

Solution Structures of Two Zinc-finger Domains from SWI5 Obtained Using Two-dimensional ^1H Nuclear Magnetic Resonance Spectroscopy

A Zinc-finger Structure with a Third Strand of β -Sheet

David Neuhaus, Yukinobu Nakaseko[†], John W. R. Schwabe and Aaron Klug

MRC Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England

(Received 16 March 1992, accepted 14 July 1992)

This paper describes the detailed three-dimensional structures of two zinc-finger domains from the yeast transcription factor SWI5, calculated using the results of the n.m.r. experiments described in the accompanying paper. The structure of finger 2 is essentially similar to those previously obtained by others for isolated, synthetic single zinc-finger domains in solution, and for the three zinc-finger peptide Zif268 in its crystalline complex with DNA. The N-terminal half of the sequence forms a two-stranded, irregular β -sheet containing both of the metal-binding cysteine residues, while the remainder of the structure forms a helix. Approximately the first half of this helix is α -helical, whereas the C-terminal portion, including the two metal-binding histidine residues, is 3_{10} helical. Four invariant hydrophobic residues form a core to the structure. In contrast to all previously described structures of zinc-finger domains, finger 1 has an additional strand in the β -sheet, formed by residues N-terminal to the formal start of the finger motif. This additional strand plays a role in stabilising the folded form of finger 1, since a two-finger peptide lacking the N-terminal residues showed folded structure in finger 2 but not in finger 1.

Keywords: DNA-binding proteins; three-dimensional structure; transcriptional activator proteins; two-dimensional nuclear magnetic resonance spectroscopy; zinc-fingers

1. Introduction

In the accompanying paper (Nakaseko *et al.*, 1992) we described the preparation and characterization of peptides containing one, two and three zinc-finger domains from the yeast transcriptional activator protein SWI5. In particular, it was shown that adjacent finger domains in multi-finger peptides from SWI5 do not have a single preferred relative orientation when the peptide is in solution, in the absence of DNA. In this paper we go on to describe in more detail the three-dimensional structures calculated for finger 1 and for finger 2 of SWI5.

As summarized in the introduction to the accompanying paper, several n.m.r. structures of synthetic single zinc-finger peptides have been published (see accompanying paper for references), and recently the crystal structure of a three zinc-finger peptide from Zif268 bound to its DNA recognition site has been solved (Pavletich & Pabo, 1991). Despite this,

the motif is so widespread that it would be premature to assume that all such motifs have identical structures. This point is reinforced by the structure of SWI5 finger 1 described here, which is the first zinc-finger structure reported to have a significantly different architecture from those previously determined. In this paper we describe the calculations that led to this structure, and that of finger 2 of SWI5, and go on to consider the likely mode of binding of the zinc-fingers of SWI5 with their DNA recognition site.

2. Materials and Methods

Details of the n.m.r. measurements, including the acquisition of DQF-COSY[‡], PE-COSY, TOCSY, RELAY

[‡] Abbreviations used: DQF-COSY, double-quantum filtered correlation spectroscopy; PE-COSY, primitive exclusive correlation spectroscopy; TOCSY, total correlation spectroscopy; RELAY, relayed correlation spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; NOE, nuclear Overhauser effect; r.m.s.d., root-mean-square deviation(s); p.p.m., parts per million; S, angular order parameter.

[†] Present address: Department of Biophysics, Faculty of Science, Kyoto University, Kyoto, Japan.

and NOESY experiments, are given in the accompanying paper. NOESY spectra were acquired using mixing times of 30, 60, 90, 120 and 200 ms, and these were used to derive the distance constraints employed in the structure calculations as described below.

(a) Structure calculations

NOE constraints were derived from the NOESY spectra as follows. Cross-peak intensities were classified as strong, medium, weak or very weak on the basis of their appearance (number of contour levels in an evenly contoured set of spectra); strong peaks were appreciable (2 contours or more) at 30 ms mixing time, medium peaks were appreciable at 60 ms, weak peaks were appreciable at 120 ms and very weak peaks were appreciable at 200 ms. The corresponding distance ranges were set to 1.8 to 2.3 Å, 1.8 to 2.9 Å, 1.8 to 3.5 Å and 1.8 to 5.0 Å (1 Å = 0.1 nm). This calibration was based on the expectation that the strongest sequential d_{NN} cross-peaks correspond to a distance of approximately 2.3 Å, and the $d_{\text{NN}}(i, i+3)$ cross-peaks within the helices correspond to a distance of approximately 3.5 Å; the value of 5 Å for the upper limit in the very weak category is to make allowance for possible effects of spin diffusion. Pseudo-atom corrections (Wüthrich *et al.*, 1983) were applied to the upper bounds (1.0 Å for CH_2 where not stereospecifically assigned and for CH_3 , 2.0 Å for C^{β}H and $\text{C}^{\alpha}\text{H}$ of Phe and Tyr, 2.4 Å for $\text{C}^{\gamma}\text{H}_3$ of valine and $\text{C}^{\delta}\text{H}_3$ of leucine where not stereospecifically assigned), and 0.5 Å was added to the upper bound for any constraint involving a methyl group to take into account multiplicity (Wagner *et al.*, 1987; Clore *et al.*, 1987). For finger 1, the 203 distance constraints used to calculate the structure comprised 4 intraresidue NOE connectivities (all medium) relating to the orientation of aromatic rings, 78 sequential NOE connectivities (21 medium, 32 weak and 25 very weak), 36 medium-range (i to $i+2$, $i+3$, or $i+4$; 1 strong, 3 medium, 9 weak and 23 very weak), and 86 long-range (13 medium, 21 weak and 52 very weak). For finger 2, the 188 distance constraints comprised 5 intraresidue NOE connectivities (1 strong, 4 medium) relating to the orientation of aromatic rings, 79 sequential NOE connectivities (5 strong, 22 medium, 28 weak and 24 very weak), 56 medium-range (i to $i+2$, $i+3$ or $i+4$; 8 medium, 15 weak

and 33 very weak), and 47 long-range (5 medium, 17 weak and 25 very weak). These NOE constraints are summarized in Fig. 2. In addition, a small number of lower limit constraints were identified and used, all corresponding to sequential distances. In each such case, the corresponding cross-peak was expected in a position free from overlap and was either absent or of not more than 1 contour intensity in the 200 ms mixing time NOESY spectrum, and each of the signals involved exhibited at least 1 normal cross-peak elsewhere in the NOESY spectrum. These lower limit constraints were set to 3.3 Å, and comprised the sequential d_{NN} distances Phe12-Glu13, Glu13-Cys14, Cys14-Leu15, Thr22-Phe23, Ser43-Cys44, Cys44-Asp45, Ala52-Phe53 and Phe53-Val54.

In order to establish dihedral angle constraints, several vicinal $J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H})$ coupling constants were measured from the high resolution PE-COSY spectra (Mueller, 1987). A few others were estimated from the DQF-COSY spectra by comparing intensities and internal multiplet structure patterns between cross-peaks, using those measured from the PE-COSY spectra as a calibration. Those couplings measured by the latter method were set more loosely (± 2 Hz) than those measured directly (± 1 Hz). A small number of further ($\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}$) couplings were estimated to be less than 4 Hz on the basis that the corresponding cross-peaks were completely absent from the DQF-COSY spectrum under the conditions used (couplings of 4 Hz, measured as passive splittings in the PE-COSY spectrum, gave just detectable cross-peaks in the DQF-COSY spectrum); in all such cases, each of the signals involved exhibited at least 1 normal cross-peak elsewhere in the DQF-COSY spectrum, showing that the absence was not due to an atypical linewidth. The coupling constant estimates are collected in Table 1.

The program HABAS (Güntert *et al.*, 1991) was used to derive stereospecific C^{β}H assignments and χ_1 angle constraints for some residues of finger 2. The input to HABAS comprised the ($\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}$) coupling constants, together with all the intraresidue and sequential NOE constraints, and loose torsion angle constraints ($-120^\circ < \phi < -30^\circ$ and $-60^\circ < \psi < +30^\circ$) within the region His57 to His66 that was found to be helical in preliminary structure calculations. Some stereoassignments and χ_1 angle constraints derived for finger 2 were also used for the equivalent locations in finger 1, but only

Table 1
Coupling constants (in Hz) measured for SWI5 fingers 1 and 2 (peptide m10FS) at pH 6.5 and 9 °C

Residue	Coupling	Value	Coupling	Value
Tyr42	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_3)$	Absent (<4.0)	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_2)$	11.0 ± 2.0
Cys44	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_3)$	10.0 ± 1.0	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_2)$	4.0 ± 1.0
Cys49	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_3)$	10.0 ± 1.0	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_2)$	3.5 ± 1.0
Asp50	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}$ down)	6.0 ± 1.0	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}$ up)	5.0 ± 1.0
Lys51	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_2)$	Absent (<4.0)		
Phe53	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_3)$	Absent (<4.0)	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_2)$	11.0 ± 2.0
Asn56	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_3)$	10.0 ± 1.0	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_2)$	4.0 ± 1.0
His57	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}$ down)	7.0 ± 1.0	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}$ up)	9.0 ± 1.0
Asp58	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_3)$	12.0 ± 1.0	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_2)$	3.0 ± 1.0
Leu59	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_2)$	4.0 ± 1.0		
His62	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_3)$	10.0 ± 1.0	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}_2)$	8.0 ± 1.0
His66	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}$ up)	Absent (<4.0)		
Gln67	$J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H}$ down)	4.0 ± 1.0		

Stereoassignments are given for C^{β}H signals where available; where no stereoassignment has been made, signals are referred to as down(field) and up(field).

after checking that there was a close correspondence in all the relevant observed couplings and NOE enhancements. In this way, χ_1 angle constraints were obtained for Cys14, Cys19, Lys21, His32, Cys44, Cys49, Lys51, Leu59 and His62.

