Nucleic Acids Research, 1993, Vol. 21, No. 15 **3341–3346**

A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA

Yen Choo and Aaron Klug
Medical Research Council, Laboratory of Molecular Biology, Cambridge CB2 2QH, UK

Received June 7, 1993; Accepted June 22, 1993

ABSTRACT

Zinc fingers of the TFIIIA type are connected by short linker sequences between the structural units. Structural investigations by 2D NMR in solution and by X-ray crystallographic analyses of complexes with DNA point to a passive role for the linkers. We have therefore investigated the influence of the linker sequence on DNA binding using as a model the first three fingers of the protein TFIIIA. Insertion of certain heterologous linkers abolishes binding, and replacement of individual amino acids can reduce binding by factors of up to twenty-four.

INTRODUCTION

The zinc finger is an ubiquitous eukaryotic DNA binding motif first identified in *Xenopus* transcription factor IIIA (TFIIIA) and now known to occur in hundreds of proteins (1). The consensus sequence for TFIIIA-type zinc finger domains is Tyr/Phe-X-Cys-X_{2-4}-Cys-X_3-Phe-X_5-Leu-X_2-His-X_{3-5}-His (where X is any amino acid), and these are present in from 2 upto 37 copies per protein, usually arranged in tandem and connected by linkers (2). Each zinc finger is an autonomously folding mini-domain which is dependent on zinc for stability: the tertiary structure is an anti parallel β-sheet packed against an (predominantly) α-helix, with the invariant cysteines and histidines chelating a zinc ion and the three conserved hydrophobic residues forming a core. This structure, first predicted assuming sequence homology between recurring substructures in metalloproteins (3), was later solved by 2D NMR methods (4), and has been recently confirmed by X-ray crystallography (5).

TFIIIA is a transcription factor of 40kD, containing nine zinc fingers, which when bound to the internal control region of the 5S RNA gene (ICR) upregulates the production of 5S RNA in *Xenopus* oocytes (6). Pending its incorporation into ribosomes, the 5SRNA is found in the cytoplasm of the oocytes as a ribonucleoprotein storage complex, bound to TFIIIA, or to another nine-fingered protein called p43 (7). Unlike TFIIIA, which has equal affinity for both 5SDNA and 5SRNA, p43 can apparently bind only the 5SRNA (8).

The means by which zinc finger proteins recognise nucleic acids in a sequence-specific manner have been the subject of much inquiry. Recently, the crystal structure of the mouse protein Zif268 bound to an 11-mer oligonucleotide has revealed that the three fingers occupy the major groove, wrapping around the DNA for almost one turn of the double helix (5). Each finger is an independent binding domain which recognises a three basepair subsite by two base-specific contacts via residues on the exposed face of the α-helix. The pattern of contacts to the DNA is simple and repetitive, and is the consequence of the relationship between one finger and the next being a function of the helical period.

Nakaseko *et al* (9) have shown by NMR that the first two of three fingers of the yeast transcription factor SWI5 are flexibly linked in solution, and have no preferred orientation relative to each other. Hence on binding DNA these fingers of SWI5, and presumably others, acquire a preferred relationship where previously there was none. Of particular interest in this matter are the short peptide sequences, commonly called 'linkers', which connect one zinc finger to another. However, in the Zif268-DNA crystal structure these linkers appear to play no further structural role since no interactions with the DNA or any part of the fingers can be seen.

The two linkers of Zif268 are of the type found in many zinc finger proteins, typified by the *Drosophila* protein *Kruppel*, which has linkers of the sequence Thr-Gly-Glu-Lys-Pro. On current databases this emerges as the most highly conserved series of residues in the consensus primary sequence of zinc fingers. Considering the evolutionary conservation, it has been expected that this sequence will have an obvious functional role. Here we investigate the question of whether these linkers play a role in the interaction with DNA. Fingers 1–3 of TFIIIA, but no others, are connected by sequences with substantial homology to the *Kruppel*-type linkers, and moreover these first three fingers are necessary and sufficient for binding to the 5SRNA ICR with full specificity and near-full affinity (10,11). We have therefore used these in an analysis of the effects of mutations on DNA binding, and find that in a panel of TFIIIA 3-finger peptides with variant linkers, deviations from the consensus linker sequence cannot be reconciled with high affinity binding to the ICR.

