Proc. Indian Acad. Sci. (Chem. Sci.), Vol. 106, No. 5, October 1994, pp. 1071–1088.

Synthetic nucleases crafted from L-lysine

DARSHAN RANGANATHAN*[1], RAKESH K MISHRA[2], BHISMA K PATEL[3] and NARENDRA K VAISH[3]

[1]Biomolecular Research Unit, Regional Research Laboratory (CSIR), Trivandrum 695019, India

[2]Centre for Cellular and Molecular Biology, Hyderabad 500007, India

[3]Department of Chemistry, Indian Institute of Technology, Kanpur 208016, India

Abstract. A central strategy for the design of chemical nucleases is presented. This involves the utilization of the α-amino carboxylate unit of L-lysine to form Cu(II) templates to function as the cleaving centre on the one hand and ω-amino group for the attachment of DNA recognition elements on the other, thus giving rise to a duplex recognition termini, harbouring a centrally placed Cu(II) for potentiation of oxidative scission. The recognition element studied encompasses a spectrum of structures ranging from quinazolines and purine residues to specifically crafted peptide segments that have potential to form secondary structures. These could be represented as R–K–Cu–K–R wherein R is the recognition system and K–Cu–K, a composite crafted from lysine, consisting of the cleaving centre from metal complexation of α-amino acid unit and the spacer consisting of the four methylene groups of the side chain. The binding and DNA scission profile of the sixteen chemical nucleases thus prepared and fully characterized have been probed by UV, fluorescence quenching and electrophoretic studies. Their binding to calf thymus DNA is associated with a decrease in ε and an ~ 10–15 nm red shift. The involvement of GC sequence in binding is indicated from studies with poly[d(G–C)·d(G–C)] and poly[d(A–T)·d(A–T)], wherein the hypochromicity and red shift were found to be quite pronounced in the former.

Fluorescence quenching studies with Bz–Trp–Trp–K–Cu–K–Trp–Trp–Bz demonstrated the binding of one ligand at approximately every stretch of 112 bp and approximately a stretch of 80 bp in the presence of salt. The DNA cleaving properties of all the nucleases were demonstrated with pBR 322 and p blue script 11KS using standard protocols. In all cases, covalently closed supercoiled (form I DNA) is converted largely into open circular (form II) suggesting nicking of the single strand at binding sites. Sequence specificity experiments with the nuclease. Bz–Ala–Gly–K–Cu–K–Gly–Ala–Bz in a ^{32}P 3'-end labeled 117 bp restriction fragment (Eco RI/Hind III) of pUC-18 showed almost exclusive attacks at thymidylate residues in particular, thymines corresponding to 5'T of the CTAT(3'–5') box. Whilst the most preferred site of attack is found at T of 3'–ATC–5' at the trinucleotide level, cleavage studies at low concentration have shown that at pentanucleotide level, the lone sequence 3'–GATCT–5' (a part of the inverted repeat –GAGATCTC–) is favoured (fragment 92) over the more frequently occurring 3'–TATCT–5' segment.

Keywords. Synthetic nucleases; DNA recognition elements; DNA scission profile.

1. Introduction

Design of synthetic molecules which can recognize specific sequences of double helical DNA and bring about scission at targetted locations continues to be an attractive goal

*For correspondence

in nucleic acid research (Corey *et al* 1990; Pei and Schultz 1990; Pyle *et al* 1990; Strobel and Dervan 1990; Uhlmann and Peyman 1990; Nagai *et al* 1991; Keck and Lippard 1992; Sitlani *et al* 1992). In order to achieve high sequence specificity (sequence specificity of 12–15 base pairs is needed for mapping the human genome, Dervan 1986), a general approach thus far has been to combine the DNA binding and cleaving capabilities into a single molecule by linking a nucleolytic agent (for recent reviews see Sigman 1990; Sigman and Chen 1990), generally a redox active metal co-ordination complex which can cut DNA by an oxidative attack on the ribose or deoxyribose unit under physiological conditions, at one end of the termini of a known DNA binding drug or a protein or an oligonucleotide (Mack *et al* 1988; Mack and Dervan 1990; Sluka *et al* 1990; Distefano *et al* 1991; Dervan 1992; Koh and Dervan 1992). The work reported in this paper outlines endeavours related to the design and synthesis of a chemical nuclease model patterned on the working of restriction enzymes (most restriction enzymes exhibit a twofold rotational symmetry in their recognition sequences and so recognise palindromic sequences in DNA) Nature's own molecular scissors which are known to recognize specific sequences of DNA and cut at or near that point – which can recognize increasing segments of double stranded DNA and subsequently effect scission at specific sites, under physiological conditions. The ultimate goal here is the crafting of a family of synthetic restriction enzymes with tailored specificity.