The NOE constraints and the χ_1 angle constraints derived from HABAS were used as the n.m.r. input to simulated annealing calculations, which were run using the YASAP protocol (Nilges, 1990) within the program XPLOR (Brünger, 1988). The version of YASAP used here comprises the following steps: (1) generation of a random starting conformation; (2) energy minimization; (3) equilibration at 1000 K for 15 ps with soft-square NOE potential and reduced van der Waals repel function; (4) slow increase of NOE potential towards square-well (increasing asymptote of NOE soft-square function) and slow increase of van der Waals repel function, carried out at constant temperature; (5) slow cooling to 300 K (25 K steps at 0.9 ps intervals) with square-well NOE potential and further increase in van der Waals repel function, and, finally; (6) energy minimization. For the NOE and dihedral constraints, the force constants were set to 50 kcal mol⁻¹ Å⁻² (1 cal = 4.184 J). The force field for the calculations comprised only geometric terms, i.e. bond lengths and angles, van der Waals repulsive terms, and planarity constraints for peptide bonds and aromatic rings. Electrostatic terms, van der Waals attractive terms and hydrogen bonding terms were not included.

In all but 1 set of calculations (see Results, below), constraints between the metal binding ligands were also used to aid the program to converge. These constraints were set to 3.1 < (HisN^e–HisN^e) < 3.5, 3.4 < (HisN^e–CysS^y) < 3.8, and 3.6 < (CysS^y–CysS^y) < 3.9 (all distances given in Å), and were included as extra entries in the list of NOE constraints with force constants of 250 kcal mol⁻¹ Å⁻²; their values were derived on the basis of tetrahedral co-ordination, using literature values of the (HisN^e–Zn) and (CysS^y–Zn) bond lengths (2.0 Å and 2.3 Å, respectively), based on EXAFS measurements with TFIPIIA (Daikun *et al.*, 1986). Following the YASAP calculations, the structures were modified by addition of a zinc atom at the average of the co-ordinates of the 4 metal ligating atoms. This was followed by a further 1000 cycles of energy minimization in the same force field as before, but with the addition of geometric constraints around the zinc as follows: bond lengths, (Zn–HisN^e) = 2.0 Å, (Zn–CysS^y) = 2.3 Å; bond angles, (HisN^e–Zn–HisN^e) = (HisN^e–Zn–CysS^y) = (CysS^y–Zn–CysS^y) = (Zn–CysS^y–CysC^b) = 109.5°, (Zn–HisN^e–HisC^c) = 126°; dihedral angle (improper constraint): (Zn–HisN^e–HisC^c–HisN^b) = 180°.

Following the generation of initial structures in this way, some further assignments were deduced. Ambiguities between alternative assignments for some NOESY cross-peaks could be resolved by inspecting the initial structures, and also a further round of stereospecific assignments was made. Thus, spin systems that showed 1 large and 1 small (C^aH, C^bH) coupling in the DQF-COSY spectra, and for which the χ_1 angle was well conserved within 1 rotamer through all the initial structures, were stereospecifically assigned on this basis. In 1 case, Asn56, the stereoassignment was based on shift data; from the initial structures it was quite clear that the high field shift (δ 1.79) of 1 of the C^bH protons must be due to its proximity to the ring of Tyr42, the spatial location of which was quite well determined. In conjunction with the observed (C^aH, C^bH) couplings, this effectively established the rotamer preference for the side-chain of Asn56, allowing the stereoassignment to be made. Other stereo-

specific assignments and χ_1 angle constraints derived at this stage comprised Phe12, Phe23, Tyr42, Phe53, Asp58 and Leu59.

Comparisons of the structures were computed relative to an average structure, which was itself computed following alignment of the structures for simultaneous multiple best fit superimposition. This multiple fit was achieved using a novel procedure developed by R. Diamond (1992). The few calculated structures that had significantly greater NOE violations than the majority were classed as outliers, and not included in calculation of the average (see Results, below). Prior to superimpositions involving side-chain atoms, particular atom names were transposed to achieve consistent labelling of symmetrically related atoms in planar groups, so avoiding the unnecessary increase of the root-mean-square deviation (r.m.s.d.) values that occurs if this is not done. This step involved transposing the labels of the following pairs of atoms as indicated (assuming the angles to be defined in the range –180° to +180°): Phe and Tyr C^{δ1}↔C^{δ2}, H^{δ1}↔H^{δ2}, C^{ε1}↔C^{ε2} and H^{ε1}↔H^{ε2}, transposed when χ_2 negative; Asp O^{δ1}↔O^{δ2}, transposed when χ_2 negative; Glu O^{ε1}↔O^{ε2}, transposed when χ_3 negative; Arg N^{η1}↔N^{η2}, H^{η11}↔H^{η21} and H^{η21}↔H^{η22}, transposed when –90° < χ_5 < +90°.

In order to assess the degree to which angles were conserved across the calculated structures, the method developed by Hyberts was employed (Detlefsen *et al.*, 1991; Hyberts *et al.*, 1992). In this method, each occurrence of a particular angle θ (e.g. ϕ , ψ or χ_1 for some particular residue, or the N–O–C angle for a putative hydrogen bond) is represented by the phase of a vector of unit length at a common origin, and the vectors are summed. The normalized magnitude of the resulting vector sum, given by $N^{-1}((\Sigma \cos \theta)^2 + (\Sigma \sin \theta)^2)^{0.5}$ where N is the number of structures, reflects the degree to which θ is conserved across the N structures. This quantity has been called the angular order parameter, abbreviated S^{angle} or S ; if θ is identical in all N structures, $S(\theta)$ has a value of 1, while if θ is completely random $S(\theta)$ is expected to take the value $N^{-0.5}$; values of less than $N^{-0.5}$ indicate anti-correlation, and $N^{-0.5}$ tends to zero for large N . The relationship between $S(\theta)$ and the standard deviation $\sigma(\theta)$ has been determined numerically by Hyberts *et al.* (1992), and is given approximately by $\sigma(\theta) = 2 \arccos(1 + 0.5 \log_{10}(S(\theta)))$.

3. Results

Figure 1 shows the sequence of the two zinc-finger peptide m10FS used for most of the present work (see accompanying paper for nomenclature), and Figure 2 summarizes the NOE constraints that were used as input to the structure calculation programs. The structural results for finger 2 will be discussed first because they form a convenient basis from which to discuss the more complicated results for finger 1.

(a) Structure of finger 2

Figure 3 shows a superimposition of the backbone atoms of 45 from a set of 50 calculated structures of finger 2 from SWI5, calculated using the YASAP simulated annealing protocol as described in Materials and Methods (Nilges, 1990). Figure 4E and F show the NOE energies, some other energy

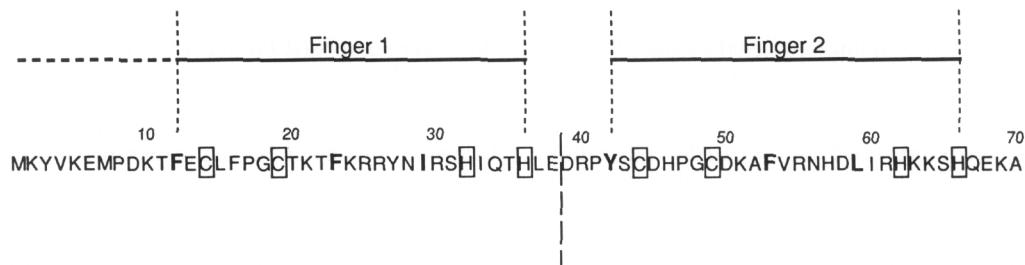


Figure 1. Amino acid sequence of the two zinc-finger construct from SWI5 used in this work (m10FS in the nomenclature of the accompanying paper). The zinc-finger motifs are defined as extending from the 1st conserved hydrophobic residue (Phe12 and Tyr42 in fingers 1 and 2, respectively) to the 4th metal binding residue (His36 and His66 in fingers 1 and 2, respectively), and the broken line over the 11 residues N-terminal to finger 1 indicates that these residues form part of the 3-dimensional structure of this particular finger (see the text for discussion). Metal-binding residues are shown boxed, and the conserved hydrophobic residues are shown in enlarged bold type. In order to compute the structures of the two domains separately, the sequence entered as input for the calculations was divided between Glu38 and Asp39, as indicated by the vertical broken line, and separate calculations carried out corresponding to each portion.

terms, root-mean-square difference from the mean (backbone atoms only), and number of violations for all 50 of these calculated structures. One of these calculations (number 50) employed a completely extended initial conformation, while the others each employed initial conformations with randomly different ϕ and ψ angles. Schematic views of the

structure having the lowest NOE energy (structure 38) are shown in Figure 5C and D.

Of the 50 calculated structures, five (4, 10, 21, 24 and 44) are clearly statistical outliers, in that their NOE energies are much higher than the spread of NOE energies amongst the others (see Fig. 4E and F; these outliers were excluded from the calculation of the average structure and the various average r.m.s.d. values throughout the paper. No tertiary NOE constraints were found for finger 2 outside the region Pro41 to His66, but for completeness these calculations included residues 39 to 70. As expected, the calculated structures show no preferred conformation in the regions Asp39 to Pro41 and Gln67 to Ala70; these regions are not displayed in the Figures, nor are they included in the comparative statistics shown in Figure 4. The average r.m.s.d. for the backbone atoms (N , C^α and C') of residues Tyr42 to His66 is $0.61 (\pm 0.19)$ Å, while for all heavy atoms of the same residues and the zinc ion the average r.m.s.d. is $1.28 (\pm 0.22)$ Å. If only the well-ordered side-chains are included (residues 42, 44, 46, 47, 49, 51, 53, 54, 56, 58, 59, 60, 62, 63 and 66; see below), this figure falls to $0.98 (\pm 0.26)$ Å. Of the energy terms included in "remaining terms" in Figure 4E, the term $E(\text{angle})$ dominates, accounting for more than 80% of the total for all the structures except 10, 21 and 44 (which are in any case excluded as outliers). The angle distortions from which this energy term arises are small and are distributed throughout the structures; apart from the outliers, the structures each have between 32 and 42 angles distorted by more than 5° relative to ideality, of which between none and two angles are distorted by more than 10° .