MATERIALS AND METHODS

Assembly of synthetic genes

The genes for the 3-finger constructs TF3A, TF43A, TF43B and TF321 were each assembled from 8 oligonucleotides made on an Applied Biosystems 380B DNA synthesiser. TF3A comprises residues 1 to 99 of TFIIIA, and the others are varied as shown in Fig. 4. The oligonucleotide design was such that the annealing of the noncoding strand was directed by oligonucleotides

specifying the coding strand and *vice versa*. Gel purified oligonucleotides were phosphorylated, annealed and ligated in a mixture containing 100pmol. Ligated strands were gel isolated prior to cloning.

Gene cloning and site-directed mutagenesis

The synthetic genes were cloned between the *Nde*I and *Eco*RI sites of the phagemid vector pGM484 (12) and propagated in *E.coli* strain TG1. For production of DNA templates suitable for oligonucleotide-directed mutagenesis, *E.coli* strain CJ236 was transformed with the appropriate constructs and phage were rescued by infection with helper phage M13K07 as described (13). Mutagenic oligonucleotides were annealed, extended using T7 DNA polymerase, and ligated essentially as described. Mutants were passaged to *E.coli* TG1 and identified by sequencing plasmid DNA (Sequenase).

Expression of zinc finger constructs

All constructs were expressed *in vitro* by coupled transcription and translation in a TNT Coupled Reticulocyte Lysate System according to the manufacturer's instructions (Promega), but made 500μM in $ZnSO_4$. Transcription was driven by the promoter for T7 RNA polymerase (14), which is found upstream of the *Nde*I cloning site in pGM484. The translation products were labelled by the addition of ^{35}S-Cys to the reticulocyte lysate, and were subsequently visualised by autoradiography following resolution by SDS-PAGE. Protein concentrations were determined by densitometer scanning (Molecular Dynamics Computing Densitometer Model 300A) of the autoradiogram in the linear range.

Gel retardation assays

All constructs were assayed using a ^{32}P end-labelled synthetic oligonucleotide duplex which encompases base-pairs 70–99 in the sequence of the 5SRNA gene (15). Reactions were in 10μl volumes of binding buffer (20mM HEPES pH7.5, 50mM NaCl, 10mM $MgCl_2$, 5mM DTT, 50μM ZnOAc, 5% glycerol) containing 1μl translation extract and 1μg poly dIdC as competitor. These were incubated at 20°C for 20min and subsequently at 4°C for 20min, then loaded directly onto prerun 8% polyacrylamide gels (38:1 acrylamide/bisacrylamide) for electrophoresis in 45mM Tris Borate at 5°C. These electrophoretic conditions can be seen not to dissociate the wild type complex (Figs 1 and 2). Gels were dried onto DE81 paper (Whatman) and exposed to preflashed film at −70°C. The autoradiograms presented herein are typically overexposed for clarity (see Fig.2). Densitometry was carried out using a Molecular Dynamics Computing Densitometer Model 300A, and data were plotted using KaleidaGraph Version 2.0 (Abelbeck Software) as described below.

Determination of the dissociation rate constant

Reactions in a final volume of 50μl were prepared in binding buffer containing 5μl translation extract and 25nM of the above end-labelled ICR oligonucleotide. Following a 40min incubation period as described, reactions were put onto wet ice and a zero time point was removed and loaded onto a prerun gel. The reaction mixtures were then mixed manually when made 2μM in unlabelled ICR oligonucleotide. Time points were taken at appropriate intervals when 5μl aliquots were loaded onto a running gel.

RESULTS

The wild-type three-finger gene (denoted TF3A) and those of mutants with multiple amino acid substitutions (TF43A, TF43B and TF321) were prepared *de novo* from synthetic oligonucleotides and cloned. Single amino acid substitutions were subsequently introduced to the wild-type gene by oligonucleotide-directed mutagenesis on single stranded templates containing small amounts of dU. Small quantities of soluble protein were conveniently prepared *in vitro*, by coupled transcription and translation. Peptides were titrated with DNA and binding was analysed by gel retardation, the results of which were treated quantitatively.

In vitro transcription and translation

All constructs were initially assayed for protein expression in *E.coli* strain BL21(DE3), where the overexpressed proteins were found in inclusion bodies. Soluble protein was then prepared from plasmid DNA given to the TNT Reticulocyte Lysate System.

The manufacture of *in vitro* translation extracts generally involves the prior digestion of endogenous mRNA by treatment with micrococcal nuclease, which is then inactivated by the metal ion chelating agent EGTA. As EGTA will bind zinc ions, we titrated the lysate with $ZnSO_4$. Maximal zinc finger activity, although associated with poorer expression, was found to occur when $ZnSO_4$ was added to 500μM.

