

Review

The discovery of zinc fingers and their development for practical applications in gene regulation

By Aaron KLUG^{*)}

MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, United Kingdom

(Communicated by Takashi SUGIMURA, M. J. A.)

Abstract: An account is given of the discovery in 1985 of the classical Cys₂His₂ (C₂H₂) zinc finger, arising from the interpretation of biochemical studies on the interaction of the *Xenopus* protein transcription factor IIIA with 5S RNA, and of subsequent structural studies on its 3D structure and its interaction with DNA. Each finger constitutes a self-contained domain stabilized by a zinc ion ligated to a pair of cysteines and a pair of histidines, and by an inner structural hydrophobic core. This work showed not only a new protein fold but also a novel principle of DNA recognition. Whereas other DNA binding proteins generally make use of the two-fold symmetry of the double helix, functioning as homo- or heterodimers, zinc fingers can be linked linearly in tandem to recognize nucleic acid sequences of different lengths. This modular design offers a large number of combinatorial possibilities for the specific recognition of DNA (or RNA). It is therefore not surprising that this zinc finger is found widespread in nature, in 3% of the genes of the human genome. It had long been the goal of molecular biologists to design DNA binding proteins for control of gene expression and we have adopted the zinc finger design and principle for this purpose. We demonstrated that the zinc finger design is ideally suited for such purposes, discriminating between closely related DNA sequences both *in vitro* and *in vivo*, and we have therefore adapted this natural design for engineering zinc finger proteins for targeting specific genes. The first example of the potential of the method was published in 1994 when a three-finger protein was constructed to block the expression of an oncogene transformed into a mouse cell line. In the same paper we also showed that we could activate a reporter gene by targeting a nine base pair promoter which we had inserted. Thus by fusing zinc finger peptides to repression or activation domains, genes can be selectively switched off or on. By combining the targeting zinc fingers with other effector or functional domains e.g. from nucleases or integrases, to form chimeric proteins, genomes can be manipulated or modified. Several applications of such engineered zinc finger proteins are described here, including some of potential therapeutic importance.

Key words: Zinc fingers; transcription activation; transcription inhibition; modular design; combinatorial principle; protein engineering; human genome; gene targeting; gene correction.

Introduction. After ten years of research on the structure of chromatin which had led to the discovery of the nucleosome and an outline of its structure, as well as the next level of folding of DNA in the 300 Å chromatin

fibre,^{1,2)} I became interested in the then so-called “active chromatin,” the chromatin which is involved in transcription or poised to do so. I looked for a system which was tractable, that is, offered the possibility of extracting relatively large amounts of material for biochemical and structural studies.

I became intrigued by the work of Robert Roeder and Donald Brown on the 5S RNA genes of *Xenopus laevis*, which are transcribed by RNA polymerase III (reviewed in Brown).³⁾ They discovered that the correct

Abbreviations: DBD, DNA binding domain; ZFP, zinc finger protein; VEGF, vascular endothelial growth factor; DTT, dithiothreitol; 7S RNP, 7S Ribonucleoprotein particle; HR, homologous recombination; HDR, homology directed repair; DSB, double stranded break; ZFN, zinc finger protein nuclease.

^{*)} Honorary Member of the Japan Academy.

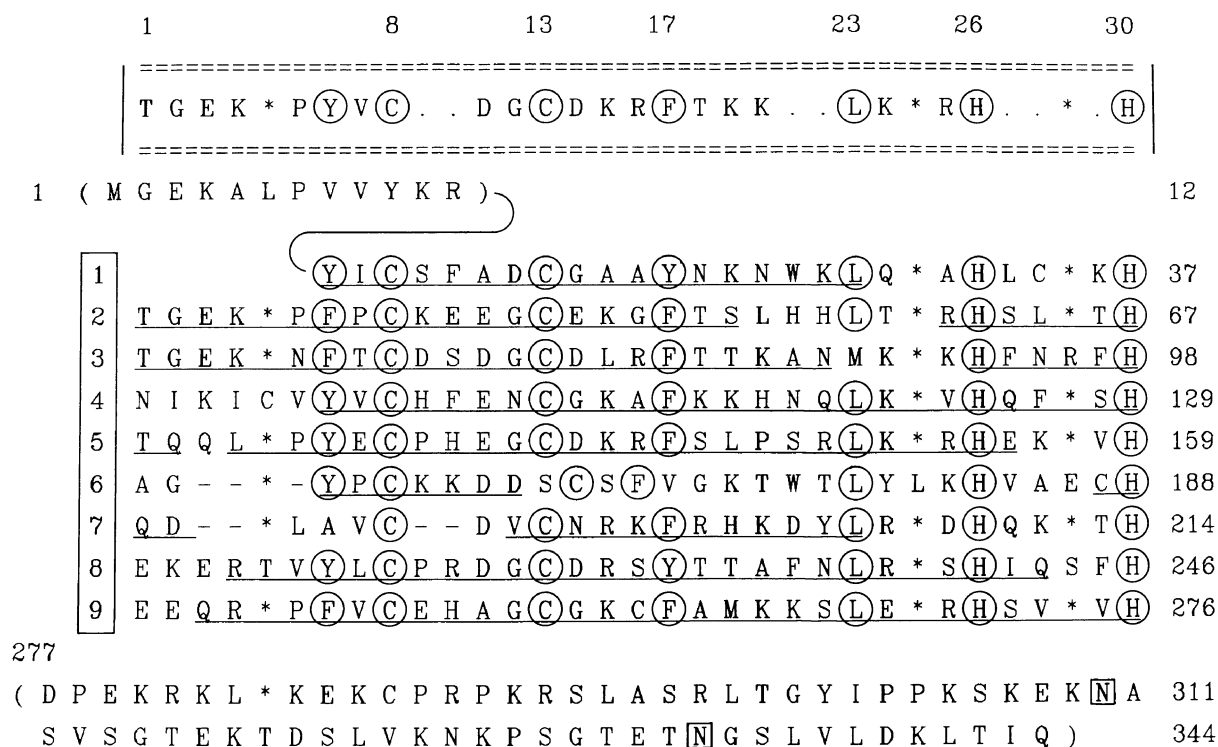


Fig. 1. Amino acid sequence⁹⁾ of Transcription Factor IIIA from *X. laevis* oocytes aligned to show the repeating units.⁶⁾ The best conserved residues are ringed. The boxed-in consensus sequences is shown at the top, indexed on a repeat unit of 30 amino acids. The structural units or finger modules (Fig. 2) begin at amino acid number 6, separated by a five amino acid linker.

initiation of transcription requires the binding of a 40 kD protein factor, variously called factor A or transcription factor IIIA (TFIIIA), which had been purified from oocyte extracts. By deletion mapping it was found that this factor interacts with a region about 50 nucleotides long within the gene, called the internal control region. This was the first eukaryotic transcription factor to be described.

Immature oocytes store 5S RNA molecules in the form of 7S ribonucleoprotein particles,⁴⁾ each containing a single 40 kD protein which was later shown⁵⁾ to be identical with transcription factor IIIA. TFIIIA therefore binds both 5S RNA and its cognate DNA and it was therefore suggested that it may mediate autoregulation of 5S gene transcription.⁵⁾ Whether this autoregulation occurred *in vivo* or not, the dual interaction provided an interesting structural problem which could be approached because of the presence of large quantities of the protein TFIIIA in immature *Xenopus* oocytes.

In the autumn of 1982, I therefore proposed to a new graduate student, Jonathan Miller, that he begin

studies on TFIIIA. This led to the discovery of a remarkable repeating motif within the protein, which we later, in laboratory jargon, called zinc fingers, because they contained zinc and gripped or grasped the DNA. The full story of the experiments is told in our first paper⁶⁾ and I will only summarize it here. I should however emphasize that the repeating structure was discovered through biochemistry, not, as some reviews have stated, by computer sequence analysis. When the sequence was published I looked for, and found by eye, a repeating pattern, which was then confirmed and aligned as a repeating motif of 30 amino acids by Andrew McLachlan's computer analysis⁶⁾ (Fig. 1).

Preparation and characterisation of TFIIIA from the 7S RNP particle. When Miller repeated the published protocols for purifying the 7S particle, he obtained very low yields, which we attributed to dissociation. Brown and Roeder had used buffers which contained variously dithiothreitol (DTT), used because the protein had a high cysteine content, and EDTA to remove any contamination by metals which hydrolyse

nucleic acids. We observed that gel filtration of the complex in 0.1 mM DTT resulted in separate elution of protein and 5S RNA. However, when we found that the strong reducing agent sodium borohydride did not disrupt the complex, we realized that the protein was not being held together by disulfide bridges, and that therefore a metal might be involved. When the particle was incubated with a variety of chelating agents, particle dissociation could be prevented only by prior addition of Zn^{2+} , and not by a variety of other metals. Analysis of a partially purified 7S preparation by atomic absorption spectroscopy also revealed a significant concentration of Zn, with at least 5 mol Zn per mol particle.