Central to the strategy for the design, was the realization that most sequence specific DNA binding proteins, including restriction enzymes, bind to DNA as a dimer. Interestingly, dimerisation is emerging as a unifying property of most of the sequence specific regulatory proteins. Dimerisation in these proteins is effected during the process of binding to DNA, with the help of specific structural elements called the dimerisation elements, the nature of which differ in each case. For example, in Met-J repressor, the two antiparallel β-sheet strands make up the dimerisation element, in GAL-4, the two amphipathic α-helices are packed together to form the dimerisation unit and in leucine zipper, the interdigitation of leucine residues placed at every fourth position helps to form the coiled coil motif. The utility of dimerisation can be understood to reflect the fact that if dimer acts as a single structural unit, the binding should double the DNA contact area thereby square the affinity constant.

Another feature which was considered important to incorporate in the design was the presence of two fold rotational symmetry – a common feature in most restriction enzymes which enables them to recognize palindromic sequences in DNA.

Figure 1 presents the design of a highly versatile chemical nuclease model, wherein, the DNA cleaving agent "C" is positioned in the center of a duplex recognition termini "R" linked through an appropriate spacer "S". This model has similarities with DNA recognition proteins such as GAL–4 (Marmorstein *et al* 1992) wherein the recognition takes place at the termini of a symmetric duplex, assembled by a central dimerisation element linked via spacer to the end units.

It was envisaged that the model would be constructed from three sub systems, namely, the recognition module "R", the spacer "S" and the cleaving centre "C" and much flexibility can be achieved by judicious permutations and combinations of these sub systems which would enable the introduction of structural units with a predictable profile towards DNA, such as groove binding interaction, intercalation and sequence recognition.

Coded amino acid L-lysine, with its α and ω NH_2 groups separated by a four methylene spacer, was considered an ideal choice and appeared tailor-made for

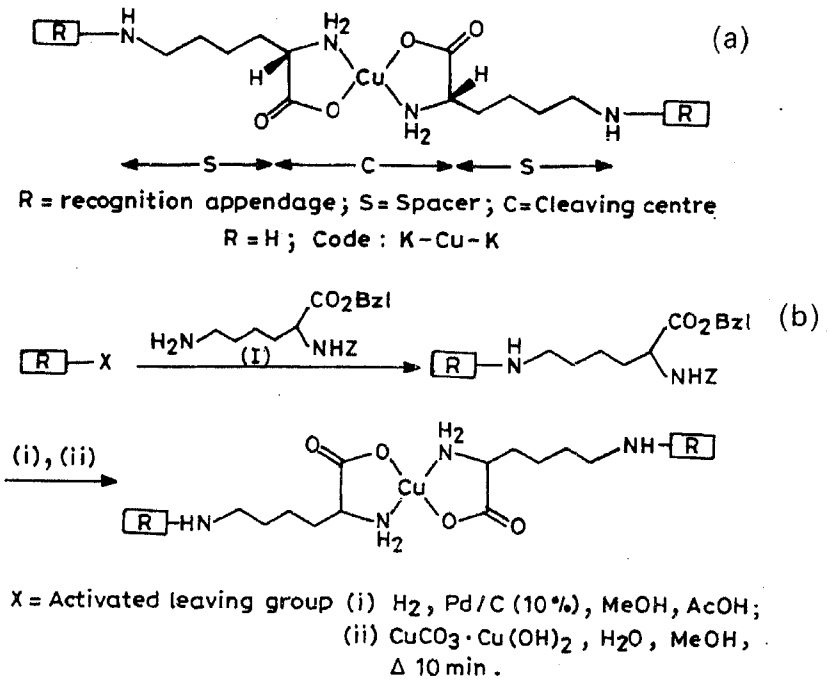


Figure 1. (a) Design of chemical nuclease model crafted from L-lysine. (b) General strategy for construction of model.

crafting such a design (figure 1). The strategy envisaged the utilization of the α-amino carboxylate unit of L-lysine to form the desired metal template acting as the cleaving centre on the one hand and the ω-amino group for the attachment of DNA recognition elements on the other, thus giving rise to a duplex recognition termini harbouring a centrally placed Cu(II) for potentiation of oxidative scission.

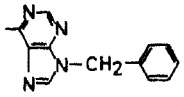
One of the main objectives of the present work was to achieve targetted scission of DNA at pre-determined sites by controlling carrier ligands – recognition modules to which the cleaving centre is attached – with the ultimate goal of creating a family of synthetic restriction enzymes having tailored specificity.

The flexibility of the design here permits the attachment of a variety of recognition systems ranging from planar, aromatic quinazolines and purines having an intercalative profile, to the whole spectrum of coded amino acids assembled as small peptide units with potential for forming secondary structures and possibly with a predictable profile.

The "recognition systems" used in the present work are so chosen to create diverse environments with an aim to assess the role of these elements in DNA sequence read out. The common strategy (figure 1) used for assembling the nuclease model, envisaged the coupling of individually constructed recognition appendages "R" with the ω-NH_2 of N^{α}, C protected L-lysine to afford the spacer-linked recognition systems, which, via sequence involving deprotection and treatment with basic copper carbonate were converted into the dimeric units, with Cu(II) acting as the dimerisation element.