Protein produced by this method migrated as a sharp band when analysed by SDS-PAGE. Yields were calculated from the fractional incorporation of ^{35}S-Cys as measured by autoradiogram scanning, and were typically approximately 20fmol/μl lysate.

Lysate which has not been programmed with vector DNA does not produce ^{35}S labelled protein and does not cause significant quantities of the ICR fragment to be retarded on gel electrophoresis as described above (see Fig. 1).

Evaluation of the apparent equilibrium dissociation constant

The mass action equation describes the concentrations of complex (DNA.P), free protein (P) and free DNA (DNA) at equilibrium:

$$K_d = \frac{[DNA]\,[P]}{[DNA.P]}$$

The dissociation constant (K_d) of DNA binding proteins is most frequently estimated as the concentration of free protein at half-maximal binding (16). In cases where the K_d far exceeds the concentration of available protein this method of determination is not possible. In the converse situation, the protein concentration is kept constant, while increasing the input of oligonucleotide until all protein is in complex. The apparent K_d is then the concentration of free DNA when half the protein has been bound. For both cases, the K_d is only an approximation since the concentrations of both free components in the gel cannot be known; for this reason, one component of the reaction must be in large excess over the other. The K_d is also said to be apparent as the equilibrium may be perturbed by mass action during electrophoresis.

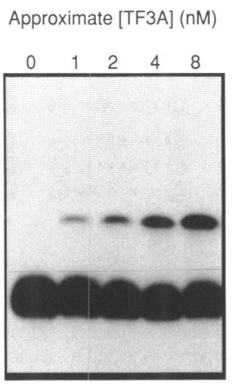

Figure 1. Titration of the ICR oligonucleotide with the wild-type 3-finger protein (TF3A). Autoradiogram of a gel retardation assay where equal volumes of rabbit reticulocyte lysate containing varying amounts of TF3A, in the indicated approximate concentrations, were added to equal amounts of the ICR oligonucleotide (25nM).

We have used the latter approach in calculating the apparent K_d, since it requires very little protein, the concentration of which need not be known. The mass action equation can be rearranged in terms of the total protein concentration (P_o) to a form similar to that of the Michaelis-Menten equation:

$$[DNA.P] = \frac{[P_o]\,[DNA]}{[DNA] + K_d}$$

Since the concentration of input DNA far exceeds that of the protein, we have assumed, in fitting the data to the above equation, that the concentration of free DNA approximates the total. Values for complex formation (autoradiography density units) at known DNA concentrations were fitted to the equation using the programme KaleidaGraph Version 2.0 (Abelbeck Software), from which K_d values were derived.

DNA binding by wild-type 3-finger polypeptide

In our hands, the 3-finger peptide TF3A will bind to the 5SRNA ICR oligonucleotide with an apparent K_d of 15nM in the presence of 1μg competitor DNA and the impurities associated with the expression system. This is comparable to the value obtained in a previous study of the three fingers which has employed purified peptide and less competitor DNA in experiments which yield a K_d value of 6nM (11).

Figure 1 shows a sample band shift obtained when increasing amounts of TF3A (*i.e.* the natural sequence) were added to a fixed concentration of DNA. It will be seen that addition of rabbit reticulocyte lysate alone does not cause a band shift. Figure 2A shows a sample band shift for TF3A in which decreasing amounts of DNA were added to a fixed amount of peptide, while Figure 3A shows a saturation curve used in the determination of the apparent dissociation constant from an experiment of this kind.

DNA binding by 3-finger polypeptides with alien linkers

In order to test whether the linker regions of TFIIIA fingers 1–3 are truly passive, we have replaced both linkers in their entirety by sequences derived from different linkers.

