phase the data with the remote data set as the reference (MLPHARE). The hand of the heavy atom was revealed during solvent flattening (DM).

Prominent secondary structure elements were manually fitted to the electron density with MAIN (Turk, 1996). Subsequently, these regions were decorated with side chains guided by the prominent features of the aromatic residues and by four cysteine residues, which coordinated the Zn^{2+} ion. The connecting loops were sequentially traced by alternating refinement of the partial model (CNS) (Brunger et al., 1998) and model building guided by the experimental

and $2F_o$-F_c electron density maps. Refinement included standard procedures of positional and B factor optimization, a round of simulated annealing, and positioning of 56 water molecules into spherical peaks of the F_o-F_c maps. The final step was a TLS refinement with REFMAC5 with the structure divided into four rigid bodies (the Zn^{2+} ion, residues 2–19 and 51–54, residues 20–50, and residues 55–81). No data within 15.0–2.0 Å were excluded, and during all refinement steps the same set of reflections (5%) was used to monitor the R_{free} factor (Table 1). The geometry of the final model was analyzed with PROCHECK (Laskowski et al., 1993) (Table 1), secondary structure elements were extracted with PROMOTIF (http://www.biochem.ucl.ac.uk/bsm/promotif/promotif.html).

CD Spectroscopy

For far UV-CD measurements, the sample was dialyzed against 10 mM Na_2HPO_4, pH 7.9. The spectra (190–260 nm) were recorded with a JASCO J-720 spectropolarimeter (Jasco Corporation, Tokyo, Japan) at 20°C in a 0.1 cm cuvette with a scanning speed of 50 nm/min. 30 scans were averaged and baseline corrected. The spectra were analyzed using the neural network-based CD deconvolution software CDNN v2.1 (Böhm et al., 1992).

Chemical Crosslinking

The crosslinking reactions were carried out at 25°C in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl with 15 μM samples of ZAD_{Grau}, GST (26 kDa), and SNAP-25 (25 kDa) (Fasshauer et al., 1999) and 1 mM glutaraldehyde. The reactions were stopped at selected time points by boiling in SDS sample buffer.

Multiangle-Laser-Light Scattering

The experiment was performed on a HR-10/30 Superdex-200 size exclusion column (Amersham) connected to a UV spectrometer and the Dawn and Optilab instruments XY (Wyatt Technology Corp.). 200 μl of a 25 μM ZAD_{Grau} sample were chromatographed in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl. The UV_{280} absorption, the light scattering at 632.8 nm and the differential refraction of the elution profile were monitored. Spectra were analyzed with the Astra software package (Wyatt, 1993). We also checked ZAD_{Grau} migration on an analytical Superdex-75 column in comparison to protein molecular weight standards.

Identification of a Putative Grauzone ZAD Homolog in the *Anopheles gambiae* Genome

We searched the whole *Anopheles gambiae* genome using the ZAD HMM described in Chung et al., 2002 and the Wise package 2.2.0 (Birney et al., 1996). All identified ZADs were aligned with all *Drosophila melanogaster* ZADs using ClustalW 1.8.1. (Thompson et al., 1994), false positive hits were eliminated. The alignment was used to construct a neighbor-joining tree with ClustalW. Based on this tree, we identified the closest relative of the Grauzone ZAD in the *Anopheles gambiae* genome. The alignment of the Grauzone ZAD with the closest paralog, CG15073, of *Drosophila melanogaster* and the putative Anopheles homolog was performed with ClustalW using default parameters.

Acknowledgments

We thank our colleagues in the labs for their various contributions, T. Siddiqui for kindly providing SNAP-25, G. Dowe for sequencing, and Stefan Pabst for help with MALLS analysis. Work was supported by the Max Planck Society. H.-R.C. is a fellow of the Boehringer Ingelheim Fonds.

Received: June 30, 2003
Revised: July 7, 2003
Accepted: July 9, 2003
Published: November 4, 2003

References

Berg, J., and Shi, Y. (1996). The galvanization of biology: a growing appreciation for the roles of zinc. Science *271*, 1081–1085.

Bhaskar, V., Valentine, S., and Courey, A. (2000). A functional interaction between dorsal and components of the smt3 conjugation machinery. J. Biol. Chem. *275*, 4033–4040.

Birney, E., Thompson, J., and Gibson, T.J. (1996). PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. Nucleic Acids Res. *24*, 2730–2739.

Blanton, J., Gaszner, M., and Schedl, P. (2003). Protein:protein interactions and the pairing of boundary elements in vivo. Genes Dev. *17*, 664–675.

Böhm, G., Muhr, R., and Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng. *5*, 191–195.

Brunger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.

Chen, B., Harms, E., Chu, T., Henrion, G., and Strickland, S. (2000). Completion of meiosis in *Drosophila* oocytes requires transcriptional control by grauzone, a new zinc finger protein. Development *127*, 1243–1251.

Chu, T., Henrion, G., Haegeli, V., and Strickland, S. (2001). Cortex, a *Drosophila* gene required to complete oocyte meiosis, is a member of the Cdc20/fizzy protein family. Genesis *29*, 141–152.

Chung, H., Schäfer, U., Jäckle, H., and Böhm, S. (2002). Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in Drosophila. EMBO Rep. *3*, 1158–1162.

Collins, T., Stone, J., and Williams, A. (2001). All in the family: the BTB/POZ, KRAB, and SCAN domains. Mol. Cell. Biol. *21*, 3609–3615.