Results and discussion.

For crafting of models containing quinazolines and purines as the recognition units, 4-chloroquinazoline (Armarego 1961) was condensed with N^{α}Z–Lys–OMe·HCl (Costopanagiotis *et al* 1968) in CH_2Cl_2 in the presence of Et_3N to give quinazolyl

Entry	R	Code [a]
1.		Pur-K-Cu-K-Pur
2.		Qu-K-Cu-K-Qu
3.	$H_2N-CH(CH_2OH)-CO$	S-K-Cu-K-S
4.	$Boc-NH-CH(CH_2CONH_2)-CO$	N-K-Cu-K-N
5.	$H_2N-CH(CH_2CH_2CONH_2)-CO$	Q-K-Cu-K-Q
6.	$Boc-NH-CH(CH_2-C_6H_4(OH)-CO$	Y-K-Cu-K-Y
7.	$Bz-NH-CH(CH_2-Indolyl)-CO$	W-K-Cu-K-W
8.	$Bz-NH-CH(CH_2CH_2CH_2NHCONH_2)-CO$	Cit-K-Cu-K-Cit
9.	$Boc-NH-CH(CH_2CH_2CH_2-NH-C(=NH)-NHNO_2)-CO$	R-K-Cu-K-R
10.	$Bz-NH-CH(CH_3)-CONH-CH_2-CO$	A-G-K-Cu-K-G-A
11.	$Bz-NH-CH(CH_2-Indolyl)-CONH-CH(CH_2-Indolyl)-CO$	W-W-K-Cu-K-W-W
12.	$Bz-NH-CH(CH_3)-CONH-CH_2-CONH-CH(CH_2-Indolyl)-CO$	A-G-W-K-Cu-K-W-G-A
13.	$Bz-NH-CH_2-CONH-CH_2-CONH-CH(CH_2-Imidazolyl)-CO$	G-G-H-K-Cu-K-H-G-G
14.	$H_2N-CH(CH_2CH(CH_3)_2)-CONH-CH(CH_2OH)CONH-CH(CH_2OH)-CO$	L-S-S-K-Cu-K-S-S-L
15.	$Bz-NH-CH_2-CONH-CH_2-CONH-CH_2-CONH-CH_2-CO$	G-G-G-G-K-Cu-K-G-G-G-G
16.	$Bz-NH-CH(CH_3)-CONH-CH_2-CONH-CH(CH_3)-CONH-CH_2-CO$	A-G-A-G-K-Cu-K-G-A-G-A

[a] one letter code has been used for amino acids, K: lysine; S: Serine; N: asparagine; Q: glutamine; Y: tyrosine; W: tryptophan; Cit: citrulline; R: arginine; A: alanine; G: glycine; Pur: purine; Qu: quinzoline.

Figure 2. Recognition ligands (R) used in the nuclease model.

(4-N^{ω})–N^{α}Z–Lys–OMe in 63% yields. Z-Deprotection (H_2, Pd/C (10%)) to the free amino ester quinazolyl (4-N^{ω})–Lys–OMe (99%) followed by hydrolysis with 2N NaOH and treatment with basic cupric carbonate ($CuCO_3 \cdot Cu(OH)_2$) in aqueous MeOH afforded tiny blue crystals of N^{ω}-quinazolyl lysine copper complex (2, figure 2) in 95% yields.

In an alternate method compound (2) was conveniently prepared in > 90% yields by direct condensation of 4-chloroquinazoline with lysine copper complex in aqueous acetone in the presence of bicarbonate.

For the preparation of chemical nuclease with purine as duplex recognition termini, N^{α}Z–Lys–OH was condensed with 9-benzyl 6-chloropurine (Montgomery and Temple 1961) in dry DMSO in the presence of anhyd. K_2CO_3 to yield 9-benzyl purinyl(6-N^{ω})–Lys(N^{α}Z)–OH (85%). Z-Deprotection to 9-benzyl purinyl(6-N^{ω})–Lys–OH (91%) followed by brief heating with aqueous basic copper carbonate (0·5 equiv.) afforded N^{ω}–(9-benzyl purinyl)–lysine copper complex (1) as a light blue micro crystalline solid (63%).

The DNA binding capabilities of (1) and (2) were tested by UV absorption spectroscopic studies, using calf thymus DNA. A decrease in the extinction coefficient combined with the shift of the wavelength of maximum absorbance to the lower

energy range (~ 15 nm shift) indicated significant binding of both compounds to DNA double helix. In binding measurements with poly [d(A–T)·d(A–T)] and poly [d(G–C)·d(G–C)], both (1) and (2) exhibited relatively more affinity for poly [d(G–C)·d(G–C)]. In the UV titration of (1) with poly [d(G–C)·d(G–C)] and poly [d(A-T)·d(A-T)], the following profile was observed at saturation point.

