Volume 262, number 2, 179–184 FEBS 08230 March 1990

Sequence-specific [^{1}H]NMR resonance assignments and secondary structure identification for 1- and 2-zinc finger constructs from SWI5

A hydrophobic core involving four invariant residues

David Neuhaus, Yukinobu Nakaseko, Kiyoshi Nagai and Aaron Klug

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

Received 2 January 1990

Complete [^{1}H]NMR resonance assignments are presented for the second of the three zinc fingers from SWI5 as it appears in both 1- and 2-finger constructs. Signals from finger 2 are unaffected by the presence or absence of finger 1, showing that the protein has a modular construction. The structure of finger 2 comprises a helix running from N56 to Q67, and a region approximating to an anti-parallel β-sheet running from Y42 to F53. These features combine to produce a hydrophobic core in the structure involving the invariant residues Y42, F53, L59 and H62.

DNA binding; [^{1}H]Nuclear magnetic resonance, 2-dimensional; Secondary structure; Sequence-specific resonance assignment; Zinc finger

1. INTRODUCTION

Recently, the short sequence motif known as 'zinc finger' has emerged as a new and widespread class of DNA binding domain in transcription regulatory proteins [1]. The first zinc finger protein to be identified, TFIIIA from *Xenopus laevis* [2], is the archetype for one class of zinc finger, the essential features of which are 4 metal binding residues (2 Cys and 2 His) and 3 invariant hydrophobic residues (Hydr). These are arranged in the sequence Hydr-X-Cys-$(X)_{2-5}$-Cys-$(X)_{12}$-His-$(X)_{2-4}$-His, with the other two hydrophobic residues near positions 4 and 10 beyond the second Cys. Miller et al. [2] proposed that the motif folds into one domain around the zinc ion and a hydrophobic core comprising the invariant residues, and Diakun et al. [3] showed by EXAFS that the zinc ion was indeed ligated to the 4 ligands proposed. The first 3-dimensional model for the zinc finger was put forward by Berg based on these considerations and the structures of other metalloproteins [4]. He proposed that the key structural elements are an α-helix near the C-terminus and an anti-parallel β-sheet near the N-terminus, which together bring the 3 conserved hydrophobic residues together to form a cluster at the core of the structure. In the absence of crystals of any zinc finger-containing protein, NMR studies have been undertaken by several workers to determine the 3-dimensional structure in solution, and results for single fingers of the TFIIIA class from ADR1 [5] and Xfin [6,7] have appeared.

Correspondence address: D. Neuhaus, MRC Laboratory for Molecular Biology, Hills Road, Cambridge CB2 2QH, England

One of the key structural questions that must be answered before the protein-DNA binding interaction can be understood is the relationship between successive fingers in multiple-finger proteins, and to tackle this problem we have begun a study of various constructs from the transcription factor SWI5. This protein was one of the first containing zinc fingers to be studied in this laboratory; it is one of a cascade of 'switch' proteins involved in transcription regulation leading to differentiation of mating type in yeast cells, and it contains 3 zinc fingers of the TFIIIA class near its C-terminus [8]. The complete 3-finger domain was expressed as a fusion protein and purified from *E. coli*, and it was shown that the re-folded 3-finger construct binds specifically to a target sequence of DNA which is part of the DNA-binding site in the HO gene recognised by intact SWI5 [9]. This shows that the 3 fingers are sufficient to direct the whole binding of SWI5 to its binding site.

For the NMR studies, we required the simplest construct containing a linker sequence between two adjacent fingers. As in the previous system, large scale preparation from *E. coli* proved difficult, since the required protein appeared in the insoluble fraction. Various alternative systems were tried, and the T7 system was found suitable for producing 1-, 2- and 3-finger constructs all in the soluble fraction; these maintained the specific DNA-binding at least in the 2- and 3-finger cases. Experiments have shown that a 70 amino acid construct comprising the first two fingers (fig.1) binds specifically to about two-thirds of the DNA target mentioned earlier, albeit at a much higher protein concentration than in the 3-finger case [10]. It