As mentioned in Materials and Methods, the identity of the zinc ligand atoms was checked by running further calculations in which no inter-ligand or metal constraints were used. Although the resolution of these structures was lower, they were closely similar to those calculated with metal constraints. The relevant interatomic distances are summarized in Table 2; for both fingers 1 and 2, it is quite clear

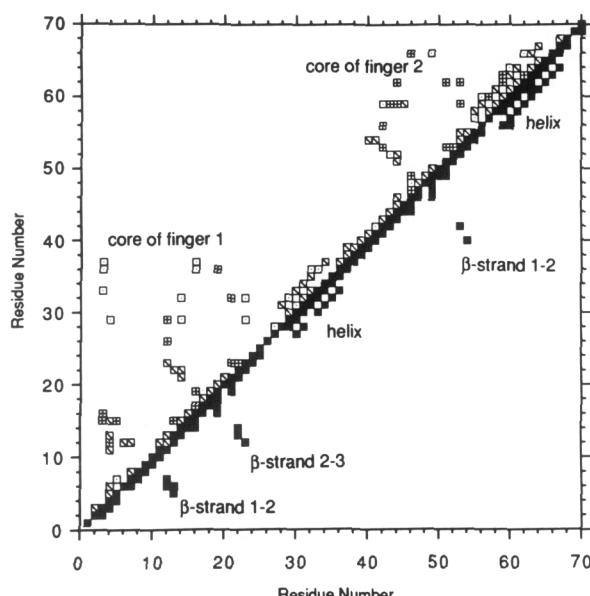


Figure 2. Diagonal plot of the NOE constraints used in the structure calculations for the two zinc-finger domains from SWI5. (■) below the diagonal signify 1 or more backbone-backbone connectivities; backbone protons are here defined as NH , C^αH and, in the case of proline residues, C^βH protons. Above the diagonal, (□) signifies 1 or more backbone-side-chain connectivities, (□) signifies 1 or more side-chain-side-chain connectivities, and (▨) signifies that both backbone-side-chain and side-chain-side-chain connectivities were observed. In the calculations described in the accompanying paper, where the two-finger domains were treated together, the sequential constraints between residues Glu38 and Asp39 were included; in the calculations described in this paper they were omitted.

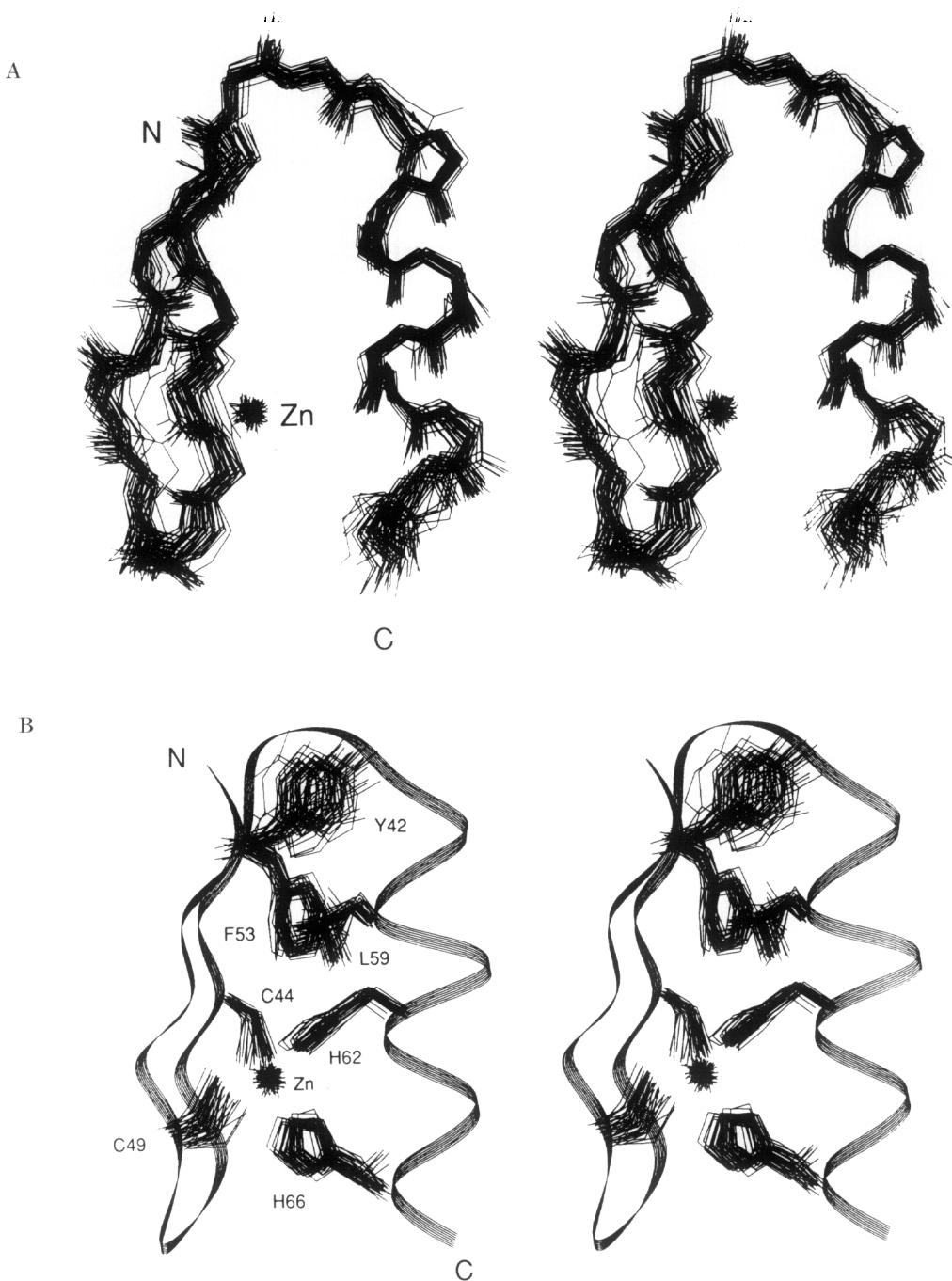


Figure 3. Stereoviews of a superimposition of 45 of the calculated structures of finger 2 of SWI5. The superimposition was carried out over the N, C^α and C' atoms of residues Tyr42 to His66, and has an r.m.s.d. to the average structure of 0.61 ± 0.19 Å. In A, the backbone atoms (N, C^α, C' and O') are shown for these same residues, while in B the side-chains of the 4 metal binding and 3 conserved hydrophobic residues are shown, together with a ribbon representation of the backbone of the structure having the lowest NOE violations (structure 38). In both A and B, the positions of the zinc ions are also shown. It is clear from these views that the structure is well-defined, as discussed in the text. The topography of the β-sheet and the transition from α- to 3₁₀-helix are clearly visible in the ribbon shown in B.

that both histidine residues co-ordinate the zinc using their N^e atoms.

The structures show the familiar architecture of the zinc-finger motif, with a β-sheet at the N terminus linked to a helix at the C terminus. The β-sheet in finger 2 of SWI5 is irregular, having only one moderately well-defined hydrogen bond (Phe53 NH to Tyr42 C=O; average H–O distance = 2.74

(±0.42) Å, N–H–O angle = 115.7°, S = 0.98 (see Materials and Methods for definition of angular order parameter, S), N–O–C angle = 138.2°, S = 0.95) conserved over most of the calculations; most of the other peptide groups in the β-sheet region have orientations that suggest they interact with solvent. There is some residual disorder within the β-sheet in the calculated structures, particularly