While these experiments were in progress, Hanas *et al.*⁷ reported the presence of Zn in the 7S particle, at a ratio of two per particle. We thought this must be an underestimate since their buffers contained 0.5 mM or 1 mM DTT, which has a high binding constant for Zn of about 10^{10} . We therefore repeated the analysis with pure and undissociated particle preparations, taking great care to ensure no contamination. We concluded that the native 7S particle contains between 7 and 11 zinc ions. This result was consistent with the fact that the protein was known to contain large numbers of histidine and cysteine residues, the commonest ligands for zinc in enzymes and other proteins. This hinted at some kind of internal substructure.

A natural step was therefore to see if any such substructure could be revealed by proteolytic digestion, and Miller had already begun such studies. He had found two products, an intermediate 33 kD fragment, and a limit 23 kD. At about that time Brown's group⁸ also showed that, on treatment with proteolytic enzymes, the 40 kD TFIIIA protein breaks down to a 30 kD product, which is then converted to a 20 kD product. They proposed that TFIIIA consists of three structural domains which they identified as binding to different parts of the 50 base-pair internal control region of the 5S RNA gene.

Carrying our proteolytic studies further, we found that on prolonged proteolysis the TFIIIA product breaks down further, finally to a limit digest of about 3 kD. In the course of this breakdown, periodic intermediates differing in size by about 3 kD could be seen. The correspondence in size between these last two values suggested that the 30 kD domain of TFIIIA might contain a periodic arrangement of small, compact domains each of 3 kD. If each such domain contained one Zn atom, then the observed high Zn content would be accounted for.

This novel idea of a small Zn-stabilized domains was

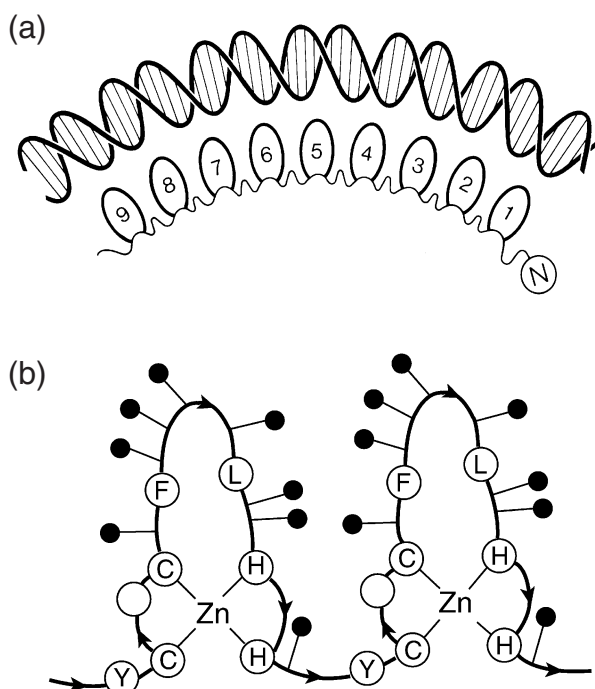


Fig. 2. (a) Interpretation⁶ of the main structural feature of the protein TFIIIA and its interactions with 50 bp of DNA, showing combinatorial recognition by the modular design. (b) Folding scheme for a linear arrangement of repeating structural units ("zinc finger modules"), each centred on a tetrahedral arrangement of zinc ligands, Cys_2His_2 . Also shown are the three hydrophobic groups which were proposed to form a structural core, as confirmed later (Fig. 4 below).

strengthened by the timely publication by Roeder's group⁹ of the sequence of TFIIIA derived from a cDNA clone. By inspection, it could be seen that the large number of cysteines and histidines present in the protein appeared to occur in a more or less regular pattern. A rigorous computer analysis showed that, of the 344 amino acids of the TFIIIA sequence residues, numbers 13-276 form a continuous run of nine tandemly repeated, similar units of about 30 amino acids, each containing two invariant pairs of histidines and cysteines⁶ (Fig. 1). Repeating patterns in the sequence were also noticed by Brown, Sander & Argos¹⁰ who concluded, however, that the whole protein was divided into twelve repeats, indexed on a 39 amino acid unit (although their abstract states "about 30").

A repeating structure for TFIIIA. From the three different lines of evidence described above, namely (1) a 30 amino acid repeat in the sequence, which (2) corresponds in size to the observed periodic

Number of C2H2 & C4 zinc finger genes

	Total number of genes	C2H2	C4
● Human	23,299	709(3%)	48(0.21%)
● Mouse	24,948	573(2.3%)	42(0.17%)
● Rat	21,276	466(2.2%)	43(0.2%)
● Zebrafish	20,062	344(1.7%)	53(0.26%)
● Drosophila	13,525	298(2.2%)	21(0.16%)
● Anopheles	14,653	296(2.0%)	20(0.14%)
● C.elegans	19,564	173(0.88%)	270(1.3%)
● C.briggsae	11,884	115(0.9%)	167(1.4%)

Fig. 3. Number of Cys₂His₂ and “Cys₄” genes in the genomes of various organisms.¹³⁾ The Cys₄ genes refer to the steroid and thyroid hormone family of nuclear receptors which were also later named zinc fingers because of a similarity in the amino acid sequences suggesting the presence of zinc tetrahedrally ligated to four cysteines. The structures are however quite different from the classical C₂H₂ finger. Note how the proportion of C₂H₂ genes increases with greater complexity on the evolutionary scale.

intermediates and the limit-digest product of 3 kD, and (3) the measured Zn content of 7-11 atoms, we proposed⁶⁾ that the 30 kD region of the TFIIIA protein has a repeating structure consisting of nine 30 amino acid units (Fig. 2a). 25 of the 30 amino acids in the repeat fold around a Zn ion to form a small independent structural domain or module, the “finger”, and the five intervening amino acids provide the linkers between consecutive fingers (Fig. 2b). The Zn ion forms the basis of the folding by being tetrahedrally coordinated to the two invariant pairs of cysteines and histidines. Each repeat also contains, besides this unique conserved pattern of Cys-Cys His-His, three other conserved amino acids, namely Tyr6 (or Phe6), Phe17 and Leu23, all of which are large hydrophobic residues (Fig. 1). We suggested that these might interact to form a hydrophobic cluster stabilising the compact finger module. The whole of the 30 amino acid repeat is rich in basic and polar residues but the largest number are found concentrated in the region between the second cysteine and the first histidine, implicating this region in particular in nucleic acid binding. This was later found to be the case.

Formally, when indexed on a 30 amino acid repeat, the repeating structure could be written as

1 – 5 6 8 13 17 23 26 30
linker h X₁ C X_{2,4} C X₃ h X₅ h X_{2,3} H X_{3,4} H

where h represents a conserved (large) hydrophobic residue. The proposal that each 25 amino acid module formed an independently folded, Zn-stabilized domain soon gained support from two lines of research. First, we carried out a study using EXAFS (extended X-ray absorption fine structure) confirming that the Zn ligands are two cysteines and two histidines.¹¹⁾ Secondly it was found by Tso *et al.*¹²⁾ that, in the DNA sequence of the gene for TFIIIA, the position of the intron-exon boundaries mark most of the proposed finger module domains.

In evolutionary terms, the multi-fingered TFIIIA may have arisen by gene duplication of an ancestral domain comprising about 30 amino acids. Because one such self-contained small domain would have had the ability to bind to nucleic acids, and could be passed on by exon shuffling, we suggested⁶⁾ that these domains might occur more widely in gene control proteins than in just this case of TFIIIA. The extent to which this prediction has been borne out (3% of the genes of the human genome, at the latest count¹³⁾) still, on occasion, astonishes me. Indeed, within months of the publication of our paper, I received word of sequences homologous to the zinc finger motif of TFIIIA. The first two were from *Drosophila*, the serendipity gene from Rosbash's group¹⁴⁾ and Krüppel from Jäckle's group.¹⁵⁾

A new principle of DNA recognition. The key point that emerged from our first paper was that, not

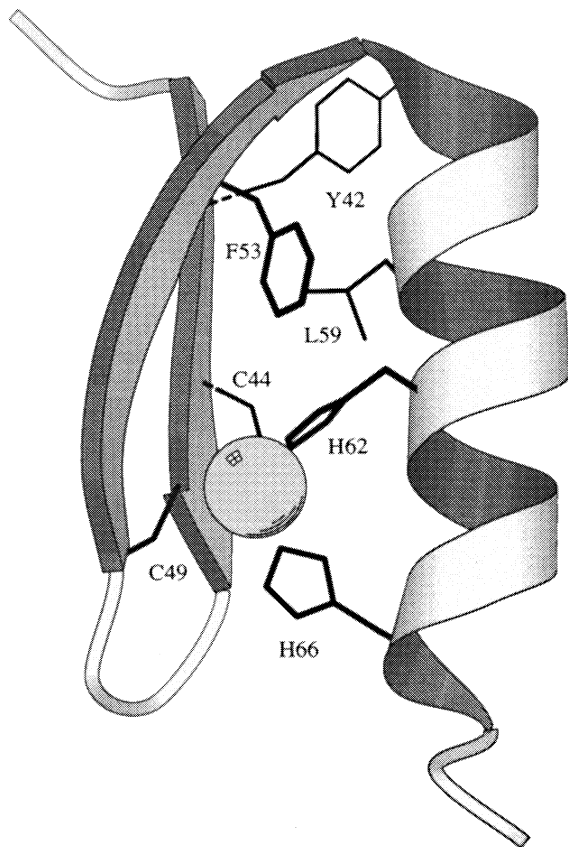


Fig. 4. The structure of a zinc finger from a 2D-NMR study of a two-finger peptide in solution.¹⁹⁾ The same study showed that the linker between the two modules is highly flexible.¹⁸⁾

only had there emerged a new protein fold for nucleic acid binding, but also a novel principle of DNA recognition. The overall design for specific DNA recognition was distinctly different from that of the helix-turn-helix motif, found in the first DNA binding proteins to be described. The latter binds to DNA as a symmetric dimer to a palindromic sequence on the DNA, thus making use of both the 2-fold symmetry of the DNA helix backbones and also the nucleotide sequence. Heterodimeric variations of this and other 2-fold symmetric designs were found later, but they still make use of the double helix symmetry.