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

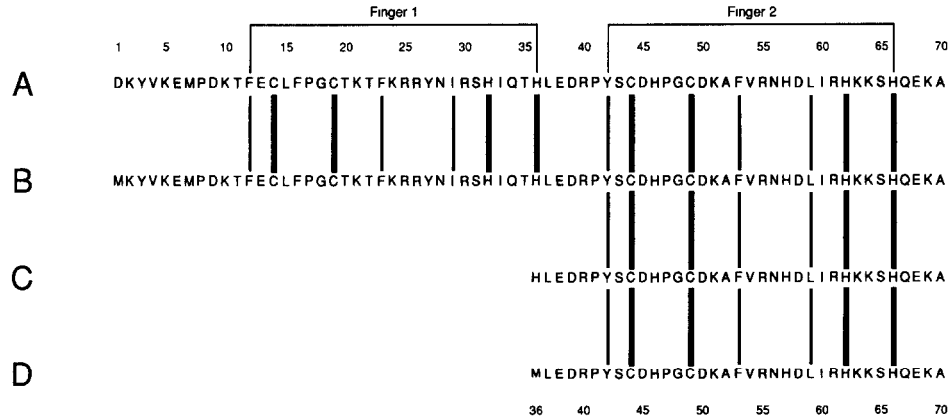


Fig.1. Constructs used in this work. (B) and (D) are, respectively, the 2-finger and 1-finger constructs used for the bulk of the assignment work. (A) and (C) were both obtained fortuitously from attempts to prepare a 2-finger construct beginning at M7. Presumably, they result from proteolysis by endogenous proteases. All sequences are numbered using the first residue of construct (A) or (B) as a starting point. The thick vertical bars indicate the conserved metal-binding residues, and the thin vertical bars indicate the conserved hydrophobic residues. The zinc finger motifs are also marked, the linker between them being taken as the region between the last histidine of finger 1 and the invariant tyrosine of finger 2 (see text).

is therefore presumably folded in its natural conformation. As shown below, NMR spectra of the second finger alone are closely analogous to the signals observed for the second finger within the 2-finger construct, showing that this single finger has the same fold as in the 2-finger construct which binds specifically to DNA.

In this paper we give a preliminary account of our NMR study, giving resonance assignments and secondary structure identifications for the second finger as it appears in both the 1- and 2-finger constructs. Although work on the 3-dimensional structure is still in progress, we are able to show that the fingers of SWI5 are modular in nature, and to present results additional to those of Lee et al. [6,7], in particular concerning interactions amongst the residues of the hydrophobic core.

2. MATERIALS AND METHODS

2.1. *Sample preparation*

SWI5 polypeptides were expressed in *E. coli* [10] under the control of the T7 gene 10 promoter [11]. BL21(DE3) was used for the host cell. The protein synthesis was induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside when the culture reached an optical density of 0.6 at 600 nm. SWI5 peptides were purified by carboxymethyl Sepharose followed by Sephadex G50 column chromatography. For the NMR experiments, approx. 5 mM (1-finger) or 2–3 mM (2-finger) solutions were prepared at pH 6.5 in either ^2H_2O or 85% ^1H_2O/15% ^2H_2O, containing 40 mM pyrophosphate and 50–500 μM $ZnSO_4$.

2.2. *NMR measurements*

All NMR measurements were carried out using a Bruker AM500 spectrometer equipped with digital phase shifting hardware and an ASPECT 3000 computer. Two-dimensional NMR spectra were acquired using time-proportional phase incrementation (TPPI) in t_1 to achieve F_1 quadrature detection [12,13]. Experiments yielding in-phase signals (NOESY and TOCSY) were optimised before execution to achieve a flat baseline by adjusting the absolute receiver reference phase [14] and the delay between receiver gate opening and the first acquisition point [15]. For all experiments except 2Q, the spectral width in both dimensions was 8 kHz, the filter width was 12 kHz, and the data size was 2K × 512 increments (time domain), 4K × 1K (frequency domain). For 2Q spectra the spectral width in F_1 was 16 kHz. Water suppression was achieved using presaturation (1.5 s) during the relaxation delay (and mixing time in the case of NOESY), using an irradiation field of the same frequency as and phase coherent with the hard pulses [16]. During data processing, time-domain data were multiplied in both dimensions by a sine-bell window shifted through $\pi/2$ radians, and the phase of the 2D spectrum was adjusted to double absorption. For in-phase spectra (NOESY and TOCSY), F_2 baseline offset errors were corrected by subtracting from every F_1 cross-section a one-dimensional trace constructed by adding 16 F_1 cross-sections free from cross peaks and diagonal peaks [17]. In a few cases a similar correction was required for F_1 baseline offset errors. No other baseline correction was used.