DNA *duplexes*	*%Decrease in absorbance*	*Red shift* (nm)
Poly [d(G–C)·d(G–C)]	22	10
Poly [d(A–T)·d(A–T)]	08	02

The two duplexes used in UV studies were composed of the following alternating sequences.

-----ATATATATAT-----
-----TATATATATA-----
and
-----CGCGCGCGCG-----
-----GCGCGCGCGC----- respectively

OD measurements further showed that at saturation point the DNA/ligand molar ratio is ~ 150 in terms of phosphate, which would mean that at the intervals of every 75 base pair there is at least one ligand binding to double helical DNA. The compounds (1) and (2) do not intercalate as shown by plasmid relaxation assay. Molecular modelling in the case of (2) using DTMM (Desk Top Molecular Modelling) shows a helical nature of the molecule, which is a pointer to the possibility that these compounds may be groove binders.

The DNA cleaving properties of (1) and (2) were tested with pBR-322 supercoiled plasmid. Upon incubation in the presence of 3-mercaptopropionic acid (MPA) as exemplified in figure 3, covalently closed supercoiled (form I) DNA is converted largely into open circular (form II) suggesting nicking of the single strand by both 1 and 2 around the binding sites. Although it has been shown that cutting of DNA is associated with binding of 1 and 2 (*vide supra*), the scission may be occurring at a single binding site or at one of the several different available sites in DNA. These possibilities can be distinguished by experiments with high sequencing gels using ^{32}P-end labelled DNA. The efficiency of cleavage was found to be higher in the case of (2) probably arising from a better groove binding interaction of the quinazoline ring system. Appearance of linear form (form III) DNA upon treatment with higher concentrations of cleaving agents indicate a double strand cleavage.

Having demonstrated the feasibility of the design with the DNA binding and cleaving capabilities of C_2 symmetric templates 1 and 2, it was considered logical to use recognition appendages which would show predictable profile in DNA interactions. These may, for example, comprise selected sequences of amino acids, involved in DNA sequence recognition and form a part of well identified DNA recognition motifs, such as helix-turn-helix (Pabo and Sauer 1984; Gehring *et al* 1990; Harrison and Aggarwal 1990; Kissinger *et al* 1990) (HTH), leucine zipper (Landshultz *et al* 1988; O'Shea *et al* 1989; Vinson *et al* 1989; Oakly and Dervan 1990; O'Neil *et al* 1990; Talanian *et al* 1990), zinc fingers (Klug and Rhodes 1987; Pavletich and Pabo 1991; Vallee *et al* 1991; Baleja *et al* 1992; Kraulis *et al* 1992; Marmorstein *et al* 1992) and helix-loop-helix (Lasser *et al* 1989; Murre *et al* 1989; Prendergest *et al* 1989; Davis *et al* 1990; Sun and Baltimore 1991; Anthony-Cahill *et al* 1992) (HLH) in regulatory proteins. A common feature of these sequence specific proteins is that, most bind to

DNA as dimers, and in the native dimeric proteins, the chemical structure of the dimerisation motif determines the geometry of each monomer subunit. This constrained positioning of the DNA binding regions would facilitate the direct interaction between amino acid residues of the protein and nucleic acid base pairs.

Figure 2 presents a profile of chemical nucleases (entry 3 → 16), which have been assembled by attachment of the appropriate recognition termini comprising of small peptide units especially chosen to create diverse environments and constructed from almost the whole spectrum of coded amino acids. One of the long range objectives for the choice of amino acids and peptide units as recognition appendages is to explore the possibility of any correlation that may exist between the amino acid side chains and peptide structures with specific DNA sequences.

The general strategy for the construction of these nucleases envisaged the condensation of respective amino acid or peptide unit – comprising the recognition element and synthesized by sequential step wise condensation of constituent amino acids using conventional solution phase coupling procedures – with N^{α} and C terminal protected lysine. The recognition system thus assembled was then dimerized with Cu(II) as the central metal template to afford the nucleases 3–16 (figure 4).

Thus, Ser containing recognition system was constructed via the sequence condensation of Z–Ser–OH with N^{α}Z–Lys–OBzl.OTs$^-$ (I, figure 1) in the presence of DCC and HOBt to give Z–Ser–Lys(N^{α}Z)–OBzl in 57% yields. Deprotection (H_2, Pd/C (10%)) afforded Z–Ser–Lys(ω-pep) in quantitative yields which on brief treatment with basic cupric carbonate afforded template $(Ser–Lys)_2Cu[(SK)_2Cu]$ (3, figure 2) in 71% yields.

In a similar fashion, potential nucleases harboring recognition systems containing polar amino acids Asn and Gln were constructed.