In contrast, the zinc finger is a DNA-binding module that can be linked tandemly in a linear, polar fashion to recognize DNA (or RNA) sequences of different lengths. Each finger domain has a similar structural framework but can achieve chemical distinctiveness

through variations in a number of key amino acid residues. This modular design thus offers a large number of combinational possibilities for the specific recognition of DNA (or RNA). It is not surprising that it is found widespread throughout so many different types of organisms (Fig. 3).

The structure of the zinc finger and its interaction with DNA. We had noted,⁶⁾ that in addition to the characteristic arrangements of conserved cysteines and histidines which are fundamental in the folding of the finger by the coordinating Zn, there are three other conserved amino acids, notably Tyr6, Phe17 and Leu23, and suggested that they were likely to form a hydrophobic structural core of the folded structure. In other words, the seven conserved amino acids in each unit would provide the framework of tertiary folding, whereas some of the variable residues determined the specificity of each domain. Jeremy Berg¹⁶⁾ built on these original observations by fitting known structural motifs from other metallo-proteins to the consensus sequence of the TFIIIA finger motifs. His proposed model consisted of an antiparallel β -sheet, which contains a loop formed by the two cysteines, and an α -helix containing the His-His loop. The two structural units are held together by the Zn ion. In analogy with the way in which the bacterial helix-turn-helix motif binds DNA, DNA recognition was postulated to reside mostly in the helical region of the protein structure.

Berg's model was soon confirmed in outline by the NMR studies of Peter Wright's group¹⁷⁾ on a single zinc finger in solution, and by David Neuhaus in our laboratory^{18),19)} on a two-finger peptide (Fig. 4). Our work took longer to solve the structure but it had the merit of showing that adjacent zinc fingers are structurally independent in solution, being joined by flexible linkers.

The question remained of the precise pattern of amino acid interactions of zinc fingers with DNA. The breakthrough came in 1991 when Nikola Pavletich and Carl Pabo,²⁰⁾ solved the crystal structure of a complex of a DNA oligonucleotide specifically bound to the three-finger DNA binding domain of the mouse transcription factor Zif268, an early response gene. The primary contacts are made by the α -helix which binds in the DNA major groove through specific hydrogen-bond interactions from amino acids at helical positions -1, 3 and 6 to three successive bases (a triplet) on one strand of the DNA (Fig. 5a). Later, the second zinc finger-DNA complex to be solved, by Fairall in Rhodes' group in our laboratory,²¹⁾ revealed an important secondary interaction from helical position 2 to the other strand (Fig. 5b and

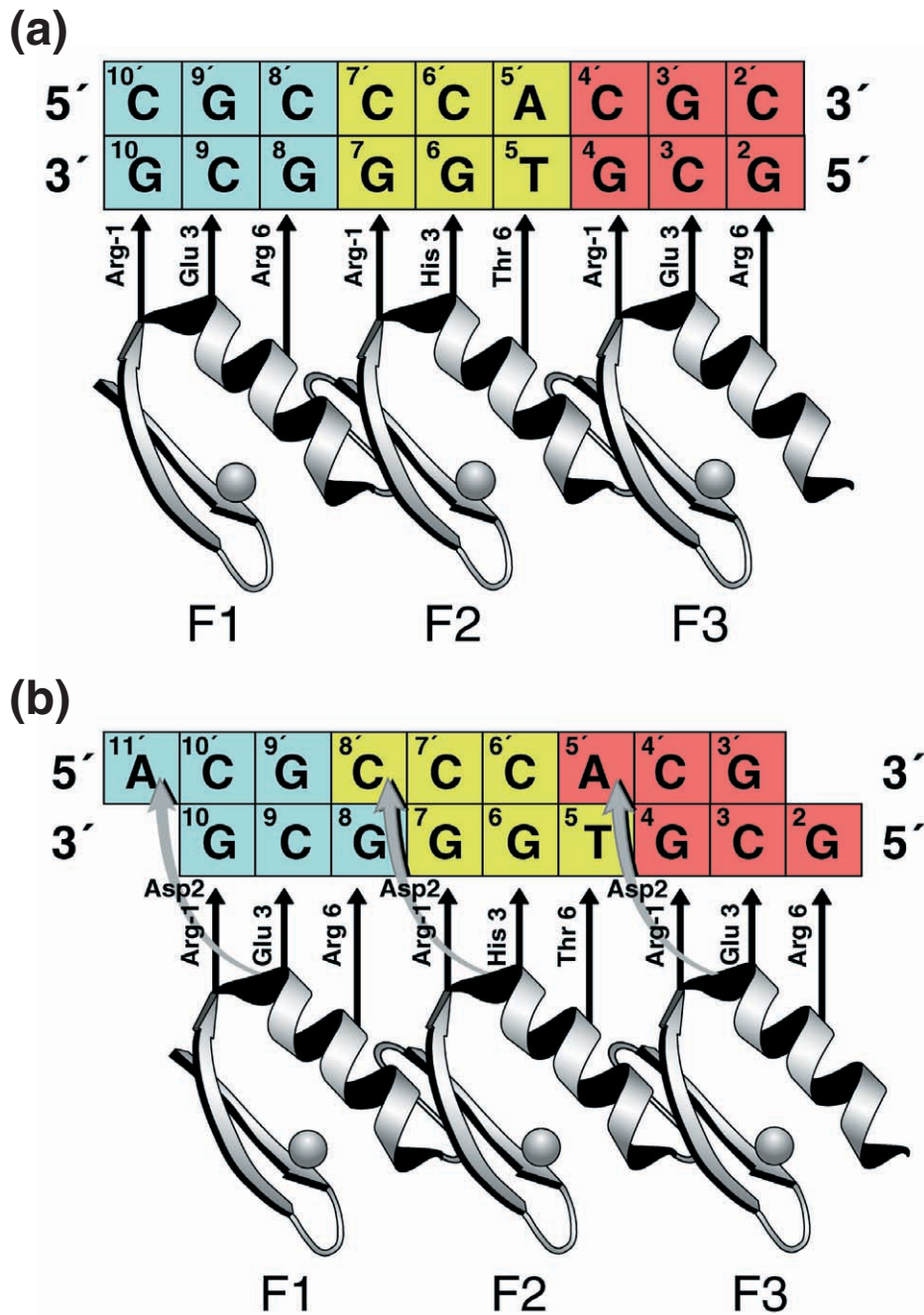


Fig. 5. (a) Schematic diagram²⁹⁾ of the first model of modular recognition of DNA by a three-zinc finger peptide, illustrating the results of the first crystal structure determination of the complex between the DNA binding domain of the transcription factor Zif268 and an optimised DNA binding site.²⁰⁾ Each finger interacts with a 3 bp subsite on one strand of the DNA, using amino acid residues in helical positions -1, 3 and 6. (b) Refined model of DNA recognition.^{29),30)} View of the potential hydrogen bonds to the second strand of the DNA ("cross-strand interactions") emanating from position 2 on the recognition helix. This is based on the crystal structure of the tramtrack-DNA complex,²¹⁾ the mutagenesis²⁹⁾ and phage display selection studies of Isalan *et al.*,³⁰⁾ and on the refined structure of the Zif268-DNA complex, and of variants by Pabo and his colleagues.⁴⁸⁾ The fingers ideally bind 4 bp overlapping subsites, so that adjacent fingers are functionally synergistic though structurally independent.

see below). This is the canonical docking arrangement, but there are however some wide variations from this arrangement in the family of zinc finger – DNA complexes now known. There are also, of course, still other interactions such as with the phosphates of the DNA backbones, but these do not play a direct part in specific recognition.