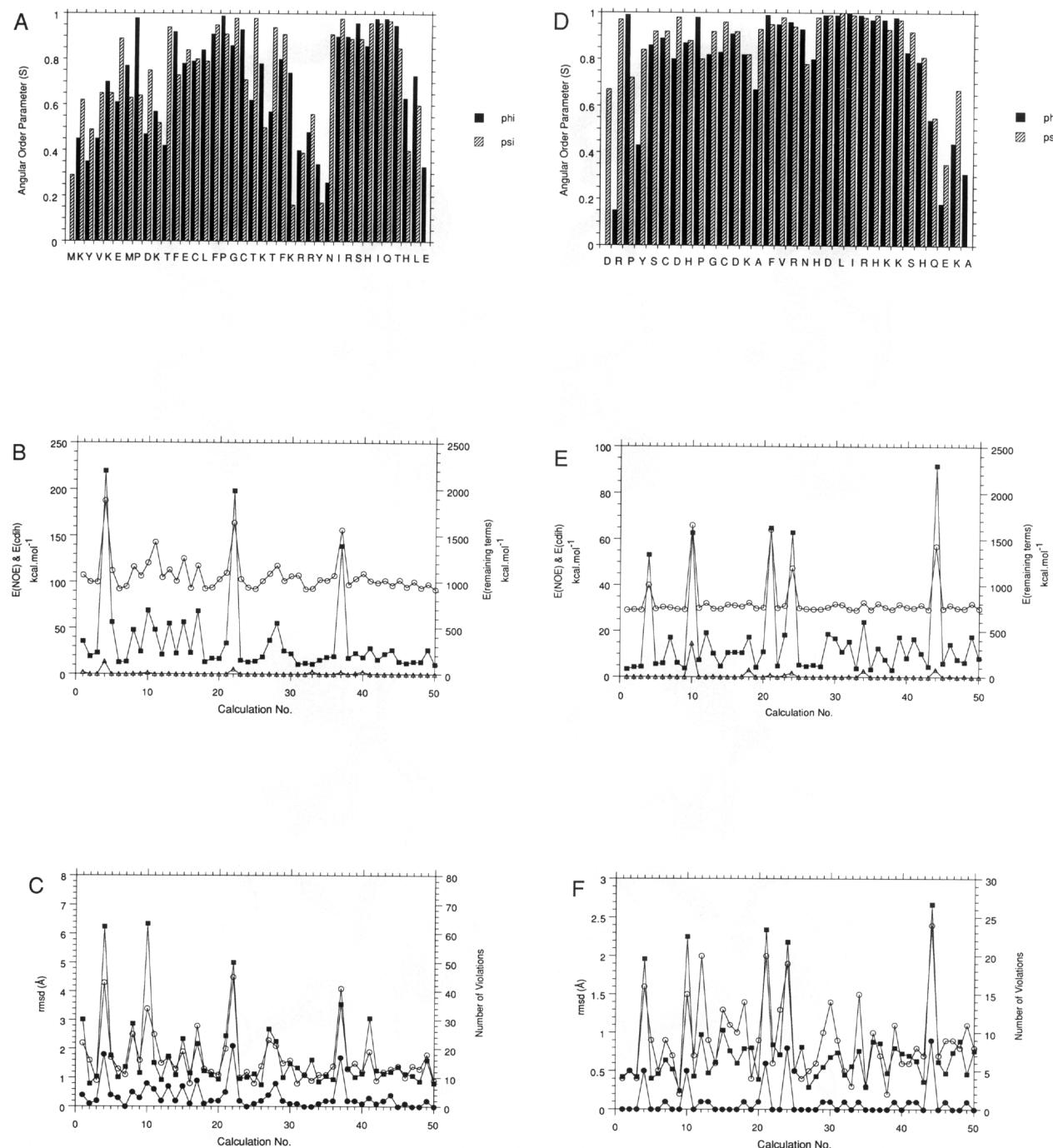


Figure 4. Plots illustrating the quality of the n.m.r. structures. Panels A, B and C relate to finger 1, panels D, E and F relate to finger 2. Panels A and D show the degree of definition of the peptide backbone for finger 1 and finger 2, respectively. Each vertical bar represents the angular order parameter (S) for the particular angle concerned over all the structures; the definition and method of calculating this parameter appears in Materials and Methods. The larger the value of S , the more closely the corresponding angle is conserved over the calculated structures. Perfect correlation corresponds to $S = 1$, whereas for a completely random distribution of a given angle over N structures, a value of $S = N^{-0.5}$ is expected (values of less than $N^{-0.5}$ correspond to anti-correlation; $N^{-0.5}$ tends to zero for large N). For finger 1 (where $N = 46$), a random distribution corresponds to $S = 0.147$, while for finger 2 ($N = 45$) the corresponding figure is 0.149. Panels B and E show, for fingers 1 and 2 respectively, the NOE energies ($E(\text{NOE})$, represented by filled squares), dihedral constraint violation energies ($E(\text{cdih})$, represented by open triangles) and the sum of the remaining energy terms ($E(\text{Total}) - E(\text{NOE}) - E(\text{cdih})$, represented by open circles) for each of the 50 calculated structures. Panels C and F show, for fingers 1 and 2 respectively, the r.m.s.d. in Å from the average (filled squares) for the backbone atoms (N , C^α and C' of residues 3 to 36 for finger 1, and of residues 42 to 66 for finger 2) for each of the 50 calculated structures, together with the number of NOE violations greater than 0.3 Å (filled circles) and 0.1 Å (open circles). As described in the text, structures 4, 10, 22 and 37 were omitted from calculations of the average structure and average r.m.s.d. figures for finger 1, while structures 4, 10, 21, 24 and 44 were similarly omitted for finger 2. Note the generally lower scales in those panels that refer to finger 2.

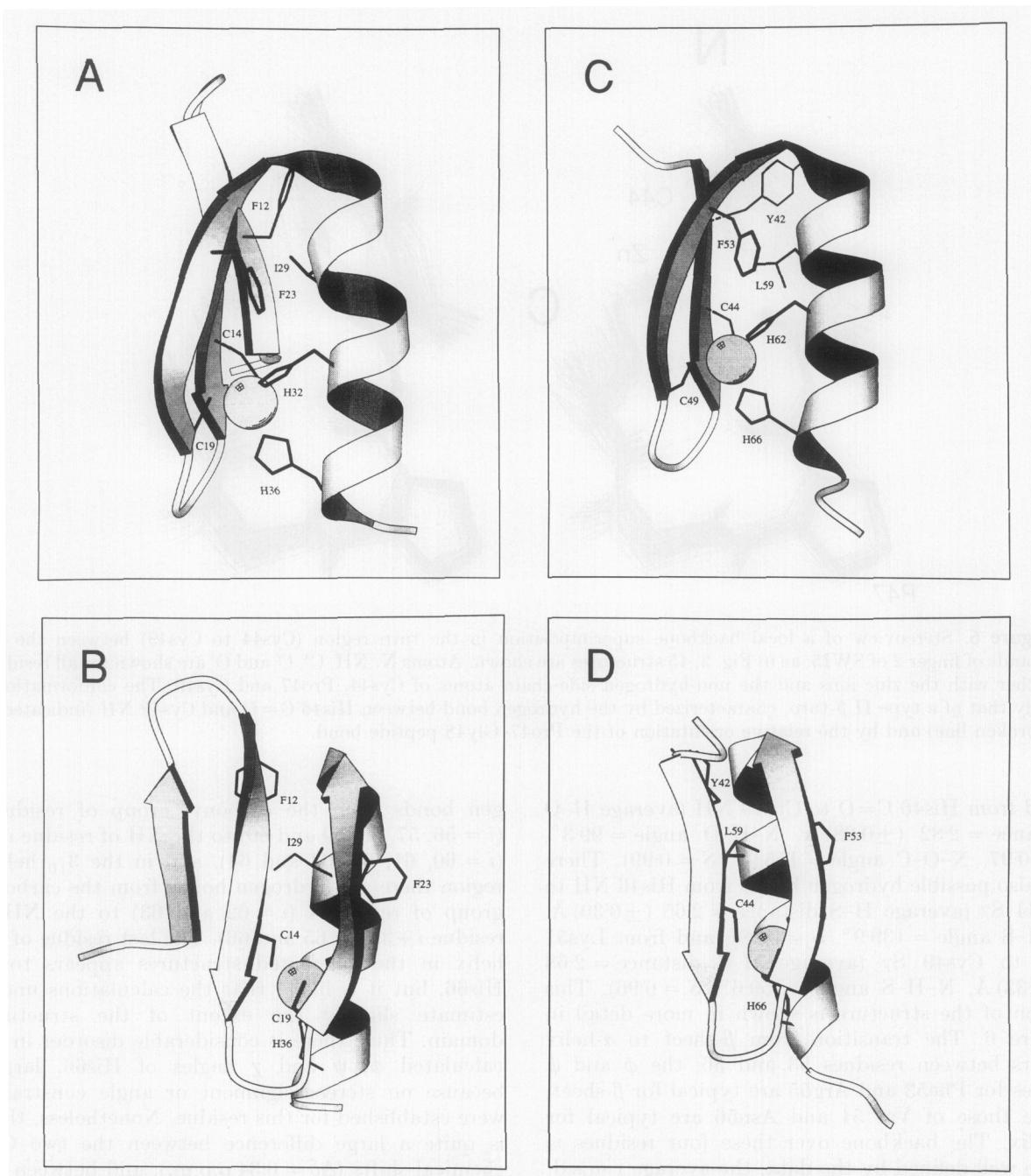


Figure 5. Schematic representation of the structures of fingers 1 (panels A and B) and 2 (panels C and D) from SWI5, prepared using the program Molscript (Kraulis, 1991). In each case, the structure having the lowest NOE violations is shown. Views A and C are related to views B and D by a 90° rotation about the vertical. In addition to the schematic representation of the backbone, bonds are shown between the non-hydrogen atoms of the side-chains of the metal binding and conserved hydrophobic residues, and these residues are labelled (note that, because the arrows representing the β -sheet do not necessarily pass through the actual C^α atom positions, the $C^\alpha-C^\beta$ bonds of some side-chains have been extended with broken lines so as to meet the backbone representation artificially).

in the region Lys51 to Ala52. As always in n.m.r. structure determination, it is difficult to establish whether such disorder is a genuine property of the molecule, or whether it merely reflects a lack of constraints in this region. In this case, however, it seems likely that the disorder is not a genuine property of the molecule, since the structures fall into two families, the minor of which (structures 12, 18, 23, 30, 34 and 46) is less plausible than the

major. Members of the minor family are associated with somewhat higher NOE energies, have positive ϕ angles for Ala52, and some (structures 18, 23 and 34) have positive ϕ angles for Lys51 also.

Between the two strands of the β -sheet, residues His46 to Cys49 form a slightly distorted type II β -turn (Pro47: $\phi = -49.5^\circ$, $S = 1.00$, $\psi = +111.3^\circ$, $S = 0.92$; Gly48: $\phi = +87.1^\circ$, $S = 0.94$, $\psi = +49.2^\circ$, $S = 0.95$), with a somewhat ill-defined hydrogen