We define a linker as follows. As originally expected, and as borne out by structural determinations, the conserved Tyr or Phe

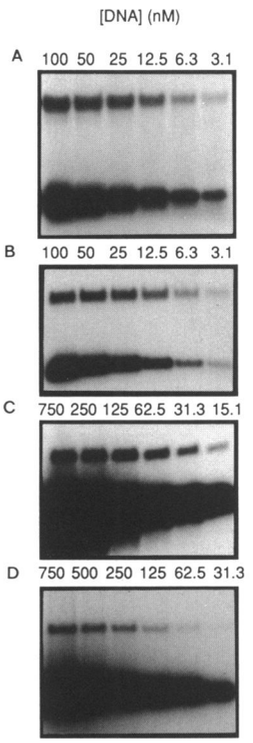

Figure 2. Titration of TF3A and mutant polypeptides with the ICR oligonucleotide by gel retardation assay. Reactions contained 1μl transcription/translation extract and the indicated ICR oligonucleotide concentrations prepared as described in Materials and Methods. The transcription/translation extracts contained: (**A**) TF3A; (**B**) N72P; (**C**) K41S; (**D**) G39P.

residue preceding the first cysteine by two positions is part of the hydrophobic cluster in the structural unit —'the finger'—, so we therefore define the linker between two tandemly repeated units as the sequence lying between the last histidine of one finger and the first conserved hydrophobic residue of the next, *i.e.* T G E K P in the case of the *Kruppel* family.

We chose to prepare two constructs in which both the linkers were changed to sequences from the quasihomologous protein p43. In the construct TF43A the entire H to C region was replaced, in accordance with other usages of the term 'linker'; whereas in TF43B only the residues of the *bona fide* linker were altered (Fig. 4).

An additional 3-finger construct designated TF321 was prepared in which the first linker was replaced by the (6 amino acid) linker found between fingers 3 and 4 in TFIIIA (Fig. 4).

In all the above cases where the linker is heterologous, no appreciable binding was detected at DNA concentrations upto 1μM, from which we infer at least hundredfold increases in the apparent dissociation constant.

DNA binding by 3-finger polypeptides with linkers mutated at single residues

To analyse the sequence requirements on the linker for strong DNA binding, single amino acid changes were introduced to the linker between fingers 1 and 2 in TF3A. In addition, since the two linkers in TF3A are not identical, one point mutant was prepared which carries P instead of N in the second linker of TF3A. A list of point mutants and their apparent dissociation

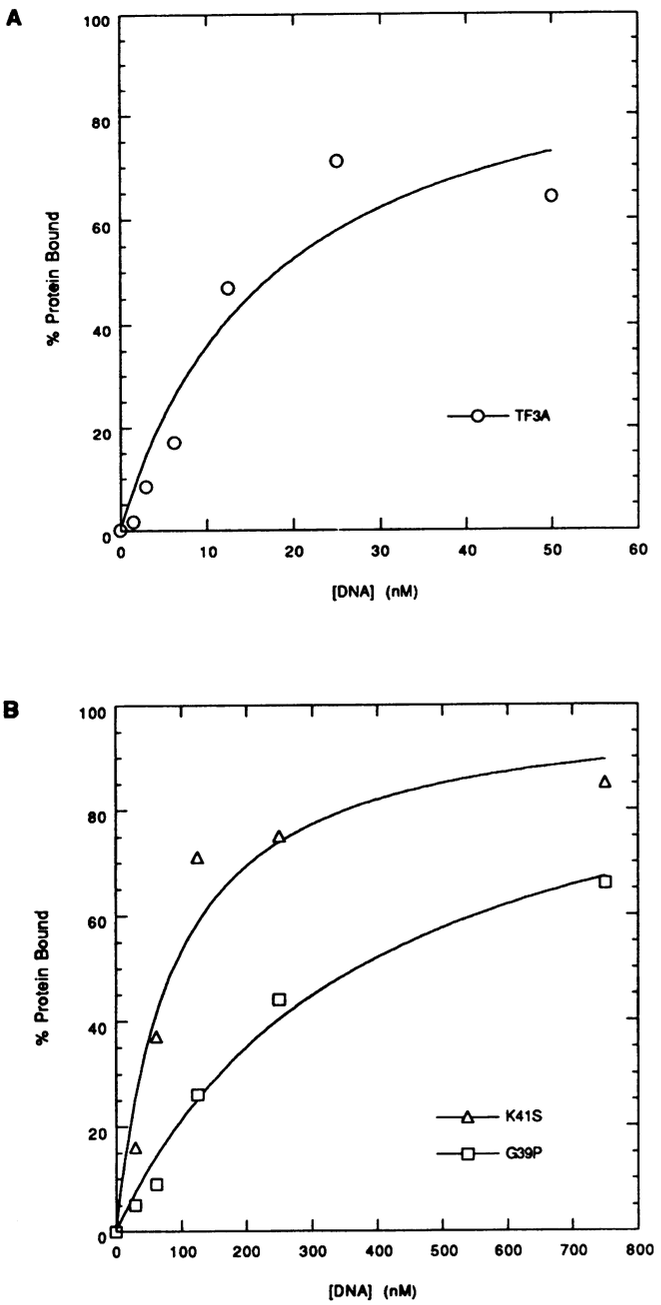

Figure 3. Full curve analysis of binding data. Binding data derived from gel retardation was fitted as described in the Results section. Graphs are presented in terms of percentage protein bound at each DNA concentration. (A) Representative analysis of TF3A binding to the ICR oligonucleotide; (B) Representative analyses of K41S and G39P binding to the ICR oligonucleotide.