Thus, Boc–Asn–OH was condensed with N^{α}Z–Lys–OBzl.OTs (I) in the presence of DPPA to afford Boc–Asn–Lys(N^{α}Z)–OBzl (46%); N, C deprotected (H_2, Pd/C) to Boc–Asn–Lys–OH(ω-pep) (~ 100%) and treated with $CuCO_3$. $Cu(OH)_2$ in aqueous MeOH to give $(Boc–Asn–Lys)_2Cu$ $[(NK)_2Cu]$ (4) in 80% yields.

The Gln containing design (5) was prepared from Z–Gln–OH via, condensation with (I) using DPPA to Z–Gln–Lys(N^{α}Z)–OBzl (41%); N, C deprotection (H_2, Pd/C) to Gln–Lys–OH(ω-pep) (~ 100%) and complexation with $CuCO_3 \cdot Cu(OH)_2$ in 79% yields.

Aromatic amino acids Tyr and Trp were incorporated as the recognition systems leading to the templates (6) and (7) respectively.

Template $(Boc–Tyr–Lys)_2Cu$ $[(YK)_2Cu]$ (6) was prepared by condensation of Boc–Tyr–hydrazide with (I) to Boc–Tyr–Lys(N^{α}Z)–OBzl (66%); N, C deprotection (H_2, Pd/C) to Boc–Tyr–Lys–OH(ω-pep) (100%) and complexation with $CuCO_3 \cdot Cu(OH)_2$ (65%).

Template $(Bz–Trp–Lys)_2Cu$ $[(WK)_2Cu]$ (7) was obtained by condensation of Boc–Trp–OH with (I) (DCC-HOBt) to Boc–Trp–Lys(N^{α}Z)–OBzl (55%); N, C deprotection (H_2, Pd/C) to Bz–Trp–Lys–OH(ω-pep) (~ 100%) and complexation with $CuCO_3 \cdot Cu(OH)_2$ (63%).

The incorporation of citrulline – the non-coded amino acid containing the interesting ureido moiety was considered attractive because of potential hydrogen bonding possibilities with DNA. The citrulline containing system, $(Bz–Cit–Lys)_2Cu$ $[(CitK)_2Cu$, 8] was prepared from Bz–Cit–OH, via the sequence, condensation with (I) to Bz–Cit–Lys(N^{α}Z)–OBzl (67%); N, C deprotection (H_2, Pd/C) to Bz–Cit–Lys–OH(ω-pep) in nearly quantitative yields and complexation with $CuCO_3 \cdot Cu(OH)_2$ (83%).

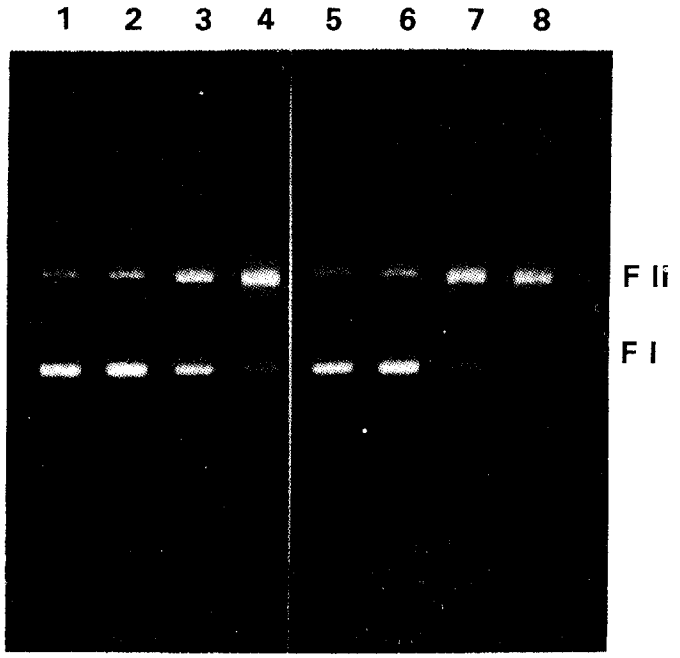


Figure 3. Cleavage of supercoiled pBR 322 DNA with $(Pur–Lys)_2Cu$ (1) and $(Qu–Lys)_2Cu$ (2). pBR 322 DNA (200 ng, form I) in a 10 mM Tris-HCl, *p*H 7·5, was incubated (total volume 10 μL) with the cleaving agent in the presence of 3-mercaptopropionic acid (MPA, 2 mM). After 10 min of incubation at 25°C, the reaction was stopped by the addition of 20 μM neocuprin and samples were analysed on 1% agarose gel. The electrophoresis was carried out at 5volt/cm in TBE (89 mM Tris-borate, *p*H 8·3, 2mM EDTA). Lanes 1 and 5: DNA alone, Lanes 2, 3 and 4: DNA + 0·37, 0·56 and 0·74 mM of (1). Lanes 6, 7 and 8: DNA + 0·17, 0·33 and 0·50 mM of (2) respectively.