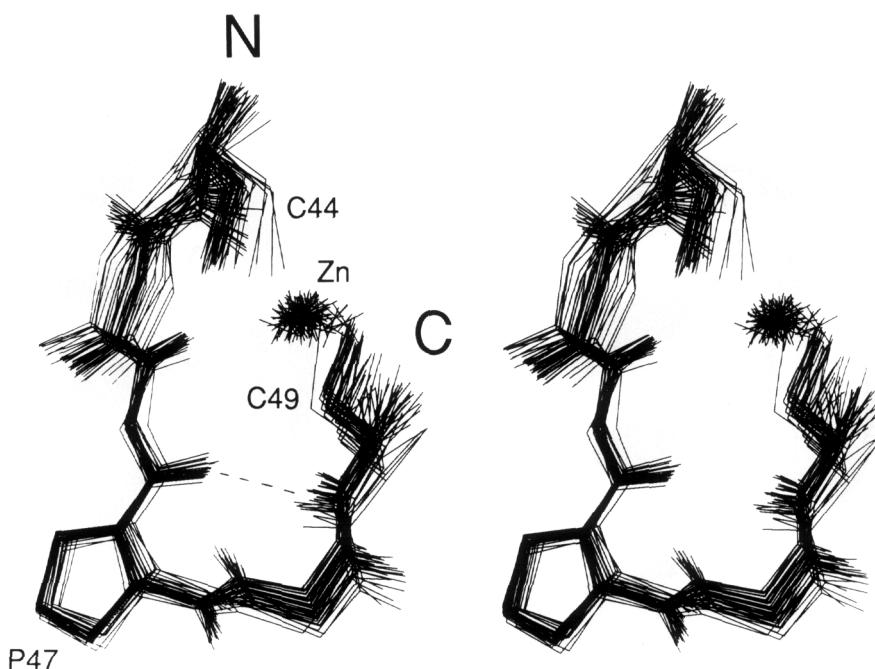


Figure 6. Stereoview of a local backbone superimposition in the turn region (Cys44 to Cys49) between the two β -strands of finger 2 of SWI5; as in Fig. 3, 45 structures are shown. Atoms N, NH, C^α , C' and O' are shown for all residues, together with the zinc ions and the non-hydrogen side-chain atoms of Cys44, Pro47 and Cys49. The conformation is clearly that of a type II β -turn, characterized by the hydrogen bond between His46 C=O and Cys49 NH (indicated by the broken line) and by the relative orientation of the Pro47-Gly48 peptide bond.

bond from His46 C=O to Cys49 NH (average H–O distance = 2.82 (± 0.28) Å, N–H–O angle = 99.3°, $S = 0.97$, N–O–C angle = 155.0°, $S = 0.99$). There are also possible hydrogen bonds from His46 NH to Cys44 Sy (average H–S distance = 2.65 (± 0.30) Å, N–H–S angle = 139.9°, $S = 0.98$), and from Lys51 NH to Cys49 Sy (average H–S distance = 2.68 (± 0.33) Å, N–H–S angle = 159.6°, $S = 0.96$). This region of the structure is shown in more detail in Figure 6. The transition from β -sheet to α -helix occurs between residues 53 and 56; the ϕ and ψ values for Phe53 and Arg55 are typical for β -sheet, while those of Val 54 and Asn56 are typical for α -helix. The backbone over these four residues is quite well-defined by the data, the average r.m.s.d. from the average for the N, NH, C^α , C^α H, C' and O atoms of these four residues being 0.26 (± 0.16) Å.

The helical region is well defined in the n.m.r. structures, and shows a transition between α -helix and 3_{10} helix at about Arg61, the carbonyl oxygen atom of which appears to have no hydrogen bonding partner. Since there are almost no systematic differences between the relevant observed n.m.r. parameters (${}^3J(\text{NH}, \text{C}^\alpha\text{H})$, $d_{\alpha\text{N}}(i, i+2)$, $d_{\alpha\text{N}}(i, i+3)$ and $d_{\alpha\text{N}}(i, i+4)$) in the two regions of the helix (for either finger 2 or finger 1), we conclude that this transition occurs in the calculated structures as a result of the geometrical requirements of zinc binding. Characteristic hydrogen bonds are found along the length of the helix, with the exception of Arg61, whose carbonyl group protrudes from the helix in the region of transition from α to 3_{10} helix. Thus, in the α -helical region there are hydro-

gen bonds from the carbonyl group of residue i ($i = 56, 57, 58, 59$ and 60) to the NH of residue $i+4$ ($i = 60, 61, 62, 63$ and 64), and in the 3_{10} helical region there are hydrogen bonds from the carbonyl group of residue i ($i = 62$ and 63) to the NH of residue $i+3$ ($i = 65$ and 66). The last residue of the helix in the calculated structures appears to be His66, but it is likely that the calculations underestimate slightly the extent of the structured domain. Thus, there is considerable disorder in the calculated ϕ , ψ and χ angles of His66, largely because no stereoassignment or angle constraints were established for this residue. Nonetheless, there is quite a large difference between the two C^βH chemical shifts ($\Delta\delta = 0.34$ p.p.m.), and between the two ${}^3J(\text{C}^\alpha\text{H}, \text{C}^\beta\text{H})$ coupling constants, suggesting that in reality the residue may have a well-defined conformation, and it is also likely that the helix extends for one further residue, as there are NOE connectivities between Gln67 and Lys64 (Lys64 C^αH to Gln67 NH, C^βH_2 and $\text{C}^\gamma\text{H}_2$), although these are of low intensity.

The absolute chirality of the tetrahedral zinc ion may be designated S , following Berg's (1988) convention that the ligand priorities be assigned according to sequence order, i.e. Cys44 > Cys49 > His62 > His66. This is the same chirality as was proposed by Berg, and has been found so far for all other zinc-fingers of the TFIIB class. As pointed out previously (Neuhaus *et al.*, 1990), residues Tyr42, Phe53, Leu59 and His62 form a miniature hydrophobic core to the structure, and all four are well-defined by the n.m.r. data. It is noticeable that

Table 2.

Distances between potential zinc-liganding atoms in a series of calculated structures for zinc-fingers 1 and 2 of SWI5 in which no metal-ligand or interligand constraints were used

Ligand atoms	Distance (Å)	Ligand atoms	Distance (Å)
<i>A. Finger 1</i>			
C14 S ^γ -H32 N ^ε	3.07-3.94 mean = 3.63	C14 S ^γ -H32 N ^δ	5.06-5.96 mean = 5.50
C14 S ^γ -H36 N ^ε	3.22-4.13 mean = 3.73	C14 S ^γ -H36 N ^δ	4.32-6.91 mean = 5.40
C19 S ^γ -H32 N ^ε	3.26-4.30 mean = 3.78	C19 S ^γ -H32 N ^δ	4.07-6.27 mean = 5.46
C19 S ^γ -H36 N ^ε	2.90-4.14 mean = 3.88	C19 S ^γ -H36 N ^δ	4.45-6.12 mean = 5.55
<i>B. Finger 2</i>			
C44 S ^γ -H62 N ^ε	3.22-3.94 mean = 3.43	C44 S ^γ -H62 N ^δ	5.36-6.07 mean = 5.56
C44 S ^γ -H66 N ^ε	3.00-3.83 mean = 3.40	C44 S ^γ -H66 N ^δ	4.24-5.83 mean = 4.90
C49 S ^γ -H62 N ^ε	3.18-4.02 mean = 3.58	C49 S ^γ -H62 N ^δ	4.37-5.67 mean = 4.82
C49 S ^γ -H66 N ^ε	3.28-3.88 mean = 3.54	C49 S ^γ -H66 N ^δ	4.07-5.88 mean = 5.24

For each pair of atoms, the range of distances found is reported together with the corresponding mean value. The expected S-N distance for tetrahedral co-ordination of 2 sulphur and 2 nitrogen atoms to a zinc atom is approximately 3.6 Å.

the side-chain of Tyr42 is not appreciably more disordered than those of the other three residues, showing that the side-chain itself participates in the structure; NOE interactions are seen from Tyr42 to residues Phe53, Asn56 and Leu59.

Several other residues have well-ordered side-chains in the calculated structures, at least as far as the χ_1 angle. Apart from Cys44 and Cys49 (the two metal-binding cysteine residues) and the proline in the turn (Pro47), these include His46, Lys51, Val54, Asn56, Asp58, Ile60 and Lys63. The side-chain of His46 packs against the side-chains of Cys49 and His66, while the side-chain of Lys51 contacts the rings of Phe53 and His62. Val54 and Asn56 both contact the ring of Tyr42, and the side-chain of Asp58 appears from the calculations to be packed against the ring of Phe53 and the backbone amide link 53-54. It is particularly noticeable that the side-chain of Lys63 is the only one of a Lys or Arg residue (in either finger 2 or finger 1) to show a measurable shift difference between the two diastereotopic methylene protons of the terminal CH₂ group ($\Delta\delta(\text{C}^{\beta}\text{H}_2) = 0.14$ p.p.m.). This large shift difference implies that the N^εH₂ group is probably involved in an interaction, either as a hydrogen bond donor or in some looser interaction such as a salt bridge, but the other partner for this interaction was not identified. Further light may be cast on this issue by the structure determination of finger 3, where the corresponding residue (Arg89) shows a similar shift dispersion between the C^βH₂ signals.

Most of the other residues have largely disordered side-chain conformations in the calculated structures, and for several of these the spectra themselves show direct evidence for conformational

averaging. Thus, the C^βH signals for Asp50 and His57 show relatively sharp lineshapes and nearly averaged values for the $^3J(\text{C}^{\alpha}\text{H}, \text{C}^{\beta}\text{H})$ coupling constants, while there is little or no diastereotopicity between the C^βH signals of Arg55, Arg61, Lys64 and Ser65. Each of these side-chains projects away from the structured domain of the peptide into solution.

(b) Structure of finger 1

Figure 7 shows a superimposition of the backbone atoms of 46 out of a set of 50 calculated structures of finger 1 from SWI5, calculated in exactly the same way as for finger 2. Figure 4B and C show the NOE energies, some other energy terms, r.m.s.d. from the mean (backbone atoms only), and number of violations for all 50 of these calculated structures. As before, one of these calculations (number 50) employed a completely extended initial conformation, while the others each employed initial conformations with randomly different ϕ and ψ angles.