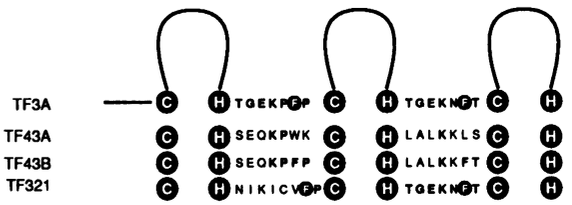

Figure 4. Amino acid sequences of highly variant linkers introduced in the 3-finger polypeptide. The 3 fingers of TF3A are shown schematically in the first line connected by the wild-type linker sequences. Other constructs, linkers of which are defined below, have fingers connected by sequences derived from either p43 (TF43A and TF43B) or from TFIIIA itself (TF321). The nomenclature is explained further in the Results section.

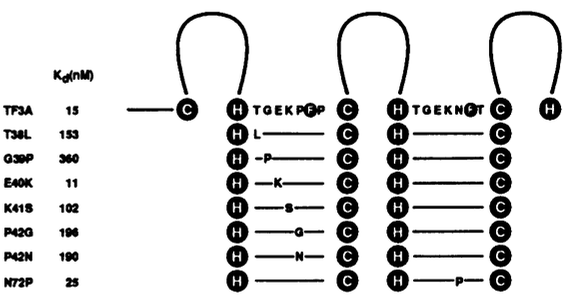

Figure 5. Amino acid sequences of linkers mutated at single residues. The three fingers of TF3A are shown schematically in the first line, connected by the wild-type linker sequences. Below, thin lines denote homology with the wild type and the mutant residues are written in the one letter code. Average K_d values are listed alongside the nomenclature.

constants is given in Figure 5. Sample bandshifts are shown in Figure 2 while binding curves for two mutants are contrasted in Figure 3. The errors in individual K_d determinations may be as high as 40%, but each quoted K_d is the mean of two to four determinations using protein made on separate occasions. Since we are demonstrating changes in TF3A affinity which differ by factors of two or more, we feel that the errors do not affect our conclusions.

In summary, the two point mutants E40K and N72P bound to the fragment of the ICR with an affinity similar to that of the wild type in our system. By contrast, the point mutants T38L, K41S and P42G had affinities which were reduced by a factor of 7–13, and the mutant G39P had an affinity reduced by a factor of 24 (Fig. 3). Single mutations in the linker therefore do not have such a dramatic effect on DNA binding as do multiple mutations.

We also carried out two experiments in which the linker sequence T G E K P was maintained but an additional serine was inserted, firstly between the Gly and Glu, and secondly between the Lys and Pro in the first linker of TF3A. Peptides with these six amino acid long linkers bound to the ICR oligonucleotide with a K_d increased by a factor of about 10.

The half life of the TF3A-ICR oligonucleotide complex

Although we have shown that the linker sequences have an effect on the binding of TF3A to the ICR, the role of the linkers has remained obscure. We wished therefore to address the question of whether the function of the linkers lies in the final form of the complex state, or alternatively in the rate of formation of this state. By comparing the rate of dissociation of the TF3A complex with those of mutant complexes, we hoped to be able to differentiate between the two possibilities. Since the dissociation rates will be independent of the binding event, any differences in these will most likely be due to effects in the complex.

An experiment to determine the dissociation rate constant of a complex is commonly performed by challenging a preformed complex of protein and radiolabelled probe, with a vast excess of unlabelled probe. The purpose of this is to trap dissociated

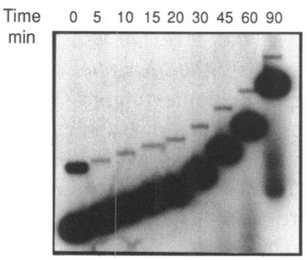

Figure 6. Determination of the off-rate of the TF3A-ICR complex. Preformed complexes were challenged with an excess of competitor ICR oligonucleotide at 4°C as described in Materials and Methods. Reactions were loaded onto a running gel at the time indicated above each lane.

protein in a new non-radioactive complex. This will occur since in addition to the rate of dissociation there is plainly a reverse association which governed the formation of the complex in the first place. The decay of radioactivity in the complex as a function of time is therefore witness to the dissociation of the original complex, and the rate can hence be derived.