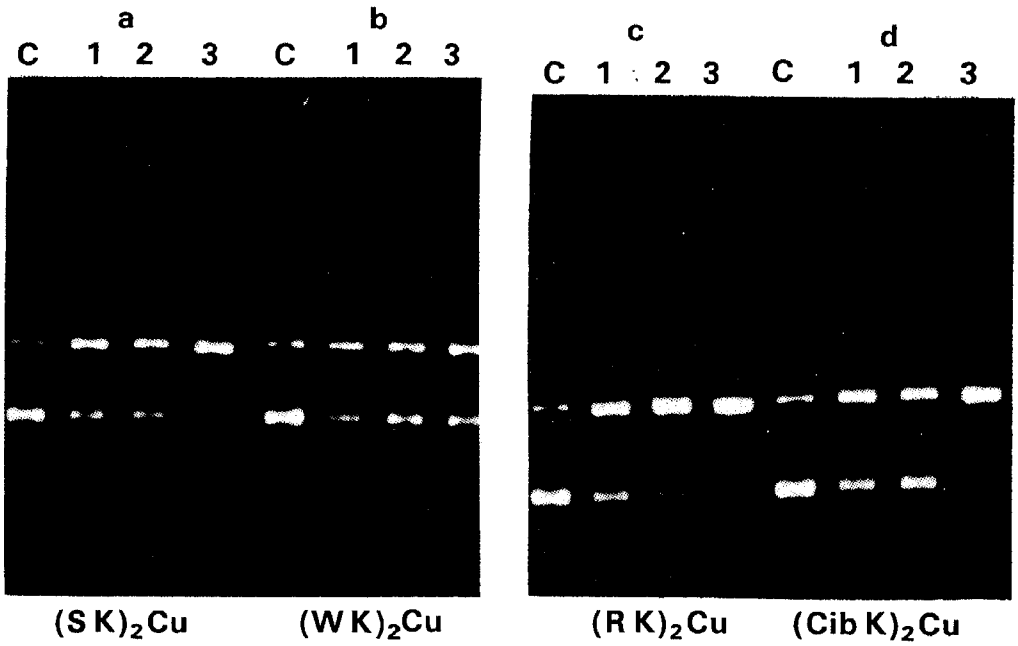


Figure 4A. Concentration dependent DNA (p Blue script 11KS plasmid cleavage by dipeptide copper complexes $(SK)_2Cu$ (3), $(WK)_2Cu$ (7), $(RK)_2Cu$ (9) and $(CitK)_2Cu$ (8). Conditions of cleavage are the same as in figure 3. *Panel a*: Lane C: DNA alone; Lanes 1–3: DNA + 133 μM, 266 μM and 533 μM of (3) respectively. *Panel b*: Lane C: DNA alone; Lanes 1–3: DNA + 100 μM, 200 μM and 300 μM of (7) respectively. *Panel c*: Lane C: DNA alone; Lanes 1–3: DNA + 133 μM, 266 μM and 533 μM of (9) respectively. *Panel d*: Lane C: DNA alone; Lanes 1–3: DNA + 133 μM, 266 μM and 533 μM of (8) respectively.

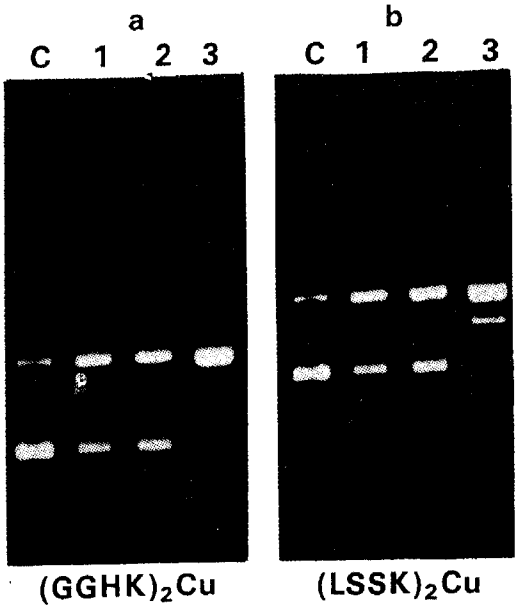


Figure 4B. Concentration dependent DNA (p Blue script 11KS) cleavage by tetrapeptide copper complexes $(GGHK)_2Cu$ (13) and $(LSSK)_2Cu$ (14), panels a and b respectively. *Panel a*: Lane C: DNA alone; Lanes 1–3: DNA + 100 μM, 200 μM and 300 μM of $(GGHK)_2Cu$ (13) respectively. *Panel b*: Lane C: DNA alone; Lanes 1–3: DNA + 133 μM, 266 μM and 533 μM of $(LSSK)_2Cu$ (14) respectively.