Three of the structures (4, 22 and 37) are clear outliers in terms of their NOE energies. Also, structure 10 differs grossly from the others in the arrangement of the β -sheet relative to the helix. The β -sheet in structure 10 corresponds approximately to the mirror image of the β -sheet in all the other structures (formed by reflection in the plane perpendicular to the vertical axis of Fig. 5B), with the result that the chirality of zinc binding in structure 10 is reversed. This arrangement is only seen in structure 10 for finger 1, is not seen at all for finger 2 (where the n.m.r.-derived constraint list is more complete, see below), and is not seen for any of the other zinc-finger structures of the TFIIB class so

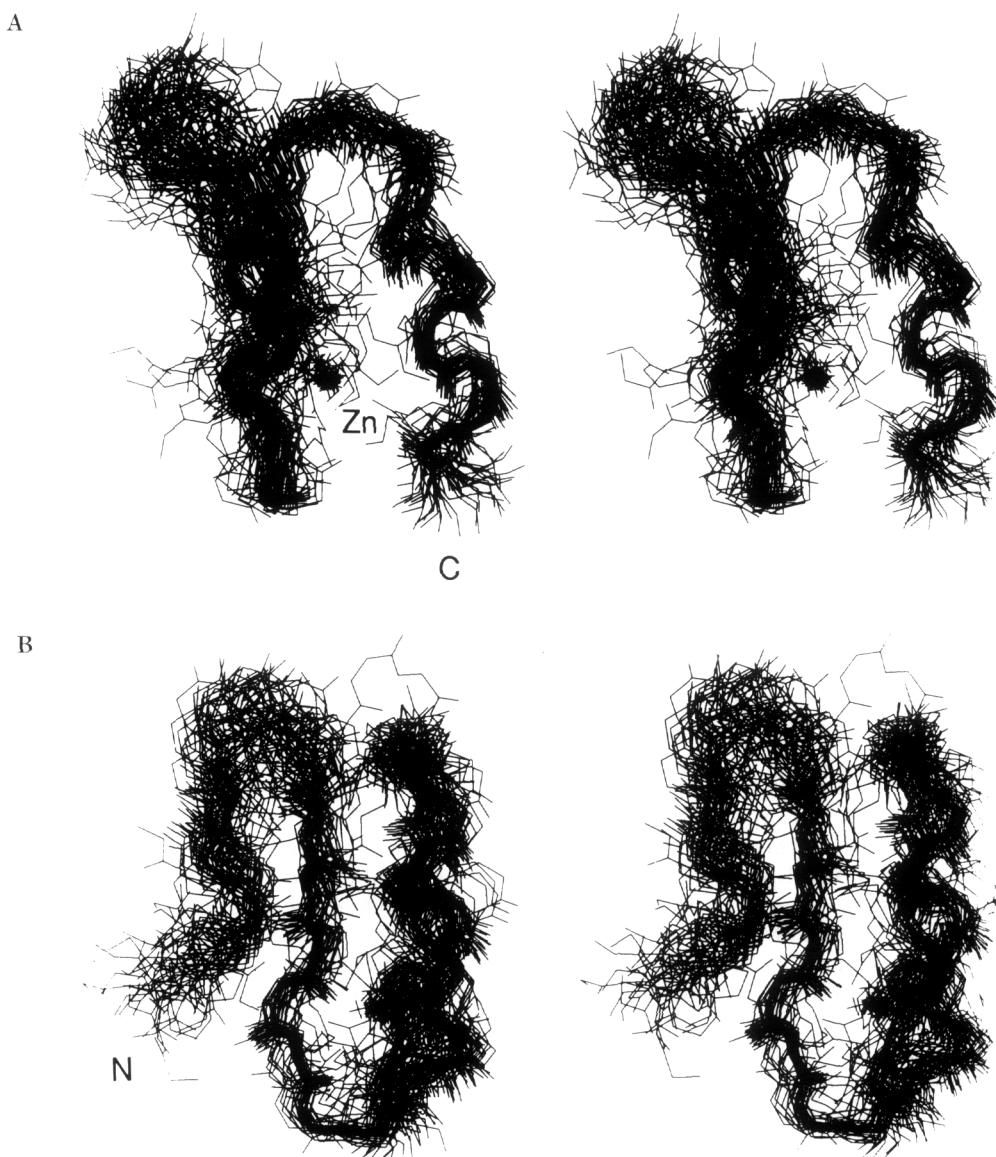


Figure 7. Stereoviews of a superimposition of 46 of the calculated structures of finger 1 of SWI5. The superimposition was carried out over the N, C^α and C' atoms of residues Tyr3 to His36, and has an r.m.s.d. from the average structure of 1.41 (± 0.61) Å; as discussed in the text, this falls to 0.84 (± 0.27) Å if 4 further outliers are excluded and just the region Phe12 to His36 is considered. The backbone atoms (N, C^α, C' and O') are shown for residues Met1 to Leu37, and the positions of the zinc ions of finger 1 are also shown. A, View is equivalent to that shown for finger 2, and principally emphasizes the helix, the β -sheet being viewed edge-on (the N terminus is not clearly visible in this view). The diffuse, disordered region at the top left of this view comprises the turn between the novel N-terminal strand of the β -sheet and the central strand. Although this is very clearly the least well-defined portion of the structure, the disorder arises mainly from "hinge" motions of the turn region with respect to the rest of the structure; locally its conformation is somewhat better defined, being a type I β -turn (see the text). In B the helix and C terminus are largely hidden from view behind the C-terminal strand of the β -sheet, thereby illustrating principally the backbone definition within the triple-stranded β -sheet. Again, it is clear that the N-terminal strand of the sheet is the least well-defined relative to the core of the structure.

far determined. It seems reasonable, therefore, to discard structure 10 as well as the three outliers mentioned above, so that, in all, structures 4, 10, 22 and 37 were excluded from calculation of the average structure and the various average r.m.s.d. figures.

The average r.m.s.d. for the backbone atoms (N, C^α and C') of residues Tyr3 to His36 is 1.41 (± 0.61) Å, while for all heavy atoms of the same

residues and the zinc ion the average r.m.s.d. is 2.25 (± 0.68) Å. Of the energy terms included in "remaining terms" in Figure 4B, the term $E(\text{angle})$ dominates, accounting for more than 80% of the total for all the structures. As was the case for finger 2, the corresponding angle distortions are small and widely distributed; apart from the outliers, the structures all have between 46 and 60 angles distorted by more than 5° relative to ideality, of

which between none and eight are distorted by more than 10°.

The relatively low resolution of these structures for finger 1 is a direct result of the difficulties described in the assignments section of the accompanying paper, which in turn caused the constraint list for finger 1 to be considerably more sparse than that for finger 2. Because the NH resonances of Lys24, Arg25, Arg26 and Asn28 are very weak, possibly due to local catalysis of their exchange with solvent caused by the run of three basic residues, very few constraints could be identified in the region Lys24 to Asn28. Similar problems arose due to the absence of assignments for the NH, C^aH and C^bH protons of Tyr3. A more general problem was that the analysis of NOESY cross-peaks involving finger 1 could only be made using spectra of the two-finger construct, whereas for finger 2, assignments were made using spectra of both two-finger and single finger constructs. For instance, for finger 1 there were very few constraints that could be identified involving residues Asp9 to Thr11.

In view of this, it seems reasonable that the poorer resolution of the calculated structures of finger 1 does not reflect a genuinely higher degree of disorder or flexibility than for finger 2, but is simply a consequence of the lack of constraints in some parts of the structure. In order to test this further, NH signals for corresponding sequence locations in each finger were compared by examining F_2 cross-sections through appropriate NOESY and COSY cross-peaks. No lineshape differences between any such pairs were apparent (excepting those residues in finger 1 for which no NH was identified), supporting the view that there is no major intrinsic difference in the dynamics of the two-fingers.

Schematic views of the structure having the lowest NOE energy (structure 50) are shown in Figure 5A and B. The most noticeable feature of the structure of finger 1 is that it has a third strand of β -sheet at its N terminus, in addition to the expected features of a zinc-finger already seen in finger 2. Thus, the first seven residues fold against the central strand of the sheet (residues 10 to 16), and also make some contacts with the helix. There is a turn involving residues 7 to 10, probably stabilized by an ill-defined hydrogen bond from Met7 C = O to Lys10 NH (average H–O distance = 3.16 (± 1.25) Å, N–H–O angle = 131.3°, S = 0.93, N–O–C angle = 112.1°, S = 0.89). The calculated structures show considerable disorder in this region (see below), and it is not possible to make a definite classification of the turn type, but it appears likely to be a type I β -turn (Pro8: $\phi = -57.9^\circ$, S = 0.98, $\psi = -8.4^\circ$, S = 0.62; Asp9: $\phi = -128.2^\circ$, S = 0.45, $\psi = +6.0^\circ$, S = 0.75). Interestingly, Glu6 has a positive ϕ angle ($\phi = +11.6^\circ$, S = 0.62, $\psi = +86.1^\circ$, S = 0.90), approximately conserved in most of the calculations. Overall, the triple stranded β -sheet region is not regular, and although the local structure is not well enough defined to identify hydrogen bonding partners with any certainty, it seems likely that the

stabilisation of this region probably arises mainly from hydrophobic interactions within the somewhat enlarged core. In particular, the side-chains of the two hydrophobic residues in the new N-terminal strand show many NOE contacts to other hydrophobic side-chains: Tyr3 contacts Val4, Leu15, Phe16, Ile33, His36 and Leu37, while Val4 contacts Phe12, Leu15 and Ile29.

Most of the differences between the calculated structures occur in the arrangement of the N-terminal strand of the β -sheet; if the structures are superimposed only in the region corresponding to the classical zinc-finger motif, that is from Phe12 to His36, then the average r.m.s.d. from the backbone over these residues only, falls from 1.41 (± 0.61) Å to 0.99 (± 0.56) Å. However, this figure is still inferior to that seen for finger 2. This is mostly due to the sparsity of constraints involving the NH signals of residues Lys24 to Arg26 and Asn28, which has the result that there is considerable local disorder in the backbone angles over this region (see Fig. 4A). Disorder in this region also propagates over nearby parts of the calculated structures, so that several of the structures differ markedly in the arrangement of the chain between Phe23 and the beginning of the helix; structures 1, 21, 27 and 41 are very different from the others, and from the structure of finger 2, in this region, and if they are excluded from the comparison, the average r.m.s.d. from the backbone between Phe12 and His36 falls further to 0.84 (± 0.27) Å.