We have found the half life of the TF3A-ICR oligonucleotide complex too fast to be able to measure without the use of specialised apparatus. Figure 6 shows how five minutes after incubating with an excess of unlabelled DNA, the rate of dissociation has already reached a plateau. Similar results are obtained when complex is sampled after one minute, but we do not feel it appropriate to draw conclusions from this as we are concerned that the time taken for the DNA to enter the gel may limit the resolution.

Liao *et al.* (11), by contrast, have recently claimed a value of 100min for the half life. We do not agree with this result, since we believe that an off-rate experiment should not be performed, as in their case, when the protein concentration (240nM) is in large excess over the concentration of DNA (21nM). Considering their reported binding constant (5.6 ± 0.9nM), the liberated radiolabelled DNA will be recaptured by the excess protein with a rate greater than that of the release, and no decay in the radioactivity of the complex should be seen. Liao *et al.* nevertheless detect a first order (or pseudo first-order) rate constant, which may represent the rate of inactivation of the protein under the experimental conditions.

DISCUSSION

Zinc fingers are largely independent mini-domains which can be used to recognise particular nucleic acid sequences. Sequence specificity is the result of mostly direct readout involving three adjacent bases on DNA and three residues found at invariant positions on the helix of a zinc finger (17). Because each zinc finger can only recognise a base triplet, fingers need to be arranged in tandem in order to have the high specificity required by certain biological functions, especially the recognition of naturally occuring operator sequences. Many zinc fingers are so arranged, connected by an invariant 'linker' of the sequence: Thr-Gly-Glu-Lys-Pro (2). We believe that the proportion of known finger proteins with such *Kruppel*-type linkers may be exaggerated over the natural occurence, owing to the frequent use of this sequence as a probe in cloning efforts. This bias notwithstanding, it is remarkable that such screens of genomic libraries predict the occurrence of thousands of *Kruppel*-linked fingers (18).

The presence of a conserved peptide sequence between the last histidine of one finger and the first residue of the hydrophobic core of the next finger is significant because this region is a division between successive structural/functional units. The relation of one unit to the next, is then to a certain extent dependent or at least constrained by the linker. Hence one has come to expect a particular mode of DNA binding to be associated with the *Kruppel*-type linker, which has popularly become a diagnostic for major groove occupancy by the linked fingers. However our understanding of the function of linkers has not been greatly advanced by previous structural studies of zinc fingers, which would imply a passive role for them.

We do not know of any previous direct mutagenesis studies on the linker region of zinc finger proteins. The importance of the linker in zinc finger function has nevertheless become apparent in scanning mutagenesis of finger regions in both ADR1 (19) and Zif268 (20). Considering the implications of the crystal structure it has been surprising that in the latter study, about one quarter of the deleterious *in vivo* mutations detected in Zif268 lie in the linker regions. Equally surprising is the subsequent finding that a number of these mutations do not affect DNA binding.

We have replaced the two linkers of TFIIIA fingers 1−3 with the linkers from the RNA binding protein p43, and have observed the loss of DNA binding. In a rather similar experiment Clemens *et al.* (21) allude to a role for the *Kruppel*-type linker in allowing TFIIIA but not p43 to bind the ICR: we find this unlikely since the primary seat of sequence discrimination in zinc fingers is the α-helix, and not the linker. Replacement of the first linker of TFIIIA fingers 1−3 with the linker connecting fingers 3 and 4 from the same protein, also has the effect of abolishing DNA binding. The reason for these replacements was to test whether the linkers can be replaced in their entirety by sequences, which although different, are derived from either the same protein or from a quasihomologous protein. In the light of our results, the linker should be viewed as a sequence which has to be compatible with the fingers which it connects.