NOESY spectra were acquired using mixing times of 30, 60, 90, 120, and 200 ms. TOCSY experiments employed the DIPSI-2 mixing sequence to achieve minimum phase distortion [18], with two z-filters flanking the spin-locking period [19]. Mixing times between 50 and 100 ms were used with a spin-locking field strength of 5 kHz. In order to limit the total length of TOCSY experiments, variation of the z-filter delays was carried out in parallel, and was also combined with the CYCLOPS phase cycle on all the hard pulses and the spin-lock. For NOESY experiments, CYCLOPS was only carried out on the third pulse. Spectra of the single finger were run using 32 or 64 transients/increment, and those of the 2-finger construct were run using up to 128 transients/increment.

3. RESULTS

Complete sequence-specific assignments for the single-finger construct D at pH 6.5 and 283 K are given in table 1. The dispersion both of the NH signals (9.55–6.80 ppm) and of the $C^{\alpha}H$ signals (4.98–3.46 ppm) is clearly consistent with a folded structure, and facilitated the assignment process. Following the normal procedure, coupling connectivities within the spin-systems were first established using 2QF-COSY, TOCSY, RELAY, and 2Q spectra [21]. Data were acquired at two temperatures, 283 K

Table 1

Chemical shift assignments for SWI5 single finger construct D at pH 6.5 and 283 K

	NH	$C^{\alpha}H$	$C^{\beta}H$	$C^{\gamma}H$	$C^{\delta}H$	Others
Met 36	–	4.12	2.08–2.24	2.52–2.69		$C^{\varepsilon}H_3$ 2.12
Leu 37	8.82	4.37	1.52–1.75 (deg. × 3)		0.90, 0.935	
Glu 38	8.695	4.24	1.92, 2.01	2.20–2.30		
Asp 39	8.52	4.48	2.56–2.67			
Arg 40	8.15	4.565	1.50–1.63	1.34, 1.51	2.92–3.04	$N^{\varepsilon}H$ 7.185
Pro 41	–	4.28	1.22, 2.07	1.68, 1.85	3.625, 3.805	
Tyr 42	7.88	4.98	2.87, 3.10		7.05	$C^{\varepsilon}H$ 6.94
Ser 43	8.59	4.91*	3.83–3.96			
Cys 44	8.73	4.265	2.83, 3.32			
Asp 45	8.44	4.81	2.72, 3.07			
His 46	9.48	4.52	2.58–2.76		6.745	$C^{\varepsilon}H$ 8.135
Pro 47	–	4.18	1.75, 2.17	1.82, 2.34	3.65–3.72	
Gly 48	8.98	3.71, 4.19				
Cys 49	8.225	4.58	2.80, 3.40			
Asp 50	8.51	4.90*	2.63, 2.87			
Lys 51	8.34	4.04	1.21, 1.41	1.29, 1.29	1.44, 1.54	$C^{\varepsilon}H_2$ 3.02
Ala 52	7.65	4.93*	1.20			
Phe 53	8.58	4.81	2.71, 3.675		7.37	$C^{\varepsilon}H$ 6.975, $C^{\delta}H$ 6.53
Val 54	8.995	4.24	2.275	1.15, 1.20		
Arg 55	8.015	4.82	1.70–1.81	1.74, 2.00	3.25–3.38	$N^{\varepsilon}H$ 7.69
Asn 56	9.05	3.68	1.79, 2.46			$N^{\delta}H$ 6.655, 7.49
His 57	9.55	4.06	2.99, 3.24		7.135	$C^{\varepsilon}H$ 8.34
Asp 58	6.80	4.31	2.64, 2.87			
Leu 59	7.00	3.46	1.43, 2.23	1.57	1.13, 1.22	
Ile 60	8.24	3.47	1.72	1.14, 1.52, 0.85	0.72	
Arg 61	7.36	3.92	1.70–1.84	1.54, 1.65	3.11–3.26	$N^{\varepsilon}H$ 7.47
His 62	7.75	4.26	2.97, 3.17		7.155	$C^{\varepsilon}H$ 7.895
Lys 63	8.85	3.65	1.90–2.04	1.80–1.90	1.87, 1.89	$C^{\varepsilon}H$ 3.12, 3.26
Lys 64	7.24	4.08	1.79–1.95	1.41–1.53	1.60–1.78	$C^{\varepsilon}H_2$ 2.93
Ser 65	7.81	4.18	3.74, 3.83			
His 66	7.125	4.62	2.62, 2.96		6.40	$C^{\varepsilon}H$ 7.955
Gln 67	7.745	4.36	2.07, 2.20	2.39–2.48		$N^{\varepsilon}H$ 7.00, 7.69
Glu 68	8.52	4.25	1.98, 2.07	2.28–2.40		
Lys 69	8.53	4.35	1.79, 1.90	1.43–1.55	1.63–1.78	$C^{\varepsilon}H_2$ 3.02
Ala 70	8.19	4.15	1.37			