Zinc finger peptides for the regulation of gene expression. The mode of DNA recognition by a finger is thus principally a one-to-one interaction between individual amino acids from the recognition helix to individual DNA bases. This is quite unlike the case of other DNA-binding proteins, where one amino acid may contact two bases and vice versa. Moreover, because the fingers function as independent modules, fingers with different triplet specificities can be linked to give specific recognition of longer DNA sequences. For this reason, the zinc finger motifs are ideal natural building blocks for the *de novo* design of proteins for recognizing any given sequence of DNA. Indeed the first experiments by Berg and others, using site-directed mutagenesis, showed that it is possible to alter rationally the DNA-binding characteristics of individual zinc fingers when one or more of the amino acids in the α -helical positions are varied. As a small collection of these mutants accumulated, it became possible to find some regularities or “rules” relating amino acids on the recognition α -helix to corresponding bases in the bound DNA sequence.

In our laboratory, my colleagues and I adopted a different approach. The reason was that the “rules” did not take into account the fact that real DNA structures are not fixed in the canonical B form but, as we had shown, there are wide departures, depending on the DNA sequence.^{22),23)} This was further brought home to us by the structure of the zinc finger – tramtrack DNA complex.²¹⁾ Here the helical position used for the primary contact by the first finger with the 3'-most base of the cognate triplet (thymine) is not the canonical –1, but 2. The cause is that the DNA helix is deformed from the B form, with the thymine followed by an adenine at a helical rotation angle of 39°, rather than the canonical 36°, and preceded by another adenine at an angle of 33°. The interaction with the finger thus occurs at an ATA sequence which has unusual flexibility, as I had noted long ago.²²⁾ DNA is not a rigid, passive participant in its interaction with proteins.

Affinity selection from a library of zinc fingers by phage display. The alternative to this rational but biased design of proteins with new specificities was the

isolation of desirable variants from a large pool or library. A powerful method of selecting such proteins is the cloning of peptides²⁴⁾ or protein domains²⁵⁾ as fusions to the minor coat protein (pIII) of bacteriophage fd, which leads to their expression on the tip of the capsid. Phage displaying the peptides of interest can be affinity purified by binding to the target and then amplified in bacteria for use in further rounds of selection and for DNA sequencing of the cloned genes. We applied this technology to the study of zinc finger-DNA interactions, after my colleague, Yen Choo, had demonstrated that functional zinc finger proteins could be displayed on the surface of fd phage, and that such engineered phage could be captured on a solid support coated with the specific DNA.^{26),27)} The phage display method was also adopted by other groups working on zinc fingers, including those of Carl Pabo and Carlos Barbas.

We created phage display libraries comprising about 10^7 variants of the middle finger from the DNA-binding domain of Zif268. A DNA oligonucleotide of fixed sequence was used to bind and hence purify phage from this library over several rounds of selection, returning a number of different but related zinc fingers which bind the given DNA. By comparing similarities in the amino acid sequence of functionally equivalent fingers, we could deduce the likely mode of interaction of these fingers with DNA. Remarkably we found that most base contacts occur from three primary positions on the α -helix of the zinc finger, correlating well with the implications of the crystal structure of Zif268 bound to DNA.²⁰⁾

This demonstrated ability to select zinc fingers with desired specificity meant that, as the data from the selections accumulated, some wider general rules could be devised for a recognition code,²⁷⁾ and hence that DNA-binding peptides could be made to measure using the combinatorial strategy exemplified by TFIIIA. In other words these general rules could be used for the “rational design” of a zinc finger peptide to recognise a short run of DNA sequence by “mixing and matching” individual specific fingers. Where the general rules for finger specificity led to an ambiguity, as in the case of closely related triplets, e.g. GCG and GTG, we showed that zinc finger modules could nevertheless be selected to discriminate between them.²⁷⁾

Use of engineered zinc finger peptides to repress gene expression in a mouse cell line. One interesting possibility for the use of such zinc finger peptides is to target selectively genetic differences in

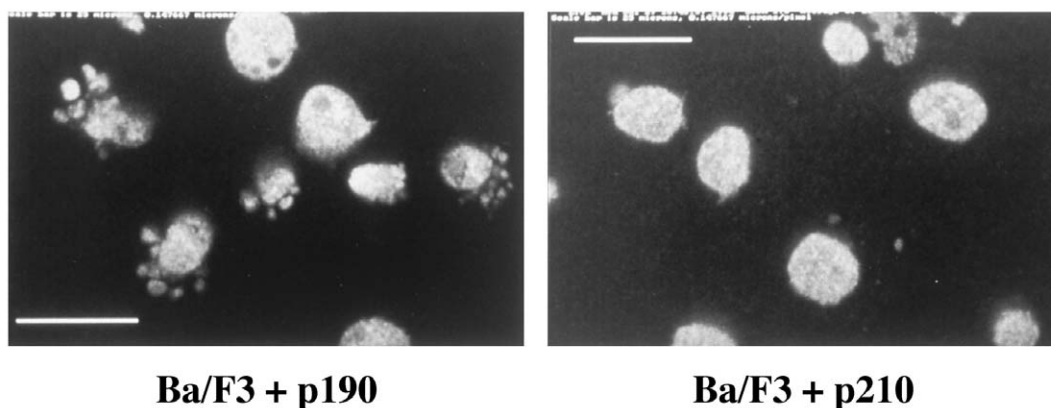


Fig. 6. **(left)** An engineered site-specific zinc finger DNA binding protein, designed *de novo* against an oncogenic BCR-ABL fusion sequence (p190) transformed into a BaF3 mouse cell line represses expression of the oncogene.²⁸⁾ Immunofluorescence image of cell nuclei eight hours after transfection with the zinc finger protein, showing apoptosis. After 24 hours 95% of the cells are destroyed. **(right)** The same on a control cell line which has been transformed by a related but different BCR-ABL oncogenic sequence p210. The cells are not affected.

pathogens or transformed cells. In December 1994 we reported the first such application,²⁸⁾ in which we built a protein which recognized a specific DNA sequence both *in vitro* and *in vivo*. This was a crucial test of our understanding of the mechanism of zinc-finger DNA recognition. The proof of principle led ourselves, and later others, to devote our future studies on zinc fingers to potential applications in gene regulation for research purposes or for therapeutic correction. It also stimulated the creation of the first biotech companies, Sangamo Biosciences and later Gendaq (see below), to exploit the new technology.

In summary, we created a three finger peptide able to bind site-specifically to a unique nine base-pair region of the p190 *bcr-abl* cDNA: this is a transforming oncogene which arises by translocation between the tips of chromosomes 9 and 22, of which one product is the Philadelphia chromosome. The latter contains a novel DNA sequence at the junction of two exons, one each from the two genomic parent *bcr* and *abl* genes. Our engineered peptide discriminated *in vitro* against like regions of the parent *bcr* and *c-abl* genes, differing in only a single base out of the nine base pair target, by factors greater than one order of magnitude.

Our peptide also contained a nuclear localization signal (NLS) fused to the zinc finger domain so that the peptide could accumulate in the nucleus. Consequently, stably transformed mouse cells, made interleukin-3 independent by the action of the oncogene, were found to revert to IL-3 dependence on transient

transfection with a vector expressing the peptide. Our construct was also engineered to contain a c-myc epitope, which enabled us to follow by immunofluorescence the localization of the peptide to the nuclei of the transfected cells. When IL-3 is subsequently withdrawn from cell culture, over 90% of the transfected p190 cells become apoptotic (that is, showing chromosome degradation) within 24 hours (Fig. 6, left). Our experiments were repeated on cells transformed by another related oncogene p210 *bcr-abl*, which served as a control. All transfected p210 cells maintained their IL-3 dependence, and remained intact on entry of the engineered peptide (Fig. 6, right).

Measurements of the levels of p190 *bcr-abl* mRNA extracted from cells treated with the peptide showed that the repression of oncogenic expression by the zinc finger peptide was due to a transcriptional block imposed by the sequence-specific binding of the peptide, which, with its highly basic NLS, presumably obstructed the path of the RNA polymerase. In later experiments to inhibit gene expression we used in addition a repression domain, such as the Kox domain from the Xenopus KRAB zinc finger family.