Kruppel-type linkers therefore appear to be ones which will accommodate fingers which wrap around DNA in the major groove. The precise nature of the interaction however cannot be inferred, since interestingly, the Pro and Asn residues in the linkers of TF3A are not interchangeable (Fig. 5) suggesting that the relationship of fingers 1 and 2 is not exactly the same as that of fingers 2 and 3. But equally it would not be correct to argue that fingers which are not connected by a *Kruppel*-type linker cannot occupy the major groove as does Zif268, since the recent crystal structure of the *tramtrack*-DNA complex shows overall Zif268-like binding but with extended linkers which are one residue longer and which are unrelated in sequence (L. Fairall, private communication).

Mutants in which single amino acid changes have been made, reveal more subtle effects of the linker on DNA binding (Fig. 5). We created T38L in order to test the importance of the -OH group in forming a hydrogen bond to the backbone of the linker, as seen in the crystal structure of Zif268, and found a 10-fold reduction in the binding constant. The largest effect we have observed is in the mutant G39P, where the binding constant was reduced 24-fold. This is to be expected as Gly has exceptional conformational freedom, whereas Pro is very constrained, and in the Zif268 structure Gly adopts a conformation ($\Phi \sim 80°$, $\Psi \sim 10°$) which is usually not attained by Pro (Φ is usually

~ −60). The mutant E40K shows wild-type binding, whereas in K41S we observe a 7-fold reduction in the binding constant. We think it may be structurally feasible for K41 to make a contact to the DNA backbone: on the basis of the position of homologous residues in the Zif268-DNA crystal and in analogy with the phosphate contact made by R2 in the flexible N-terminal sequence of Zif268 (5). If we assume that K41 makes a phosphate contact to the DNA, then it is clear that E40 must face away from the DNA and can play no part in binding. When P42 is changed to G, a 13-fold reduction in the binding constant occurs, suggesting that a certain conformational restriction is desirable in the C-terminus of the linker, to be got by limiting the F angle or by an interaction between Pro and the first conserved aromatic residue as noted (5). It is interesting that in the second linker of TF3A, Pro has drifted to Asn but can still be tolerated, whereas Pro in the first linker cannot be replaced by Asn without some reduction in the binding constant. Insertion of an extra Ser in the linker also reduces the binding affinity by a factor of about 10.

Although we have found it interesting to discuss our results in the context of the Zif268 structure (5), the precise mode of interaction of TF3A with the ICR is not known, and may be subtly different. Furthermore, while we have shown the linker sequence not to be arbitrary, we are not sure whether its function is to do with flexibility and the positioning of fingers, or with contacts to the DNA while bound. In the first case, mutations are expected to affect the equilibrium binding constant via the rate for complex association, while in the second instance the rate of complex dissociation will likely be perturbed. We have therefore tried to measure the half life of the TF3A-ICR complex and have found it to be under 5min, and hence too fast to be of use in resolving this ambiguity. We believe that the earlier value of 100min reported by Liao *et al.* is misleading as discussed above.

A weak interaction of adjacent zinc fingers in solution has been described by Omichinski *et al.* (22), who have studied a pair of fingers from the human enhancer binding protein MBP-1. This is said to be the result of an interaction between a Thr residue from the helix of one finger and a Lys residue from the edge of the β-sheet in the next finger, with an additional hydrophobic interaction from a Val residue (Val27) found in the linker. With the exception of the last contribution, a similar (reciprocal) inter-finger contact was reported for Zif268 when bound to DNA. The linker region however deviates from the canonical sequence we have been discussing, and such a stabilising effect cannot yet be extended to the *Kruppel* linker in the absence of structural studies.

From the experiments on the first three fingers of TFIIIA described above, we have been able to show that the *Kruppel*-type linker is not a passive polypeptide in the interaction with DNA; on the contrary, that the particular sequence is required for high affinity binding to the ICR. It appears likely to us that the sequence is one which is symptomatic of Zif268-type binding to DNA, although it is clear that others can function equally well in a similar capacity. We are hence unable to account fully for the remarkable evolutionary conservation of this sequence in terms of its requirement in DNA binding.

However we do not yet know much about the interaction of zinc finger proteins with other cellular components. It is therefore possible, for example, that a conserved amino acid sequence in zinc finger transcription factors interacts with other factors in the transcription apparatus (23). The possibility also exists that the particular sequence of the *Kruppel*-type linker is a requirement for the evolution of multi-fingered proteins of the *Kruppel*-type. The division of structural/functional units at the level of protein, is also apparent at the level of the gene where nucleotide sequences coding for the *Kruppel* linker delimit genetic information relating to single fingers. We have considered for some time the possible evolutionary advantage to finger proteins, of recombining modules with specific DNA binding functions, thereby giving rise to proteins with novel specificities. The existence therefore of a high homology region between fingers, could be the key to the preponderance of zinc finger DNA binding proteins. We would like to point out the striking N-terminal to C-terminal gradient in conservation of the *Kruppel* linker as revealed by Jacobs (24), which is reminiscent of meiotic gene conversion gradients caused by site-specific recombination but on a quite different scale (25). The full significance of the wide occurence of the *Kruppel*-type linker is yet to be understood.