Notwithstanding the lower resolution in finger 1, it is possible to say that the structures of finger 1 and finger 2 are closely similar over the regions corresponding to the classical zinc-finger motif (Phe12 to His36 in finger 1 and Tyr42 to His66 in finger 2). For instance, the turn between the two metal binding cysteine residues is again a type II β -turn in finger 1 (Pro17 $\phi = -52.5^\circ$, S = 0.99, $\psi = +125.4^\circ$, S = 0.92; Gly18 $\phi = +91.3^\circ$, S = 0.89, $\psi = -22.5^\circ$, S = 0.98), with a hydrogen bond between Phe16 C = O and Cys19 NH (average H–O distance = 2.09 (± 0.25) Å, N–H–O angle = 140.5°, S = 0.98, N–O–C angle = 138.8°, S = 0.98). This similarity is not only reflected in the calculated structures, but also in many of the chemical shifts.

4. Discussion

(a) The additional β -strand in finger 1

The most unexpected result presented here is that several residues N-terminal to the first zinc-finger of SWI5 fold back to become involved in the structure, forming an additional strand to the β -sheet of finger 1. This is demonstrated by many side-chain-side-chain and side-chain-backbone NOE cross-peaks, and by backbone-backbone NOE interactions linking Lys5 NH with Glu13 NH, Glu6 NH with Thr12 C^aH, Glu6 C^aH with Thr12 C^aH, Glu6 C^aH with Glu13 NH, and Met7 NH with Thr12 C^aH. Nonetheless, it is perhaps stretching a point to

describe this new feature as a β -strand, since it appears to have very few, if any, of the backbone-backbone interstrand hydrogen bonds that would be expected in a regular β -sheet, other than that expected and found in the tight turn (Met7 C = O ... Lys10 NH). Rather, the new N-terminal strand seems to be stabilised largely by hydrophobic side-chain-side-chain interactions, with Tyr3 contacting Leu15, Phe16, Ile33, His36 and Leu37, and Val4 contacting Thr11, Phe12, Leu15 and Ile29. However, this does not really distinguish the new N-terminal strand from the other two, since these also lack a backbone-backbone hydrogen bonding network away from the tight turns (both in finger 1 and in finger 2), and are also apparently stabilised largely by hydrophobic interactions, together with the participation of the two cysteine residues in zinc binding.

It is of course impossible to say whether the residues N-terminal to finger 1 form an additional strand to the β -sheet in intact SWI5. However, as described in the accompanying paper, when residues Asp1 to Glu6 are missing, finger 1 is not folded under the n.m.r. conditions (10 to 27°C, pH 6.5, approx. 2 mM-protein in 40 mM-pyrophosphate). Also, there is an increasing body of circumstantial evidence that points to structural involvement of residues formally N-terminal to the first zinc-finger in several multi-zinc-finger proteins. For at least two other zinc-finger proteins, peptides spanning the first two zinc-finger motifs only bind specifically to DNA if ten or so residues are included N-terminal to the first conserved hydrophobic residue of the first finger. The proteins in question are Tramtrack, a repressor protein from *Drosophila* that contains two zinc-finger motifs (Fairall *et al.*, 1992), and ADR1, a yeast transcription factor that also contains two zinc-fingers (Thukral *et al.*, 1991). Also, a deletion mutant of SWI5 in which only residues Leu529 to Val541 are missing (residues -8 to +4 in the present numbering scheme) is completely inactive in its normal role as a transcription factor (K. Nasmyth, personal communication); it remains to be seen if this is because its ability to bind DNA is reduced.

On the other hand, it is also fairly clear that not all zinc-finger proteins necessarily exhibit this phenomenon. For instance, sequence-specific binding of peptides including finger 1 of TFIIIA is not affected by deletion of residues N-terminal to the first finger (X. Liao & P. Wright, personal communication). Similarly, the three-zinc-finger peptide from Zif268 employed in a recent crystallographic study of a zinc-finger-DNA complex does not include sufficient residues N-terminal to the first zinc-finger for an additional β -strand to be formed, but nonetheless this peptide is clearly capable of binding to DNA (Pavletich & Pabo, 1991).

It will be interesting to see whether further examples similar to SWI5 emerge, in which the first zinc-finger motif in a series of adjacent fingers contains an additional N-terminal β -strand in its structure.

(b) Comparison of the SWI5 zinc-finger structures with others

Apart from the additional β -strand at the N terminus of finger 1, the greatest difference between the present structures and those published previously for other zinc-finger peptides arises through a difference in the primary sequence in the region between the two metal-binding cysteine residues. In both fingers 1 and 2 of SWI5, the two cysteine residues are separated by four residues, and in both cases the calculations show unequivocally that this region of the structure adopts a somewhat distorted type II β -turn conformation (see Results). The only other structure so far determined for a zinc-finger peptide containing a Cys-X₄-Cys loop is for one of three fingers of Zif268 (Pavletich & Pabo, 1991), but the loop itself was largely disordered in this crystal structure. All of the other published zinc-finger structures are of peptides that possess the more common Cys-X₂-Cys arrangement, where a variety of different conformations have been described in the different cases.

Another, smaller, difference amongst the various structures concerns the β -sheet (disregarding the new N-terminal strand of SWI5 finger 1). The only definitely identified backbone-backbone hydrogen bonds in this region of the SWI5 structures involve the two conserved aromatic residues (Phe12 and Phe23 in finger 1, and Tyr42 and Phe53 in finger 2), and the $i \leftarrow i+3$ hydrogen bond of the β -turns. Overall, the two strands form a more open structure than a regular β -sheet, with most of the remaining hydrogen bond donors and acceptors accessible to solvent. Stabilization of the structure presumably results mainly from hydrophobic interactions within the core and from zinc binding, but there may also be hydrogen bonds of the type Cys(*i*) S⁷ \leftarrow NH of residue (*i*+2) (see Results), analogous to those predicted by Berg (1988) and found for Zif268. Although the details differ due to sequence differences, this more open backbone arrangement most closely resembles that described for human enhancer binding protein. For Xfin31 (Lee *et al.*, 1989a,b), ZFY6 (Kochyan *et al.*, 1991a,c) and the three fingers of Zif268, the β -sheet is apparently more regular, although the number of additional interstrand hydrogen bonds is only one or two, while for ADR1b (Klevit *et al.*, 1990), the sheet appears again to be irregular.

The arrangement of the side-chains within the hydrophobic core of SWI5 is most similar to that seen in Xfin31 and Zif268, but this probably reflects the greater sequence similarities in these cases. Of the other structures, ZFY6 has a variant consensus sequence in which the second conserved hydrophobic residue occurs two, rather than four, residues beyond the second cysteine, and human enhancer binding protein has a Phe residue in both locations, while for ADR1b the side-chain positions vary significantly amongst the structures.

The helix in the SWI5 fingers, best resolved in

finger 2, is very similar to those found in Xfin31 and Zif268, each helix having N and C termini at almost exactly matching sequence locations. More significantly, all three cases show a transition from α to 3_{10} helix one or two residues N-terminal to the first metal-binding histidine (in Zif268, the helix returns to α -helix for its last 2 residues). This transition seems to be necessary to bring the two metal-binding histidine rings into positions suitable for their interaction with the zinc, and appears to be a characteristic of zinc-finger sequences having a His-X₃-His arrangement. Interestingly, of the crystal structures used by Berg (1988) to predict that the His-X₃-His moiety occurred within a helix, at least one (the only one to bind zinc) also adopts a 3_{10} helical conformation in the corresponding region. Residues Val139 to Asp150 of thermolysin form a helix including two histidine residues (His142 and His146) bound to a common zinc, and residues Ala141 and Leu144 to His146 adopt ψ angles between -26° and -33° , characteristic of 3_{10} helix, while residues His142 and Glu143 adopt intermediate conformations ($\psi = -45.6^\circ$ and -43.3° , respectively; for Ala141 to His146, the ϕ angles range from 64.4° to 68.7°) (Holmes & Matthews, 1982).

For ADR1b, which also has a His-X₃-His arrangement, it is unclear whether a 3_{10} helix exists in the calculated structures; no characteristic NOE enhancements were found that might distinguish α from 3_{10} helix (i.e. $d_{\alpha N}(i, i+2)$ and $d_{\alpha N}(i, i+4)$ cross-peaks in NOESY), but as the SWI5 case demonstrates, this does not mean that 3_{10} helix was absent. In the remaining zinc-finger structures, the metal-binding histidine residues are more widely separated; ZFY6 has an arrangement His-X₄-His, and human enhancer binding protein has an arrangement His-X₅-His. In both these cases, the α -helix continues to a point between the two metal-binding histidine residues, and some other geometry is adopted at the C terminus of the finger to bring the second histidine close to the zinc.

When making comparisons of this sort between the n.m.r. structures, it is important to remember that some differences of detail may well be artifacts of the process of solving the structure. Although the various structures may be very similar, the n.m.r. spectra themselves must differ considerably, since the sequence of each peptide is different. Thus, the particular problems of overlap and resolution of the n.m.r. signals, that in practice dictate which NOESY cross-peaks can be assigned and which cannot, will be completely different for each peptide, necessarily resulting in rather different constraint lists in each case. Such discrepancies between the constraint lists, which may be purely spectroscopic in origin and unrelated to conformation, could be the true origin of some of the smaller differences discussed above, particularly in such parameters as the precise N and C termini of the helix, the detailed orientation of the peptide groups in the β -sheet, and the location of some of the less well-defined hydrogen bonds.

(c) Implications for DNA-binding of SWI5

Now that a crystal structure exists for a zinc-finger peptide from Zif268 complexed with its DNA recognition element (Pavletich & Pabo, 1991), it is interesting to compare this with the SWI5 structures described here, in order to see how similar the mode of DNA binding is likely to be.