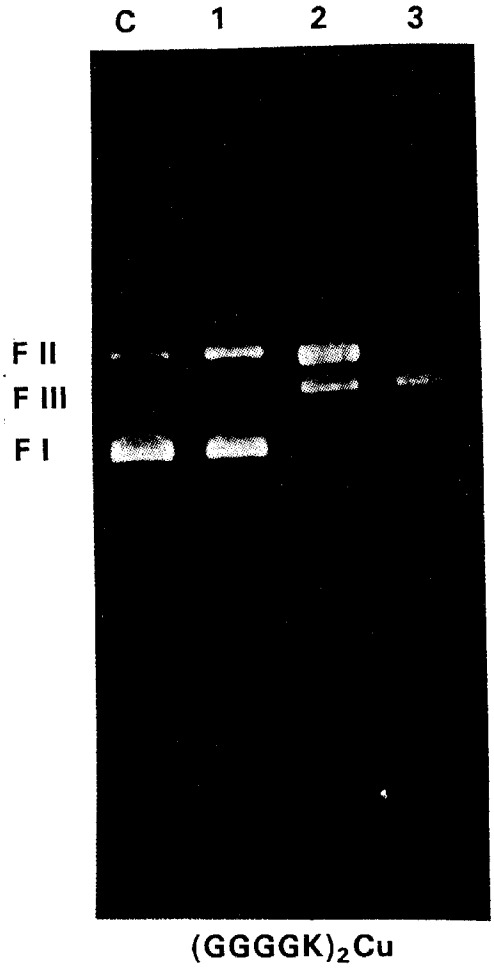


Figure 4C. Concentration dependent cleavage of supercoiled plasmid DNA, p Blue script 11KS, by $(GGGGK)_2Cu$ (15). Lane C: DNA alone; Lanes 1–3: DNA + 25 μM, 50 μM and 100 μM of (15) respectively.

In the sequence-specific interaction with several proteins with DNA, the coded amino acid Arg makes direct hydrogen bonding contacts with code base guanine(G) in the major groove. It was therefore considered interesting to incorporate Arg as the recognition system and study its DNA interaction profile. For the preparation of recognition unit containing this amino acid, namely (Boc–Arg(N^G-NO_2)–Lys)$_2$Cu [(RK)$_2$Cu, 9], Boc–Arg(N^G–NO_2)–OH was condensed with (I) (DCC/HOBt) to Boc–Arg(N^G–NO_2)–Lys(N^αZ)–OBzl (44%); N, C deprotection (H_2, Pd/C) to Boc–Arg(N^G-NO_2)–Lys–OH(ω-pep) ($\sim$100%) and complexed with $CuCO_3 \cdot Cu(OH)_2$ (80%).

The above nucleases (represented by entries (3–9, figure 2) containing dipeptide as recognition systems constructed from the wide spectrum of polar, hydrophobic and basic amino acids, were probed for their DNA scission profile. Figure 4A presents cleavage of supercoiled plasmid DNA, p-blue script 11KS by (SK)$_2$Cu (3), (WK)$_2$Cu (7), (RK)$_2$Cu (9) and (CitK)$_2$Cu (8). Upon incubation in the presence of 3-mercapto propionic acid as exemplified in figure 4A, covalently closed supercoiled (form I) DNA is converted largely into open circular (form II) suggesting nicking of the single strand at binding sites. Appearance of linear DNA (form III) with higher concentration of the agent as demonstrated in the case of (9) shows nicking of DNA at both the strands. No cleavage was observed in control experiments with Cu(II) alone in the presence of reductant (MPA).

Endeavours to incorporate larger peptide units are described below. These peptides, constructed from elements which may themselves show DNA interaction, for example, Trp by intercalation present possibilities for secondary structures, thereby enhancing DNA recognition profile. The synthesis of tri, tetra and pentapeptide recognition systems and their DNA binding and cleaving properties have been probed.

The tripeptide Bz–Ala–Gly–Lys as the recognition system was considered attractive because of the possibility of its hydrophobic interaction utilizing the backbone elements.

Bz–Ala–Gly–OMe – prepared in 73% yields by the condensation of Bz–Ala–OH and Gly–OMe using DCC-HOBt procedure – was converted, *in situ* to the azide and coupled with N^αZ–Lys–OMe. The product Bz–Ala–Gly–Lys(N^αZ)–OMe, obtained in 74% yields, was converted into the metal template (Bz–Ala–Gly–Lys)$_2$Cu [(AGK)$_2$Cu] (10) (75% yield) via sequence involving ester hydrolysis, N^αZ removal (H_2, Pd/C) to Bz–Ala–Gly–Lys–OH(ω–pep) (98%) and brief heating with $CuCO_3 \cdot Cu(OH)_2$.

Incorporation of Trp in recognition systems was considered important, not only because it has the possibility for intercalation but also because this unit can monitor DNA binding by fluorescence quenching studies.