* Observed at 300 K

and 300 K, to allow resolution of accidental degenerations of NH signals with one another and of $C^{\alpha}H$ signals with the water signal. In this way, complete spin-system assignments were obtained, with the exception of the backbone NH proton of H57 and all non-amide exchangeable protons other than $NH\epsilon$ of the 3 arginine residues. H57 NH was later found during analysis of the NOESY spectrum, and has low intensity (~20% under our conditions) due to more rapid solvent exchange than for other amide protons. When 1-1 excitation was used in place of presaturation to achieve water suppression, H57 NH was restored to full intensity in the one-dimensional spectrum.

Sequential assignment was achieved in the normal way by analysis of the NOESY spectrum [21]. Fig.2 shows a summary of the sequential and other enhancements used to make the assignments. A somewhat greater number of cross peaks were observed in the NOESY spectra recorded at 283 K than in those recorded at 300 K, albeit at the cost of some line-broadening at the lower temperature. Otherwise, the spectra were not strongly perturbed by changing the temperature, demonstrating that the structure is probably stable over this range. A sequence of NOESY spectra with increasing mixing times (30, 60, 90, and 120 ms) was also recorded at 283 K to help quantitate

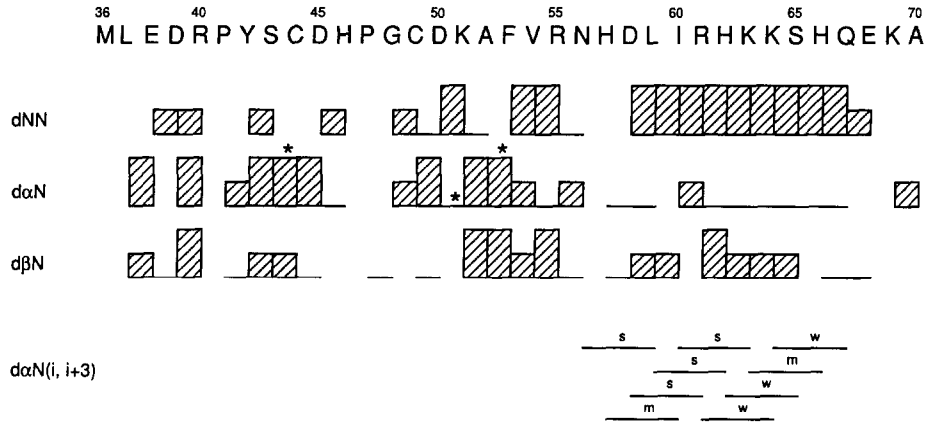


Fig.2. NOE enhancements used to make the sequence-specific resonance assignments for construct D. The height of the hatched blocks shown for the d_{NN}, $d_{\alpha N}$, and $d_{\beta N}$ connectivities represent very approximately the strength of the corresponding enhancement in the NOESY spectrum (τ = 200 ms, T = 283 K), as do the letters w (weak), m (medium), and s (strong) for the $d_{\alpha N}(i, i+3)$ connectivities. The $d_{\alpha N}$ enhancements marked * were observed in a NOESY spectrum recorded at 300 K (τ = 200 ms).

the distance constraints needed for calculation of the 3-dimensional structure, at present in progress. These spectra showed that most of the enhancements used in the assignments were appreciable at short mixing times, and thus unlikely to be due to spin-diffusion.

A similar overall procedure was followed to obtain sequential assignments for the 2-finger construct B, except that the lower protein concentrations available in this case made it necessary to acquire spectra from samples in ^2H_2O as well as those in 85% ^1H_2O/15% ^2H_2O. NOESY, TOCSY, and 2QF-COSY spectra of samples in ^2H_2O at 300 K were used in addition to those previously mentioned for the single finger during the spin-system assignment procedure. Assignments are still lacking for some 3 or 4 residues in finger 1 of B, but complete assignments were obtained for the linker region and all of finger 2.