As discussed, the structural details of the zinc-finger domains from SWI5 and Zif268 are very similar. Furthermore, the majority of the side-chains of Zif268 that are involved in DNA contacts have easily identified analogues in the SWI5 structures, suitably located spatially to carry out equivalent roles on DNA binding. In the DNA complex of Zif268, the helices of the zinc-finger domains are located in the major groove of the DNA, with their N termini at the base of the groove, and their C termini near the top of the wall of the groove. Each helix makes two specific contacts with the bases of the DNA, one using residues near the N terminus of the helix, the other from near the mid-point of the helix, and in addition each helix makes a non-specific contact with the DNA phosphate backbone using the N^δ atom of the first metal-binding histidine. In the SWI5 structures, the N^δ atoms of His32 and His62 are solvent accessible, and are oriented analogously to their counterparts in Zif268, so it seems likely that they fulfil similar roles on DNA binding. Similarly, the specific contact from the mid-point of each helix seems most likely to be made by Asn28 or Arg30 in finger 1 of SWI5, and by Arg61 in finger 2.

For the specific contact involving the N terminus of the helix, more subtle, though more speculative, parallels are possible. In Zif268, these contacts all involve a partial sequence Arg-Ser-Asp. The guanidinium group of the arginine is hydrogen-bonded on one side to N-7 and O-6 of a guanine of the DNA, but it is also hydrogen-bonded on the other side to the carboxylate group of the Asp. This forms a so-called "buttressed" interaction with the DNA, in which four out of five of the nitrogen bound hydrogen atoms of the guanidinium group act as hydrogen bond donors. In the corresponding locations for fingers 1 and 2 of SWI5, the arginine residues are both present (Arg25 and Arg55, respectively), but the Ser and Asp residues are not. However, in each case the Asp residue is replaced by another side-chain capable of acting as a hydrogen bond acceptor, namely that of Tyr27 in finger 1, and that of His57 in finger 2. If these rather different side-chains were to participate in buttressed interactions together with the arginine residues, this might be one basis for modulation of the DNA binding specificity of the zinc-finger motif. Not only would the different lengths of these alternative side-chains such as Tyr and His lead to a different spatial position for the arginine in the buttressed arrangement, but the different nature and number of hydrogen bond acceptors in the aromatic rings of Tyr and His would further modify the electronic character of the guanidinium group of the arginine.

Presumably, the His residue could only participate in such an arrangement if the imidazole ring were present in its uncharged, monoprotonated state.

Of the other DNA contacts made by the zinc-finger domains of Zif268, the only one preserved in all three fingers involves the side-chain of an arginine residue located two positions beyond the second metal-binding cysteine, in the second strand of the β -sheet. This contacts the phosphate backbone of the DNA, and it is very likely that the side-chains of Lys21 and Lys51 fulfil analogous roles in fingers 1 and 2 (respectively) of SWI5. There are a total of four other contacts between Zif268 and its DNA recognition element, but as none of these occurs in more than one of the three fingers, it is difficult to draw any parallels with SWI5 in these cases.

Overall, then, it appears that there is likely to be quite a close parallel between the Zif268-DNA complex and the as yet unknown structure of the SWI5-DNA complex. Given the similar locations of, and interrelationships between, the probable DNA binding groups, it is likely that the overall architecture of the complex is also similar, with the three fingers wound around the major groove of the DNA. If this is the case, what is likely to be the significance of the additional β -strand in the structure of finger 1? Clearly, no answer exists to this question at present, but, if the complex does indeed have a similar architecture to that of Zif268, then the additional N-terminal β -strand would come close to the opposite strand of the DNA (the top strand, in the case of SWI5), bringing it into a suitable position to make contacts with the phosphate backbone. Strong support for this hypothesis is provided by DNA-binding data for Tramtrack, one of the other zinc-finger proteins referred to earlier, which requires additional N-terminal residues in order to bind to DNA. Experiments to establish the minimum DNA recognition site for Tramtrack have shown that its two zinc-finger motifs require at least 11 base-pairs for binding, and that finger 1 appears to contact both DNA strands at points facing one another across the major groove (Fairall *et al.*, 1992). Thus, the role of the third β -strand may be to provide the zinc-finger domain with an additional means of contacting the DNA.

Clearly, another important consequence of the presence of the additional β -strand is that the relationship between the entry and exit points of the finger domain is altered. Normally, the chain enters and leaves from opposite ends of the domain, whereas in finger 1 of SWI5 the entry point is more nearly at the same end of the domain as the exit point. This may be related to the fact that residues N-terminal to finger 1 presumably form part of other domains unrelated to DNA binding, the structural relationship of which to finger 1 may be quite different to that between adjacent fingers.

In conclusion, we have presented the three-dimensional structures of two of the three zinc-finger domains from the yeast transcriptional activator SWI5. One of these structures is quite similar

to those determined previously, but the other, that of finger 1, has the novel feature that its β -sheet comprises three strands rather than the usual two. There is evidence that this may represent a more general property of first fingers in a run of adjacent zinc-finger domains.

We are grateful to Drs Kiyoshi Nagai and Daniela Rhodes for their enthusiastic help and advice throughout this work, and to Professor Laurie Hall for providing space and facilities for the n.m.r. spectrometer within the Department of Medicinal Chemistry, University of Cambridge. We would also like to thank Dr Robert Diamond for his help and advice concerning the programs for multiple superposition and angle comparisons, Dr Kim Nasmyth for helpful discussions and for allowing us to discuss results of his prior to publication, and Drs Nikola Pavletich and Carl Pabo for permission to use the coordinates of their crystal structure of the Zif268-DNA complex. We thank a referee for pointing out that symmetry-related atom label transpositions for Arg residues should be based on the range $-90^\circ < \chi_5 < +90^\circ$, rather than $-180^\circ < \chi_5 < 0^\circ$ as we had originally proposed. We are grateful to the Human Frontiers in Science Program, who in part supported this work.

References

- Berg, J. M. (1988). Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 99–102.
- Brünger, A. T. (1988). *X-PLOR Manual*. Yale University Press, New Haven, CT.
- Clore, G. M., Gronenborn, A. M., Nilges, M. & Ryan, C. A. (1987). Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *Biochemistry*, **26**, 8012–8023.
- Daikun, G. P., Fairall, L. & Klug, A. (1986). EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. *Nature (London)*, **324**, 698–699.
- Detlefsen, D. J., Thanabal, V., Pecoraro, V. L. & Wagner, G. (1991). Solution structure of Fe(II) cytochrome c551 as determined by two-dimensional NMR. *Biochemistry*, **30**, 9040–9046.
- Diamond, R. (1992). On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Science*, in the press.
- Fairall, L., Harrison, S. D., Travers, A. A. & Rhodes, D. (1992). Sequence-specific DNA-binding by a two zinc-finger peptide from the *Drosophila melanogaster* Tramtrack protein. *J. Mol. Biol.* **226**, 349–366.
- Güntert, P., Braun, W. & Wüthrich, K. (1991). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 517–530.
- Holmes, M. A. & Matthews, B. W. (1982). Structure of thermolysin refined at 1.6 Å resolution. *J. Mol. Biol.* **160**, 623–639.
- Hyberts, S. G., Goldberg, M. S., Havel, T. F. & Wagner, G. (1992). The solution structure of eglin c based on measurements of many NOEs and coupling constants

- and its comparison with X-ray structures. *Protein Science*, **1**, 736–751.
- Klevit, R. E., Herriott, J. R. and Horvath, S. J. (1990). Solution structure of a zinc finger domain of yeast ADR1. *Proteins*, **7**, 215–226.
- Kochyan, M., Havel, T. F., Nguyen, D. T., Dahl, C. E., Keutmann, H. T. & Weiss, M. A. (1991a). Alternating zinc fingers in human male associated protein ZFY: 2D NMR structure of an even finger and implications for “jumping-linker” DNA recognition. *Biochemistry*, **30**, 3371–3386.
- Kochyan, M., Keutmann, H. T. & Weiss, M. A. (1991b). Alternating zinc fingers in the human male associated protein ZFY: refinement of the NMR structure of an even finger by selective deuterium labelling and implications for DNA recognition. *Biochemistry*, **30**, 7063–7072.
- Kochyan, M., Havel, Keutmann, H. T. & Weiss, M. A. (1991c). Architectural rules of the zinc-finger motif: comparative two-dimensional NMR studies of native and “aromatic-swap” domains define a “weakly polar switch”. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 8455–8459.
- Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
- Lee, M. S., Cavanagh, J. & Wright, P. E. (1989a). Complete assignment of the ^1H NMR spectrum of a synthetic zinc finger from *Xfin*; sequential resonance assignments and secondary structure. *FEBS Letters*, **254**, 159–164.
- Lee, M. S., Gippert, G. P., Kizhake, V. S., Case, D. A. & Wright, P. E. (1989b). Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science*, **245**, 635–637.
- Mueller, L. (1987). P.E. COSY, a simple alternative to E. COSY. *J. Magn. Reson.* **72**, 191–196.
- Nakaseko, Y., Neuhaus, D., Rhodes, D. & Klug, A. (1992). Adjacent zinc finger motifs in multiple zinc finger peptides from SWI5 form structurally independent, flexibly linked domains. *J. Mol. Biol.* **228**, 619–636.
- Neuhaus, D., Nakaseko, Y., Nagai, K. & Klug, A. (1990). Sequence specific [^1H]NMR resonance assignments and secondary structure identification for 1- and 2-zinc finger constructs from SWI5; a hydrophobic core involving four invariant residues. *FEBS Letters*, **262**, 179–184.
- Nilges, M. (1990). NMR structure determination and refinement. In *X-PLOR Version 2.1 User Manual* (Brünger, A. T., ed.), chapt. 19, pp. 257–263, Yale University Press, New Haven, CT.
- Pavletich, N. P. & Pabo, C. O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2·1 Å. *Science*, **252**, 809–817.
- Thukral, S. K., Eisen, A. & Young, E. T. (1991). Two monomers of yeast transcription factor ADR1 bind a palindromic sequence symmetrically to activate *ADH2* expression. *Mol. Cell. Biol.* **11**, 1566–1577.
- Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N. & Wüthrich, K. (1987). Protein structures in solution by nuclear magnetic resonance and distance geometry; the polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611–639.
- Wüthrich, K., Billiter, M. & Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.

Edited by P. E. Wright