Promoter-specific activation by a chimeric zinc finger peptide. These experiments showed that a zinc finger peptide could be engineered to switch off gene expression *in vivo*. In the same paper²⁸⁾ we described other experiments on a different cell system (cultured mouse fibroblasts) to show that a gene could also be switched on in a similar way. We used the same

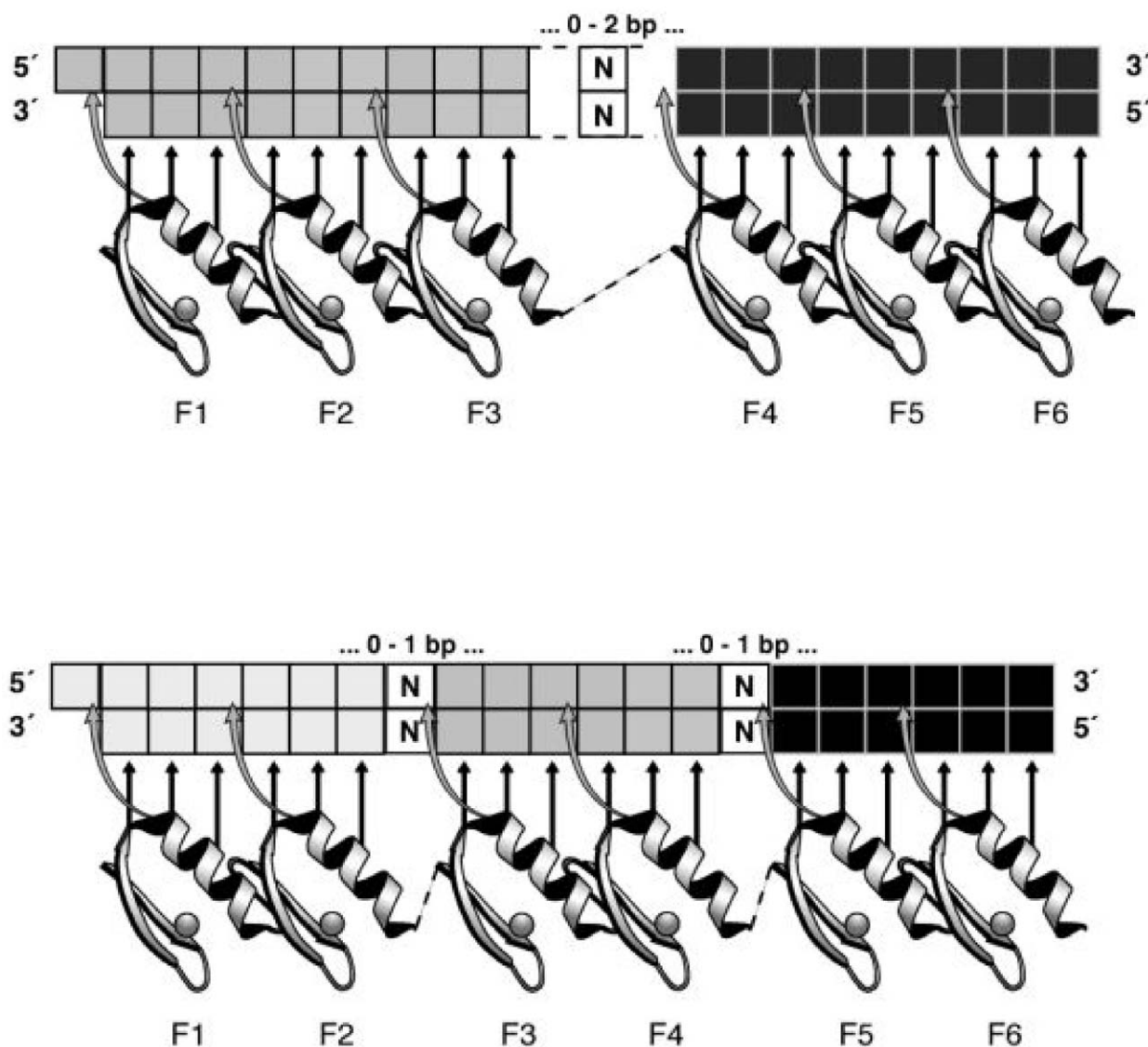


Fig. 7. Two modes of generating six-zinc finger proteins for specific recognition of 18 bp sequences.^{32),33)} (**Top**). Two three-finger peptides fused together using an extended canonical linker (2 × 3F scheme). (**Bottom**). Three two-finger peptides linked using canonical linkers extended by an insertion of either a glycine residue or a glycine-serine-glycine sequence in the canonical linkers between fingers 2 and 3 and fingers 4 and 5 respectively.

nine base pair sequence, but this time as a promoter for a CAT reporter gene contained in a plasmid. The peptide, which recognized the promoter sequence, was fused to a VP16 activation domain from Herpes Simplex virus and, on transient transfection, stimulated expression of the reporter gene by a factor of 30-fold above controls.

Improving zinc finger specificity by protein engineering. (i) Having shown proof of principle that engineered zinc finger peptides could be used to target

DNA, we focussed our subsequent work on improving the specificity of recognition. While the main source of specificity lies in the amino acids at positions -1, 3 and 6 of the recognition α -helix of a zinc finger for successive bases lying on one strand of a DNA triplet, we found that the “cross-strand” interaction described above from helical position 2 to the neighbouring base pair on the adjacent triplet (Fig. 5b) can significantly influence the specificity.²⁹⁾ Therefore it has been necessary to revise

the simple model that zinc fingers are essentially independent modules that bind three base-pair subsites to a model that considers functional synergy at the interface between adjacent independently folded zinc fingers. In this revised model, Zif268-like zinc fingers potentially bind four base-pair overlapping subsites³⁰⁾ (Fig. 5b).

We have therefore redesigned³¹⁾ our method of phage library construction to allow for the optimisation of the interaction that a finger makes with the DNA binding site of the adjacent N terminal finger. We adopted a bipartite selection strategy in which two halves of a three-finger DNA-binding domain are selected separately and then recombined *in vitro* to make a complete three-finger domain which binds nine base pairs and in which the interface synergy between its constituent three fingers has been automatically allowed. These two separate (non-overlapping) libraries are then used to perform the bipartite selections. Because all the steps are carried out *in vitro*, the method is rapid and easily adapted to a high-throughput automated format. This was applied commercially by an MRC spin-off company Gendag (later acquired by a Californian biotech company Sangamo Biosciences) for up to $2 \times 4^5 = 2048$ binding sites to create a large archive of zinc finger peptides which recognise a vast number of nine DNA base pair sequences.

(ii) An important step forward has been to increase the length of the DNA sequence targeted and hence its degree of rarity. At the same time there would be expected an increase in the binding affinity of the longer zinc finger construct. Three zinc fingers recognize nine base pairs, a sequence which would occur randomly many thousand times in a large genome. Therefore six fingers linked together would recognize a DNA sequence 18 base pairs in length, sufficiently long to constitute a rare address in the human genome. However one cannot simply go on adding fingers with the conventional linkers, because the periodicity of the packed fingers does not quite match the DNA periodicity. They thus tend to get out of register and are strained in doing so, leading to only a small increase in affinity. In recent years, we have learned how to engineer longer runs of zinc fingers which can target longer DNA sequences and have affinities a thousand times greater than three-finger peptides^{32),33)} (Fig. 7).

An early design³⁴⁾ was to use a longer more flexible linker between two different pre-existing three-finger domains to form a six-finger peptide but this has not been used much in practical applications. The method of choice in current use is that developed in Cambridge by

Moore *et al.*³³⁾ (and later transferred to Sangamo *via* Gendag). This uses two-finger binding domains which can be obtained from the archive described above. Three of these are assembled into six-finger domains, using longer variants of the conserved TGEKP linkers (Fig. 7, bottom). The linkers contain an extra glycine residue or a glycine-serine-glycine tripeptide between the constituent two-finger modules. Such six-finger domains (denoted $3 \times 2F$) bind their 18 bp targets with picomolar affinity, as also do six fingers made by linking two three-finger domains ($2 \times 3F$) made with an extended linker.³³⁾ However the advantage of the $3 \times 2F$ over the $2 \times 3F$ strategy is that it is much more sensitive to a mutation or an insertion in the target sequence, with a loss of affinity of up to a 100-fold. Thus the $3 \times 2F$ peptide discriminates more strongly than the $2 \times 3F$ peptide between closely related DNA target sequences. The logic behind this $3 \times 2F$ design was that the strain in binding longer DNA targets is more evenly distributed than in a $2 \times 3F$ construct, and it indeed turned out to have the advantage described. These six-finger peptides not only have picomolar affinities for their 18 bp targets but also give virtually single gene recognition³⁵⁾ when tested on DNA microarrays displaying twenty thousand different sequences. A library of two-finger peptides was begun in Cambridge and later transferred to the archive at Gendag-Sangamo.

Another strategy we developed³²⁾ was to target two non-contiguous 9 base pair DNA sequences separated by up to 10 bp of non bound DNA. Using a non-specific binding finger (in which all key amino acids had been mutated to serine) as a structured linker, we found that it could span a gap of 7 to 8 bp and maintain picomolar affinity. In contrast, the use of a flexible linker such as $(GSG)_n$ displayed no preference for a length of span, but the affinity was reduced to ~ 50 pM, probably attributable to the increased conformational entropy of the long peptide. These strategies have not so far been deployed.

Some applications of engineered zinc-finger proteins. Zinc finger proteins (ZFPs) can be engineered with a variety of effector domains fused to polypeptide zinc finger peptides which can recognise virtually any desired DNA sequence with high affinity and specificity. They thus form the basis of a novel technology which has increasing uses in research and medicine. An excellent summary of numerous such applications is given by Pabo and his colleagues.³⁶⁾ I will here only mention a few by ourselves and others.