Most of the samples studied during this work underwent spontaneous degradation over a period of days, resulting in a quite different spectrum, assumed to be that of the de-metallated species. This showed much reduced signal dispersion, significantly sharper lines, and no NOE enhancements transmitted over more than two residues in the sequence, all of which is consistent with a largely unfolded structure. This is in contrast to the findings of Carr et al. [20] for the case of a different zinc finger sequence, mkr2, at lower pH, but the oxidation state of the cysteines in the SWI5 case is not yet established and may differ from that in the metal-free mkr2. Sequential assignments have been obtained for about three-quarters of the 1-finger construct D in the unfolded state, and work is progressing to complete these.

4. DISCUSSION

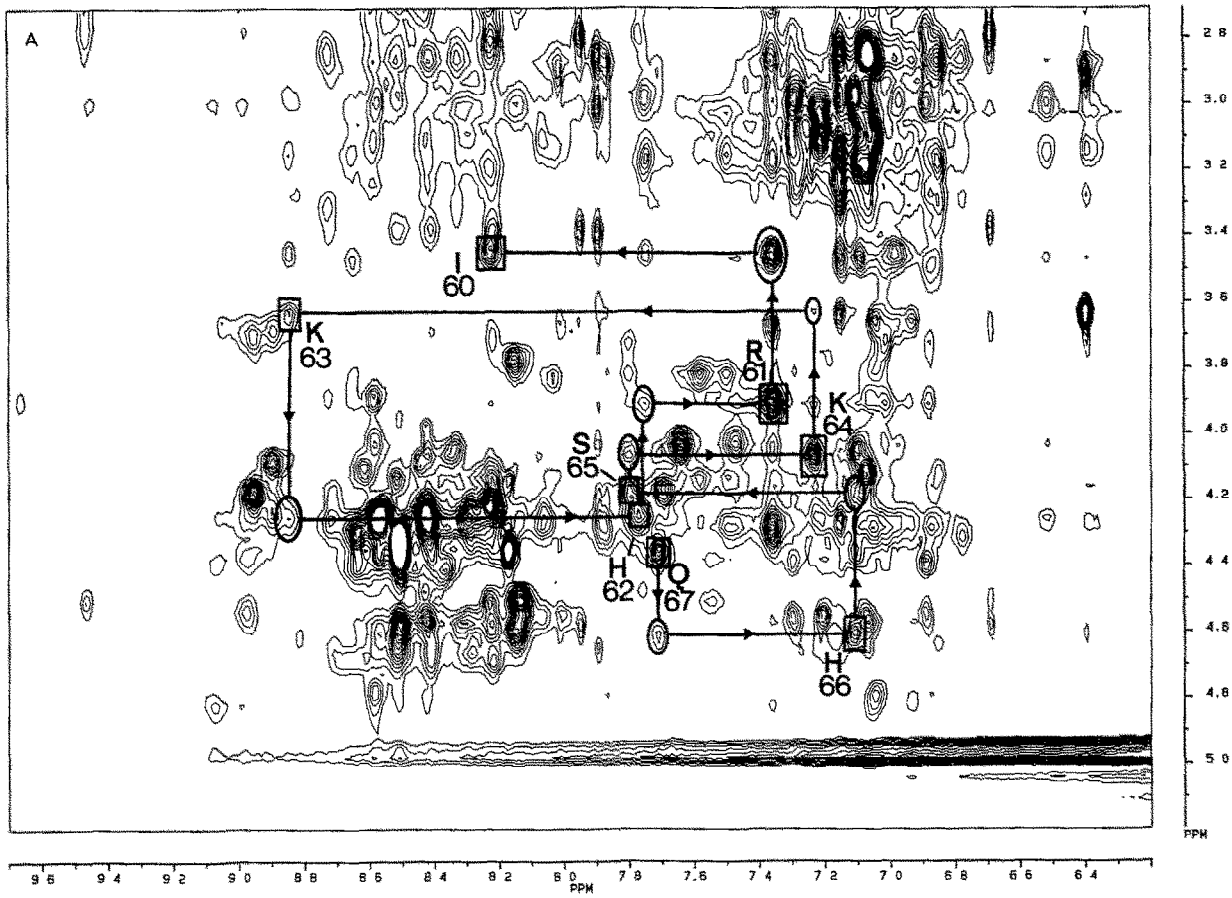
The full assignments for the 2-finger construct will be described in a subsequent paper, but the key point to make here is that all the signals and NOE enhancements involving residues R40–A70 are essentially identical in spectra of the 2-finger construct B and of the single-finger construct D (fig.1). This clearly indicates that the 2-finger structure is modular in nature, and that the 2 fingers within it interact only weakly with one another in the free protein. Consistent with this, we have so far found no NOE enhancements between the 2-finger domains, and only intra-residue and sequential enhancements in the short linker region. To establish that no circular arguments were involved in demonstrating the coincidence of signals from finger 2 in constructs B and D mentioned above, we point out that almost all the assignments for the 2-finger construct were already made before work on the single finger was begun. The only use made of data from the single finger when making assignments in the 2-finger spectra was to make the relative sequential assignments of D45 and H46, to locate some protons beyond $C^{\beta}H$ within long side-chains (Lys, Arg, Pro), and to locate the backbone NH signals of N56, H57, and E68. Fig.3 illustrates the relationship between spectra of a 2-finger construct (B) and a 1-finger construct (D).