ACKNOWLEDGEMENTS

We are grateful to P.J.G.Butler for help with data analysis. We also wish to thank J.W.R.Schwabe and D.Rhodes for discussions. Y.Choo is the recipient of a partial Medical Research Council grant.

REFERENCES

1. Miller, J., A. D. McLachlan and A. Klug (1985) *EMBO J*, **4**, 1609−1614.
2. Berg,J.M. (1993) *Curr Opin Struct Biol*, **3**, 11−16.
3. Berg, J. M. (1988) *Proc. Natl. Acad. Sci.*, **85**, 99−102.
4. Lee, M. S., G. P. Gippert, K. V. Soman, D. A. Case and P. E. Wright (1989) *Science*, **245**, 635−637.
5. Pavletich, N. P. and C. O. Pabo (1991) *Science*, **252**, 809−817.
6. Smith, D. R., I. J. Jackson and D. D. Brown (1984) *Cell*, **37**, 645−652.
7. Pelham, H. R. B. and D. D. Brown (1980) *Proc. Natl. Acad. Sci. USA.*, **77**, 4170−4174.
8. Joho, K. E., M. K. Darby, E. T. Crawford and D. D. Brown (1990) *Cell*, **61**, 293−300.
9. Nakaseko, Y., D. Neuhaus, A. Klug and D. Rhodes (1992) *J. Mol. Biol.*, **228**, 619−636
10. Christensen, J. H., P. K. Hansen, O. Lillelund and H.C. Thogersen (1991) *FEBS Letters*, **281**, 181−184.
11. Liao, X., K. R. Clemens, L. Tennant, P. E. Wright and J. M. Gottesfeld (1992) *J. Mol. Biol.*, **223**, 857−871.
12. Fairall, L., S. D. Harrison, A. A. Travers and D. Rhodes (1992) *J. Mol. Biol.*, **226**, 349−366.
13. Kunkel, T. A., J. D. Roberts and R. A. Zakour (1987) *Methods Enzymol.*, **154**, 367−382.
14. Studier, F. W., A. H. Rosenberg, J. J. Dunn and J. W. Dubendorff (1990) *Methods Enzymol.*, **185**, 60−89.
15. Fairall, L., D. Rhodes and A. Klug (1986) *J. Mol. Biol.*, **192**, 577−591.
16. Riggs, A. D., H. Suzuki and S. Bourgeois (1970) *J. Mol. Biol.*, **48**, 67−93.
17. Nardelli, J., T. J. Gibson, C. Vesque and P. Charnay (1991) *Nature*, **349**, 175−178.
18. Hoovers, J. M. N., M. Mannens, R. John, J. Bliek, V. van Heyningen, D. J. Porteous, N. J. Leschot, A. Westerveld and P. F. R. Little (1992) *Genomics*, **12**, 254−263.
19. Thukral, S. K., M. L. Morrison and E. T. Young (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9188−9192.
20. Wilson, T. E., M. L. Day, T. Pexton, K. A. Padgett, M. Johnston and J. Milbrandt (1992) *J.Biol. Chem.*, **267**, 3718−3724.
21. Clemens, K. R., V. Wolf, S. J. McBryant, P. Zhang, X. Liao, P. E. Wright and J. M. Gottesfeld (1993) *Science*, **260**, 530−533.
22. Omichinski, J. G., G. M. Clore, M. Robien, K. Sakaguchi, E. Appella and A. M. Gronenborn (1992) *Biochemistry*, **31**, 3907−3917.
23. Braun, B. R., B. Bartholomew, G. A. Kassavetis and E. P. Geiduschek (1992) *J. Mol. Biol.* **228**, 1063−1077.
24. Jacobs, G. H. (1992) *EMBO J.*, **11**, 4507−4517.
25. Malone, R. E., S. Bullard, S. Lundquist, S. Kim and T. Tarkowski (1992) *Nature*, **359**, 154−155.