(i) Inhibition of HIV expression.³⁷⁾ It was shown that ZFPs targeted to the HIV promoter long terminal

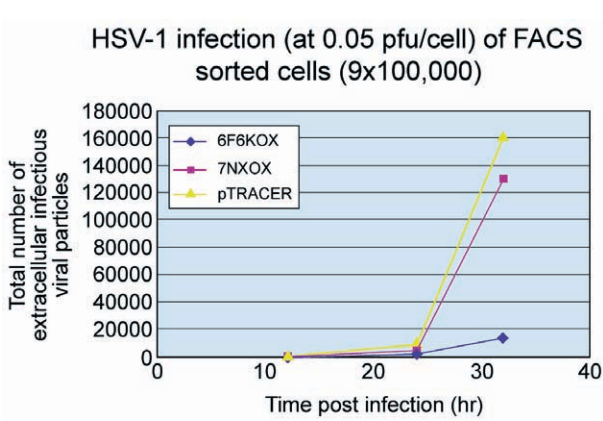


Fig. 8. Repression of herpes simplex virus infection(at 0.05 pfu/cell) of FACS sorted HeLa cells by zinc finger peptides fused to a KOX1 repression domain.³⁸⁾ The peptides are targeted to the viral gene IE175K, the first of the six immediate early genes to be expressed by the virus. A six-finger recognition peptide 6F6K0X reduces the virus titre ten-fold, while a three-finger peptide produces only a 3-fold reduction.

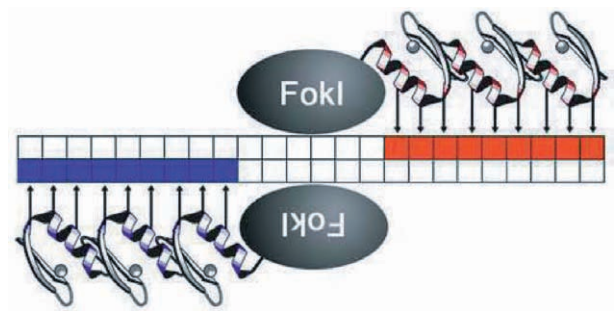


Fig. 10. Gene correction using a pair of three-zinc finger protein nucleases (ZFNs)⁴⁴⁾ to produce a double stranded DNA break. The zinc finger peptides are linked to the non-specific catalytic domain of the FokI endonuclease by a short amino acid linker. In the 3D-structure the two catalytic domains form a dimeric association.

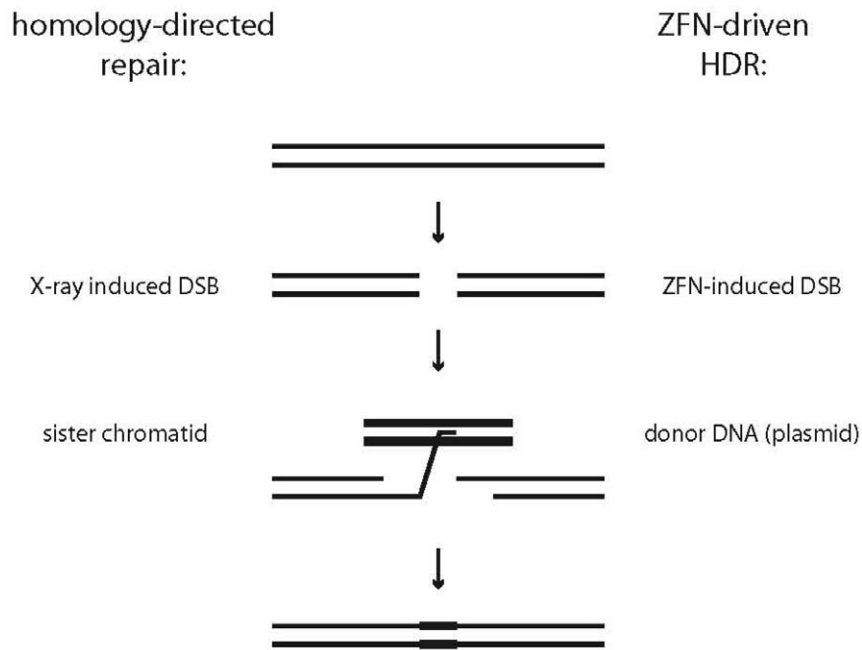


Fig. 9. Gene modification or correction using homologous recombination⁴²⁾ via “short-path gene conversion” stimulated by a double stranded break (DSB).⁴³⁾ The left side schematically depicts the repair of a random X-ray-induced DSB by homologous recombination using the sister chromatid as the repair donor. As shown on the right, a site specific DSB is created by zinc finger protein nucleases (Fig. 10 above).

repeat activated by the tat protein effectively repressed expression, and preliminary experiments in a cellular infection assay gave a 3-fold drop in infectivity.

(ii) Disruption of the effective cycle of infection of herpes simplex virus (HSV).³⁸⁾ A ZFP transcription factor designed to repress the promoter of a viral gene that is

normally activated first in the replication cycle produced a 10-fold reduction in the virus titre in an infected cell line (Fig. 8). This is a good result, considering that there are five other “immediate early” genes which contribute to the infection. Several more of these would have to be repressed to reduce the titre further.

(iii) Activating the expression of vascular endothelial growth factor (VEGF-A) in a human cell line³⁹⁾ and in an animal model.⁴⁰⁾ These experiments have led to a therapeutic application (see below).

(iv) Regulation of the level of zinc finger expression by a small molecule.⁴¹⁾ This will be important in controlling the dose and timing for a therapeutic application. An efficient way of achieving this is by fusing the ZFP to the ligand binding domain of a steroid hormone nuclear receptor. In the absence of hormone, the ZFP transcription factor is retained in the cytoplasm, but after ligand binding the ZFP translocates to the cell nucleus in active form.

Stimulation of new vasculature by engineered ZFPs. Following the work at Sangamo on the activation of the transcription factor VEGF by ZFPs in mouse and human cell lines,³⁹⁾ experiments showed that new blood vessels could be formed in a mouse ear.⁴⁰⁾ These did not leak, unlike the results which had been obtained by earlier workers who had been delivering various cDNA spliced isoforms of the gene. The reason for the success with ZFP activation is that the latter acts on the promoter of the VEGF gene, and hence all spliced isoforms are naturally produced when the gene is induced to promote angiogenesis.

Subsequent work by Frank Giordano at Yale and Brian Annex at Duke University showed increased blood flow in hindlimbs of ischaemic rabbits, with the ZFP delivered by a retrovirus or simply by injecting the DNA. On the basis of this, and other confirmatory work, Edwards Life Sciences, a licensee of Sangamo, was last year granted an IND by the FDA to begin clinical trials to evaluate the ability of an appropriate ZFP to stimulate the natural growth of normal blood vessels in treating claudication, a symptom of peripheral arterial obstructive disease that causes poor blood flow in the legs. If these trials are successful a ZFP of this kind could ultimately have possible use in ischemic cardiac disease, if an effective delivery system can be devised.

Gene Correction by homologous recombination using sequence-specific zinc finger nucleases.

Gene Correction is the process by which sequence alterations in defective or deleterious genes can be changed or “corrected” by homologous recombination

(HR)-mediated gene conversion between the target locus and a donor construct encoding the corrective sequence (Fig. 9). Monogenic disorders such as X-linked severe combined immune deficiency (SCID), sickle-cell anemia (SCA), haemophilia, and Gaucher's disease are caused by the inheritance of defective alleles of a single gene. The ability to replace this gene sequence *via* HR-mediated gene correction has the potential of fully restoring the gene function and providing a permanent cure for patients with these disorders. However, this process is highly inefficient in that the frequency of unaided HR at a specific locus occurs in only about 1 in 10^5 cells.⁴²⁾ This is far below a level that would be considered therapeutic. A double-stranded break (DSB) has been demonstrated by Jasin⁴³⁾ to potentiate HR at a specific genetic locus by ~5,000-fold. Therefore, the introduction of a corrective donor sequence together with a site-specific nuclease that would produce a DSB at or near the location of the mutation could stimulate Gene Correction to levels that would provide a therapeutic impact.

Jasin's demonstration was based on artificially introducing into an endogenous gene of a human cell line an 18 bp DNA sequence, which was the specificity binding site for the homing endonuclease *Sce*1 and which had only a small probability of occurring naturally elsewhere in the genome. However, to carry out gene correction in native cells requires the specific targeting of the mutated sequence and a zinc finger peptide fused to a nuclease domain is the natural choice. A nuclease of this type has been developed by Chandrasegaran⁴⁴⁾ using an engineered ZFP fused to the non-specific cleavage domain of the FokI type II restriction enzyme. This type of zinc finger nuclease has been used by Dana Carroll⁴⁵⁾ to produce mutants in *Drosophila* by gene “correction” and by Matthew Porteus⁴⁶⁾ in a GFP model system to study gene modification in a human cell line. Following Chandrasegaran, the three-finger ZFP nucleases are introduced as pairs with tandem binding sites engineered in opposite orientations (Fig. 10), with a 6 bp spacing separating the two half sites. In both cases the efficiency of the targeting is not very high but this was not crucial to these studies. Indeed some of the *Drosophila* mutants are lethal, but they are normally selected on the phenotype. Both studies showed that a ZFN-produced DSB can markedly increase the rate of homologous recombination between a donor DNA construct and a reporter gene in their two different systems.