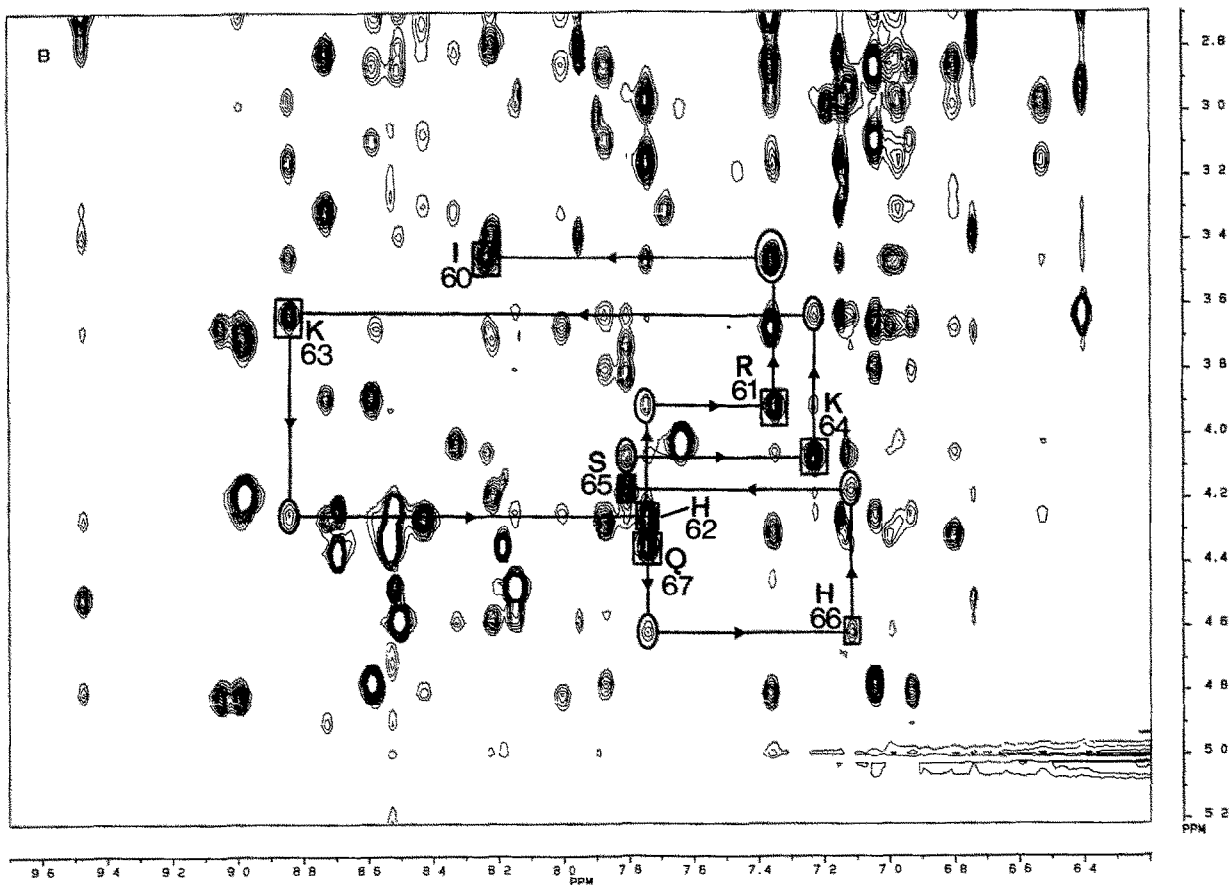
In addition to the comparison of finger 2 as it occurs in constructs B and D, we have also measured spectra of constructs A and C (fig.1). Although assignments for these species have not been obtained independently, it is clear from comparison of the spectra that, again, the signals and enhancements from residues R40–A70 are essentially identical to those seen for constructs B and D.