Crozatier, M., Kongsuwan, K., Ferrer, P., Merriam, J., Lengyel, J., and Vincent, A. (1992). Single amino acid exchanges in separate domains of the *Drosophila* serendipity delta zinc finger protein cause embryonic and sex biased lethality. Genetics *131*, 905–916.

Fasshauer, D., Antonin, W., Margittai, M., Pabst, S., and Jahn, R. (1999). Mixed and non-cognate SNARE complexes. Characterization of assembly and biophysical properties. J. Biol. Chem. *274*, 1440–1446.

Gaszner, M., Vazquez, J., and Schedl, P. (1999). The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. Genes Dev. *13*, 2098–2107.

Grishin, N. (2001). Treble clef finger—a functionally diverse zinc-binding structural motif. Nucleic Acid Res. *29*, 1703–1714.

Guex, N., and Peitsch, M. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714–2723.

Harms, E., Chu, T., Henrion, G., and Strickland, S. (2000). The only function of Grauzone required for *Drosophila* oocyte meiosis is transcriptional activation of the cortex gene. Genetics *155*, 1831–1839.

Jones, S., and Thornton, J. (1996). Principles of protein-protein interaction. Proc. Natl. Acad. Sci. USA *93*, 13–20.

Klug, A., and Schwabe, J. (1995). Protein motifs 5: zinc fingers. FASEB J. *9*, 597–604.

Krishna, S., Indraneel, M., and Grishin, N. (2003). Structural classification of zinc fingers: survey and summary. Nucleic Acids Res. *31*, 532–550.

Laity, J., Lee, M., and Wright, P. (2001). Zinc finger proteins: new insights into structural and functional diversity. Curr. Opin. Struct. Biol. *11*, 39–46.

Lander, E., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lespinet, O.W., YI. Koonin, EV., and Aravind, L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. *12*, 1048–1059.

Looman, C., Abrink, M., Mark, C., and Hellman, L. (2002). KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. Mol. Biol. Evol. *19*, 2118–2130.

Lorick, K., Jensen, J., Fang, S., Ong, A., Hatakeyama, S., and Weissman, A. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proc. Natl. Acad. Sci. USA *96*, 11364–11369.

Luisi, B., Xu, W., Otwinowski, Z., Freedman, L., Yamamoto, K., and Sigler, P. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. Nature *352*, 497–505.

Luzzati, V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. Acta Crystallogr. A *5*, 802–810.

Mackay, J., and Crossley, M. (1998). Zinc fingers are sticking together. Trends Biochem. Sci. *23*, 1–4.

Margolin, J., Friedman, J., Meyer, W., Vissing, H., Thiesen, H., and Rauscher, F. (1994). Kruppel-associated boxes are potent transcriptional repression domains. Proc. Natl. Acad. Sci. USA *91*, 4509–4513.

McCarty, A., Kleiger, G., Eisenberg, D., and Smale, T. (2003). Selective dimerization of a C2H2 Zinc Finger Subfamily. Mol. Cell *11*, 459–470.

Merritt, E., and Bacon, D.J. (1997). Raster3D: photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.

Miller, J., McLachlan, A., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. *4*, 1609–1614.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Pavletich, N., and Pabo, C. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. Science *252*, 809–817.

Payre, F., Crozatier, M., and Vincent, A. (1994). Direct control of transcription of the *Drosophila* morphogen bicoid by the serendipity delta zinc finger protein, as revealed by in vivo analysis of a finger swap. Genes Dev. *8*, 2718–2728.

Payre, F., Buono, P., Vanzo, N., and Vincent, A. (1997). Two types of zinc fingers are required for dimerization of the serendipity delta transcriptional activator. Mol. Cell. Biol. *17*, 3137–3145.

Peng, H., Begg, G., Harper, S., Friedman, J., Speicher, D., and Rauscher, F. (2000). Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. J. Biol. Chem. *275*, 18000–18010.

CCP4 (Collaborative Computational Project 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.

Rosenberg, H., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I., and Jackle, H. (1986). Structural homology of the product of the *Drosophila* Krüppel gene with Xenopus transcription factor-IIIA. Nature *319*, 336–339.

Ruez, C., Payre, F., and Vincent, A. (1998). Transcriptional control of *Drosophila* bicoid by Serendipity delta: cooperative binding sites, promoter context, and co-evolution. Mech. Dev. *78*, 125–134.

Thompson, J., Higgins, D., and Gibson, T. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Tjandra, N., Omichinski, J., Gronenborn, A., Clore, G., and Bax, A. (1997). Use of dipolar 1H–15N and 1H–13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat. Struct. Biol. *4*, 732–738.

Turk, D. (1996). MAIN 96: an interactive software for density modifications, model building, structure refinement and analysis. In Meeting of the International Union of Crystallography Macromolecular Macromolecular Computing School, P.E. Bourne and K. Watenpaugh, eds. (International Union of Crystallography).

Wolfe, S., Nekludova, L., and Pabo, C. (2000). DNA recognition by Cys2His2 zinc finger proteins. Annu. Rev. Biophys. Biomol. Struct. *29*, 183–212.

Wyatt, P. (1993). Light scattering and the absolute characterisation of macromolecules. Anal. Chim. Acta *272*, 1–40.

Accession Numbers

The structure has been submitted to the Protein Data Bank (http://www.rscb.org/pdb) under accession number 1PZW.