The aim of recent work at Sangamo Biosciences, led by Michael Holmes, Fyodor Urnov and Philip Gregory, is

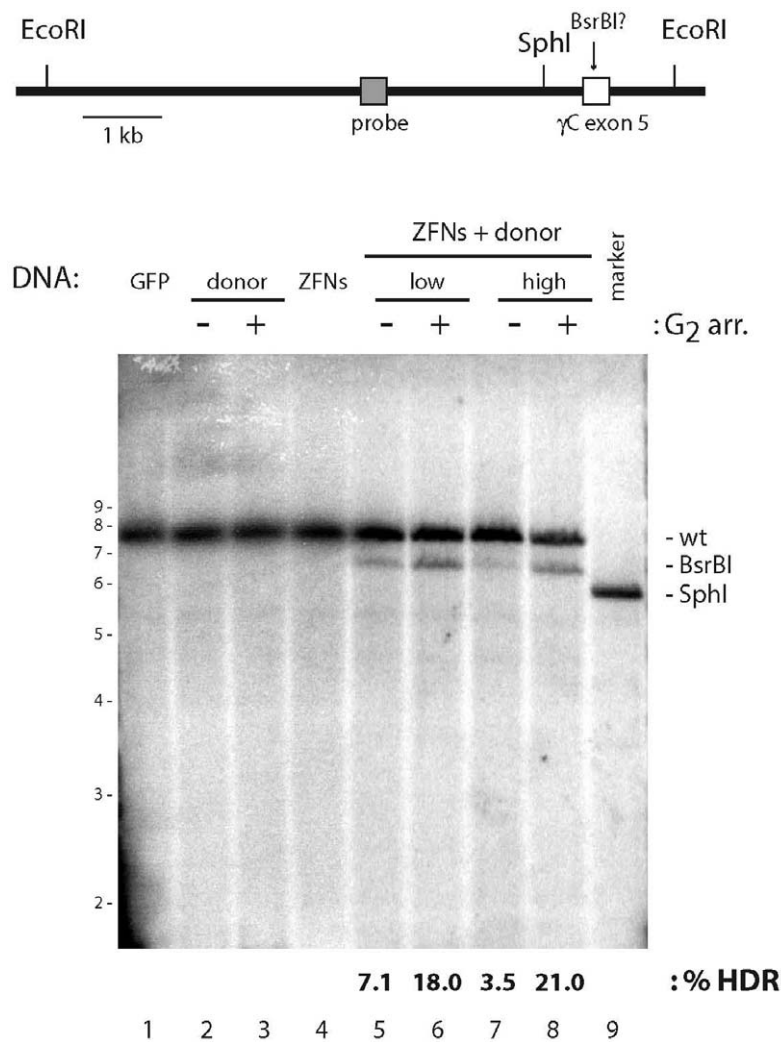


Fig. 11. Results of gene correction for SCID disease by the Sangamo group⁴⁸⁾: high-frequency homologous driven repair at the endogenous IL2R γ locus of leukaemic K562 cells driven by designed ZFNs. Cells were transfected with the indicated plasmids (at two different concentrations indicated by “low” and “high” to ensure that the exogenous DNA load was not limiting). The donor DNA was marked by a BsrBI restriction site (not present in the K562 genome) which was created by replacing a single nucleotide of one codon so as to preserve the amino acid code. After 1 month, genomic DNA was isolated, digested to completion with EcoRI, BsrBI, and DpnI, and the percentage of BsrBI-carrying chromosomes was measured by Southern blotting with the probe indicated. A SphI restriction fragment is also shown as a size marker.

to use zinc finger nucleases to create DSB's at specific sites of mutation in the human genome responsible for monogenic disorders. After experiments on model systems to optimise ZFN-driven gene correction, they set out to determine whether ZFNs could evoke a comparable increase in HR frequency at an endogenous gene. They have focussed their efforts on the IL2R γ gene in which loss-of-function point mutations cause X-linked

severe combined immunodeficiency disease (SCID). In the absence of bone marrow transplantation or gene therapy, this leads to death in early childhood. Treatment for the disease by gene therapy⁴⁷⁾ has been performed by inserting one or more copies of the normal gene in the chromosomes of a number of affected children, that is, by gene addition rather than gene correction. After successful treatment of the majority of the

patients, two of them died some years later of leukaemia, probably because the random insertion of the vector led to an LM02 gene coming under inappropriate control.

The middle exon of IL2R γ contains SCID mutation hotspots. Since maximal homology driven recombination occurs when a DSB is evoked at or close to the mutated site, the Sangamo workers engineered a pair of ZFNs specific for the exact location of the mutation on the X chromosome. Two DNA binding domains were assembled from the zinc finger archive described above, each containing four highly specific zinc-finger motifs, and thus simultaneously recognising two different 12 bp sites, separated by a fixed distance between them. The chance of this particular pattern existing elsewhere in the genome is negligible. Having confirmed *in vitro* that the proteins bind as intended, they next improved them further by single amino acid substitutions in the zinc finger recognition helices which gave a further five-fold increase in potency. The results⁴⁹ show an 18-20% rate of gene correction in the target cells, which was stable after one month in culture (Fig. 11).

This accomplishment is dramatic, with an increase by many orders of magnitude over anything achieved in the past by “gene targeting”, particularly as no selection has been used. Moreover measurements of both the mRNA and protein levels expressed by the corrected IL2R γ gene showed that the mutation had been efficiently and stably corrected. The next step in curing SCID disease would be to isolate CD34 progenitor cells from the patient’s bone marrow, and correct them *ex vivo*. After this gene correction, the cells would be allowed to expand and then re-introduced into the bone marrow to repopulate it with corrected cells. Of course an extensive study of ZFN safety (e.g. no other DSBs created outside the target? immunogenicity?) would first have to be undertaken, but there is every prospect of gene correction eventually becoming a reality for SCID and other monogenic diseases.

Acknowledgements. I thank Dr. Philip Gregory and his colleagues at Sangamo Biosciences for many helpful discussions, for sending me a preprint of their Nature paper (ref. 49) and for permission to use their figures reproduced in Figs. 9, 10 and 11). I would also like to acknowledge my former colleagues, Yen Choo, Mark Isalan and Michael Moore in Cambridge, and later at Gendaq, who helped develop zinc fingers into a robust technology for use in research and therapeutics.

References

- 1) Kornberg, R. D. (1977) Structure of chromatin. *Ann. Rev. Biochem.* **46**, 931-954.
- 2) Klug, A. (1983) From macromolecules to biological assemblies. *In* Les Prix Nobel en 1982. The Nobel Foundation, Stockholm, pp. 93-125.
- 3) Brown, D. D. (1984) The role of stable complexes that repress and activate eukaryotic genes. *Cell* **37**(2), 359-365.
- 4) Picard, B., and Wegnez, M. (1979) Isolation of a 7S particle from *Xenopus laevis* oocytes: a 5S RNA-protein complex. *Proc. Natl. Acad. Sci. USA* **76**(1), 241-245.
- 5) Pelham, H. R. B., and Brown, D. D. (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc. Natl. Acad. Sci. USA* **77**(7), 4170-4174.
- 6) Miller, J., McLachlan, A. D., and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**(6), 1609-1614.
- 7) Hanas, J. S., Hazuda, D. J., Bogenhagen, D. F., Wu, F. Y., and Wu, C. W. (1983) *Xenopus* transcription factor A requires zinc for binding to the 5S RNA gene. *J. Biol. Chem.* **258**(14), 14120-14125.
- 8) Smith, D. R., Jackson, I. J., and Brown, D. D. (1984) Domains of the positive transcription factor specific for the *Xenopus* 5S RNA gene. *Cell* **37**(2), 645-652.
- 9) Ginsberg, A. M., King, B. O., and Roeder, R. G. (1984) *Xenopus* 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. *Cell* **39**, 479-489.
- 10) Brown, R. S., Sander, C., and Argos, P. (1985) The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Lett.* **186**(2), 271-274.
- 11) Diakun, G. P., Fairall, L., and Klug, A. (1986) EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. *Nature* **324**(6098), 698-699.
- 12) Tso, J. Y., van den Berg, D. J., and Korn, L. J. (1986) Structure of the gene for *Xenopus* transcription factor TF IIIA. *Nucl. Acids Res.* **14**(5), 2187-2200.
- 13) Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) The Pfam protein families database. *Nucl. Acids Res.* **30**(1), 276-280.
- 14) Vincent, A., Colot, H. V., and Rosbash, M. (1985) Sequence and structure of the *serendipity* locus of *Drosophila melanogaster*: a densely transcribed region including a blastoderm-specific gene. *J. Mol. Biol.* **186**(1), 149-166.
- 15) Rosenberg, U. B., Schröder, C., Preiss, A., Kienlin, A., Côté, S., Riede, I., and Jäckle, H. (1986) Structural homology of the product of the *Drosophila* Krüppel gene with *Xenopus* transcription factor IIIA. *Nature* **319**, 336-339.
- 16) Berg, J. M. (1988) Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc. Natl. Acad. Sci. USA* **85**(1), 99-102.
- 17) Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., and Wright, P. E. (1989) Three-dimensional solution structure of