The work on the unfolded species mentioned earlier is not yet complete, but it is already possible to draw some tentative conclusions as to the 'boundaries' of the region which is structured by the presence of zinc. Residues E68–A70 show particularly sharp signals, and have exactly the same pattern of shifts and enhancements in the folded as in the unfolded single-finger material, suggesting strongly that they are unfolded in both. In contrast, the signals and enhancements for Q67 are quite different in the two cases, placing the C-terminal 'boundary' between Q67 and E68. The identity of signals and enhancements from residues R40–A70 in the folded 1- and 2-finger constructs mentioned earlier further suggests that the first residue of the structured region (for finger 2) is R40. The further characterisation of the linker is one of the main remaining goals of this work, and will be tackled by comparing the patterns of shifts and enhancements for these residues and their neighbours as observed in folded constructs of finger 1, finger 2, fingers 1 + 2, and unfolded material.

Certain features of the secondary structure of finger 2 are apparent from the spectra of both the 1- and 2-finger constructs. There are many enhancements indicating that the side chains of all 3 'hydrophobic core' residues Y42, F53, and L59 are in contact, and that the side chain of H62 also forms part of this core region, in agreement with the general model for zinc fingers of the TFIIIA class proposed by Berg [4]. Fig.4 shows the enhancements found involving protons in these residues. This provides a clear structural basis for including the first hydrophobic residue within the zinc finger domain, a conclusion not able to be drawn from

Fig.3. Part of the NOESY spectrum (τ = 200 ms, T = 283 K) of (A) the 2-finger construct B and (B) the 1-finger construct D. Conditions were identical for the two spectra, except that the protein concentration in (B) was approx. 2–3 times that in (A), and the measurement time in (A) was twice that in (B). Also, there was a minor difference in pH. A short stretch of sequential $d_{\alpha N}$ connectivities is shown in both (A) and (B) to emphasise the commonality between the signals due to finger 2 in each spectrum. Boxed cross peaks correspond to intraresidue NH, $C^{\alpha}H$ interactions, and are labelled according to residue type and sequence position; cross peaks within ellipses correspond to sequential $d_{\alpha N}$ interactions.


A
I 60
K 63
R 61
K 64
S 65
H 62
Q 67
H 66
PPM


B
I 60
K 63
R 61
K 64
S 65
H 62
Q 67
H 66
PPM

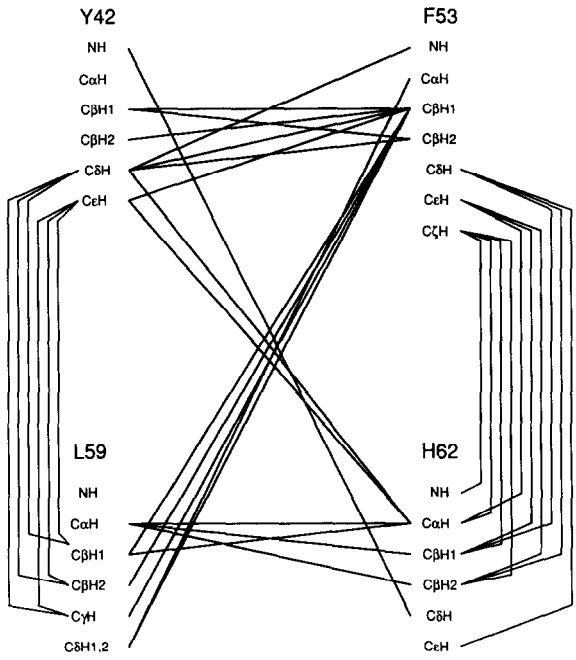


Fig.4. NOE enhancements observed between protons of residues in the hydrophobic core of finger 2 in construct D (Y42, F53, L59, H62).

the results of Lee et al. [6,7]. A pattern of strong d_{NN} connectivities and $d_{\alpha N}(i,i+3)$ connectivities indicates a stretch of helix running from N56 to Q67. The location of this helix is similar to that found in Xfin-31 by Lee et al. [6,7], but it is significantly longer than that found by Párraga et al. in ADR1a [5], or that proposed by Berg [4]. Several other features are broadly similar to the results of Lee et al. [6,7] but show differences of detail. The pattern of strong $d_{\alpha N}$ connectivities (Y42-S43-C44-D45, C49-D50, and K51-A52-F53) and large (NH, $C^{\alpha}H$) coupling constants (Y42, D50, F53, R55) suggest a region of probably irregular anti-parallel β-sheets extending roughly from Y42 to F53, centered on a turn at P47-G48. Consistent with this, we find enhancements between K51 NH and both $C^{\beta}H$ protons of C44, in addition to the many enhancements between protons of Y42 and F53 referred to earlier. The detailed pattern of enhancements is somewhat different from that found by Lee et al. [6,7], as would be expected only on the grounds that SWI5 has 4 residues between the metal-binding cysteines, whereas Xfin-31 has 2.