- a single zinc finger DNA-binding domain. *Science* **245**(4918), 635-637.
- 18) Nakaseko, Y., Neuhaus, D., Klug, A., and Rhodes, D. (1992) Adjacent zinc-finger motifs in multiple zinc-finger peptides from SWI5 form structurally independent flexibly linked domains. *J. Mol. Biol.* **228**(2), 619-636.
 - 19) Neuhaus, D., Nakaseko, Y., Schwabe, J. W., and Klug, A. (1992) 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 beta-sheet. *J. Mol. Biol.* **228**(2), 637-651.
 - 20) Pavletich, N. P., and Pabo, C. O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**(5007), 809-817.
 - 21) Fairall, L., Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993) The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **366**(6454), 483-487.
 - 22) Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Shakked, Z., and Steitz, T. A. (1979) A hypothesis on a specific sequence-dependent conformation of DNA and its relation to the binding of the lac-repressor protein. *J. Mol. Biol.* **131**(4), 669-680.
 - 23) Rhodes, D., and Klug, A. (1981) Sequence-dependent helical periodicity of DNA. *Nature* **292**, 378-380.
 - 24) Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**(4705), 1315-1317.
 - 25) McCafferty, J., Griffiths, A. D., Winter, G., and Chiswell, D. J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**(6301), 552-554.
 - 26) Choo, Y., and Klug, A. (1994) Towards a code for the interactions of zinc fingers with DNA: selection of randomised fingers displayed on phage. *Proc. Natl. Acad. Sci. USA* **91**(23), 11163-11167.
 - 27) Choo, Y., and Klug, A. (1994) Selection of DNA binding sites for zinc fingers using rationally randomised DNA reveals coded interactions. *Proc. Natl. Acad. Sci. USA* **91**(23), 11168-11172.
 - 28) Choo, Y., Sánchez-García, I., and Klug, A. (1994) *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **372**(6507), 642-645.
 - 29) Isalan, M., Choo, Y., and Klug, A. (1997) Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl. Acad. Sci. USA* **94**(11), 5617-5621.
 - 30) Isalan, M., Klug, A., and Choo, Y. (1998) Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* **37**(35), 12026-12033.
 - 31) Isalan, M., Klug, A., and Choo, Y. (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nature Biotech.* **19**(7), 656-660.
 - 32) Moore, M., Choo, Y., and Klug, A. (2001) Design of polyzinc finger peptides with structured linkers. *Proc. Natl. Acad. Sci. USA* **98**(4), 1432-1436.
 - 33) Moore, M., Klug, A., and Choo, Y. (2001) Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc. Natl. Acad. Sci. USA* **98**(4), 1437-1441.
 - 34) Kim, J. S., and Pabo, C. O. (1998) Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl. Acad. Sci. USA* **95**(6), 2812-2817.
 - 35) Tan, S., Guschin, D., Davalos, A., Lee, Y. L., Snowden, A. W., Jouvenot, Y., Zhang, H. S., Howes, K., McNamara, A. R., Lai, A. *et al.* (2003) Zinc-finger protein-targeted gene regulation: genome wide single-gene specificity. *Proc. Natl. Acad. Sci. USA* **100**(21), 11997-12002.
 - 36) Jamieson, A. C., Miller, J. C., and Pabo, C. O. (2003) Drug discovery with engineered zinc-finger proteins. *Nature Reviews Drug Discovery* **2**(5), 361-368.
 - 37) Reynolds, L., Ullman, C., Moore, M., Isalan, M., West, M. J., Clapham, P., Klug, A., and Choo, Y. (2003) Repression of the HIV-1 5' LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors. *Proc. Natl. Acad. Sci. USA* **100**(4), 1615-1620.
 - 38) Papworth, M., Moore, M., Isalan, M., Minczuk, M., Choo, Y., and Klug, A. (2003) Inhibition of herpes simplex virus 1 gene expression by designer zinc-finger transcription factors. *Proc. Natl. Acad. Sci. USA* **100**(4), 1621-1626.
 - 39) Liu, P. Q., Rebar, E. J., Zhang, L., Liu, Q., Jamieson, A. C., Liang, Y., Qi, H., Li, P. X., Chen, B., Mendel, M. C. *et al.* (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. *Activation of vascular endothelial growth factor A. J. Biol. Chem.* **276**(14), 11323-11334.
 - 40) Rebar, E. J., Huang, Y., Hickey, R., Nath, A. K., Meoli, D., Nath, S., Chen, B., Xu, L., Liang, Y., Jamieson, A. C. *et al.* (2002) Induction of angiogenesis in a mouse model using engineered transcription factors. *Nature Med.* **8**(12), 1427-1432.
 - 41) Beerli, R. R., Schopfer, U., Dreier, B., and Barbas III, C. F. (2000) Chemically regulated zinc finger transcription factors. *J. Biol. Chem.* **275**(42), 32617-32627.
 - 42) Capecchi, M. R. (1989) Altering the genome by homologous recombination. *Science* **244**(4910), 1288-1292.
 - 43) Jasin, M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* **12**(6), 224-228.
 - 44) Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA* **93**(3), 1156-1160.
 - 45) Bibikova, M., Beumer, K., Trautman, J. K., and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**(5620), 764.
 - 46) Porteus, M. M., and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**(5620), 763.
 - 47) Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., and Casanova, J. L. (2000) Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* **288**(5466), 669-672.

- 48) Elrod-Erickson, M., Benson, T. E., and Pabo, C. O. (1998) High resolution structures of variant Zif268-DNA complexes: implications for understanding zinc-finger-DNA-recognition. *Structure* **6**(4), 451-464.
- 49) Urnov, F. D., Miller, J. C., Lee, Y.-L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., and Holmes, M. C. (2005) Highly efficient endogenous human gene correction using designed zinc finger nucleases. *Nature*, in press.

(Received March 9, 2005; accepted March 14, 2005)

Profile

Aaron Klug was born in 1926 in Lithuania, but grew up in Durban, South Africa, from the age of two when his family emigrated there. He was educated at the Universities of Witwatersrand (BSc 1945), Cape Town (MSc 1946) and Cambridge (PhD 1953). He began as a medical student, transferred to science, and his PhD at the Cavendish Laboratory was in Physics. From 1954 until her untimely death in 1958, he worked with Rosalind Franklin at Birkbeck college, London on the structure of tobacco mosaic virus. In 1962, he joined the Medical Research Council Laboratory of Molecular Biology in Cambridge, became Joint Head of the Structural studies Division in 1978 and was the Director of the Laboratory from 1986 to 1996. He now continues as a member of staff, leading a research group on gene expression.

He was President of the Royal Society (1995-2000), is a member of the Order of Merit, a Foreign Associate of the US National Academy of Sciences, the French Academy of Sciences, the Max Planck Gesellschaft, the Japan Academy, and has received many honorary degrees. He is an Honorary Fellow of Peterhouse and of Trinity College, Cambridge.

His work has been on the interactions of proteins with nucleic acids and on the elucidation of the structures of large biological molecules and assemblies, including simple viruses and chromatin, by X-ray diffraction and electron microscopy, and the development of new methods for their study. The principle of his method of 3-D image reconstruction in electron microscopy from a series of 2-D tilted images later formed the basis of the X-ray CT scanner. In 1982 he was awarded the undivided Nobel Prize in Chemistry. The citation reads "for the development of crystallographic electron microscopy and the structural elucidation of protein-nucleic acid complexes of biological importance".

His current research is on the structure of DNA and RNA binding proteins which regulate gene expression and in particular on the interaction with DNA of the zinc finger family of transcription factors which he discovered in 1985. In 1994 he showed that a synthetic zinc finger protein could be used to block the expression of an oncogene in a mouse cell line and also to activate gene expression in another system. More generally, by fusing sequence-specific zinc finger peptides to various effector or functional domains, chimeric proteins can be engineered to target virtually any gene in a complex genome and manipulate or modify it. The US biotech company Sangamo has recently been granted the go ahead for trials on the use of a zinc finger construct as a therapeutic agent to stimulate blood vessel growth in human peripheral vascular disease. Promising preclinical results have been obtained on correcting the mutations in two monogenic diseases.

Aaron Klug has also considerable experience of other branches of biotechnology. He is on the Scientific Advisory Boards of several companies, including Cambridge Antibody Technology which was spun-off from the MRC when he was Director, to produce fully human therapeutic antibodies, some of which are now in clinical use. He also set in train the creation of the Sanger Centre in Cambridge which was responsible for sequencing one third of the human genome.