While the characterisation of finger 1 in the 2-finger construct is so far less complete than that of finger 2, it is possible to say that it shares the same key structural features, namely the helix, the pattern of strong $d_{\alpha N}$ connectivities within the anti-parallel β-sheet, and contacts within the hydrophobic core (enhancements between residues F12, F23, I29, and H32). Since the side chains of F12 in finger 1 and Y42 in finger 2 are both involved in their respective hydrophobic clusters, we can generalise the conclusion that the first conserved hydrophobic residue should be included within the zinc finger domain.

Acknowledgements: This work was made possible by the generosity of Professor Laurie Hall in providing space and facilities for the spectrometer within the Department of Medicinal Chemistry, University of Cambridge. During the early part of this work, some spectra were run on spectrometers at Imperial College London (Bruker AM500), the NMR centre at NIMR Mill Hill London (Bruker AM500), and Parke Davis Research Unit Cambridge (Bruker AM300), and we gratefully acknowledge the help given to us by those institutions, and particularly Mr Dick Sheppard, Drs Chris Bauer, Tom Frenkiel, Jim Feeney, Giles Ratcliffe, and Professor Alan Fersht. We also thank Drs Bauer and Frenkiel for giving us details of their implementation and extension of the baseline flattening procedure for in-phase 2D spectra mentioned in the text. We thank Dr Daniela Rhodes for help and advice.

REFERENCES

[1] Klug, A. and Rhodes, D. (1987) Trends Biochem. Sci. 12, 464–469.
[2] Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609–1614.
[3] Daikun, G.P., Fairall, L. and Klug, A. (1986) Nature 324, 698–699.
[4] Berg, J.M. (1988) Proc. Natl. Acad. Sci. USA 85, 99–102.
[5] Párraga, G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T. and Klevit, R.E. (1988) Science 241, 1489–1492.
[6] Lee, M.S., Cavanagh, J. and Wright, P.E. (1989) FEBS Lett. 254, 159–164.
[7] Lee, M.S., Gippert, G.P., Kizhake, V.S., Case, D.A. and Wright, P.E. (1989) Science 245, 635–637.
[8] Stillman, D.J., Bankier, A., Seddon, A., Gorehout, G. and Nasmyth, K.A. (1988) EMBO J. 2, 485–494.
[9] Nagai, K., Nakaseko, Y., Nasmyth, K. and Rhodes, D. (1988) Nature 332, 284–286.
[10] Nakaseko, Y., Rhodes, D. and Nagai, K., in preparation.
[11] Rosenberg, A.H., Lade, B.N., Chui, D.-S., Lin, S.-W., Dunn, J.J. and Studier, F.W. (1987) Gene 56, 125–135.
[12] Bodenhausen, G., Vold, R.I. and Vold, R.R. (1980) J. Magn. Reson. 37, 93–106.
[13] Marion, D. and Wüthrich, K. (1983) Biochem. Biophys. Res. Commun. 113, 967–974.
[14] Marion, D. and Bax, A.D. (1988) J. Magn. Reson. 79, 352–356.
[15] Hoult, D.I., Chen, C.-N., Eden, H. and Eden, M. (1983) J. Magn. Reson. 51, 110.
[16] Zuiderweg, E.R.P., Hallenga, K. and Olejniczak, E.T. (1986) J. Magn. Reson. 70, 336–343.
[17] Klevit, R.E. (1985) J. Magn. Reson. 62, 551–555.
[18] Shaka, A.J., Lee, C.J. and Pines, A. (1988) J. Magn. Reson. 77, 274–293.
[19] Rance, M. (1987) J. Magn. Reson. 74, 557–564.
[20] Carr, M.D., Pastore, A., Gausepohl, H., Frank, R. and Rösch, P., Eur. J. Biochem., in press.
[21] Wüthrich, K., NMR of Proteins and Nucleic Acids, Wiley, New York, 1986.