Construction of recognition systems containing two Trp residues adjacent to each other was achieved by condensing Bz–Trp–Trp–OMe in the form of azide, with (I) to yield the tripeptide Bz–Trp–Trp–Lys(N^αZ)–OBzl (38% yield). N, C deprotection (H_2, Pd/C) to Bz–Trp–Trp–Lys(N^αZ)–OH(ω-pep) (99% yield) and complexation with $CuCO_3 \cdot Cu(OH)_2$ yielded the final template (Bz–Trp–Trp–Lys)$_2$Cu [(WWK)$_2$Cu)] (11) in 79% yields.

Synthesis of recognition systems incorporating tetrapeptide unit are described below.

For the preparation of (Ala–Gly–Trp–Lys)$_2$Cu [(AGWK)$_2$Cu] (12) which is expected to show DNA interaction via hydrophobic binding, backbone interaction

or intercalation, Bz–Ala–Gly–azide, generated *in situ* was coupled with Trp–OMe affording the tripeptide Bz–Ala–Gly–Trp–OMe in 62% yields. Repetition of the above operations involving azide generation and coupling with (I) gave tetrapeptide Bz–Ala–Gly–Trp–Lys(N$^{\alpha}$Z)–OBzl in 74% yields. Further operation involving N, C deprotection (H_2, Pd/C) to Bz–Ala–Gly–Trp–Lys–OH(ω-pep) in quantitative yields and copper template formation ($CuCO_3 \cdot Cu(OH)_2$) afforded the template[$(AGWK)_2Cu$] (12) in 90% yields.

The construct $(Gly–Gly–His–Lys)_2Cu$ [$(GGHK)_2Cu$] (13) containing the recognition element, GGH, which represents the Cu(II) binding region of Hin recombinase, a remarkable enzyme, that effects DNA scission at two selected sites, inverts the information and re-ligates to generate an "anti sense DNA" domain (Mack *et al* 1988).

Construction of recognition system containing this interesting unit was accomplished from Bz–Gly–Gly–OMe prepared by coupling Bz–Gly–OH and Gly–OMe by DCC–HOBt procedure in 68% yields – which was condensed, via azide, with His–OMe to produce Bz–Gly–Gly–His–OMe (39%). Further condensation of this tripeptide via azide, with (I) afforded the ω-linked peptide Bz–Gly–Gly–His–Lys (N$^{\alpha}$Z)–OBzl (31%). N, C deprotection with (H_2, Pd/C) to Bz–Gly–Gly–His–Lys–OH (ω-pep) ($\sim$ 100%) followed by complexation with $CuCO_3 \cdot Cu(OH)_2$ gave the metal template (13) in 72% yields.

The tetrapeptide unit Leu–Ser–Ser–Lys has the possibility of DNA interaction via hydrophobic (Leu) or polar (Ser) units. The construction of recognition system containing this unit is described below.

Thus, Z–Leu–OH on condensation with Ser–OMe using DCC-HOBt procedure afforded Z–Leu–Ser–OMe in 78% yields. Azide coupling of this dipeptide with Ser–OMe led to Z–Leu–Ser–Ser–OMe (57%) whose azide, when coupled with (I) gave the interesting tetrapeptide Z–Leu–Ser–Ser–Lys(N$^{\alpha}$Z)–OBzl in 28% yields. N, C deprotection (H_2, Pd/C) to Leu–Ser–Ser–Lys(ω-pep) in quantitative yields, followed by complexation with $CuCO_3 \cdot Cu(OH)_2$ gave the final template $(Leu–Ser–Ser–Lys)_2Cu$ [$(LSSK)_2Cu$] (14) in 56% yields.

The pentapeptide units represented by Gly–Gly–Gly–Gly–Lys (15) and Ala–Gly–Ala–Gly–Lys (16) are anticipated to interact with DNA via their backbone elements. The construction of recognition systems incorporating these units was achieved as follows.

Thus, Bz–Gly–Gly–OH was condensed with Gly–Gly–OMe – obtained quantitatively by Z-deprotection (H_2, Pd/C) of Z–Gly–Gly–OMe – to give the tetrapeptide Bz–Gly–Gly–Gly–Gly–OMe (59%), which via azide condensation with (I) to Bz–Gly–Gly–Gly–Gly–Lys(N$^{\alpha}$Z)–OBzl (36%) was converted to the final recognition template model $(Bz–Gly–Gly–Gly–Gly–Lys)_2Cu$ [$(GGGGK)_2Cu$] (15) in 70% yields, via sequence N, C deprotection to Bz–Gly–Gly–Gly–Gly–Lys(ω-pep) and complexation with $CuCO_3 \cdot Cu(OH)_2$.

Condensation of Bz–Ala–Gly–OH, by DCC-HOBt procedure, with Ala–Gly–OMe – obtained quantitatively by Z deprotection of Z–Ala–Gly–OMe – afforded the tetrapeptide Bz–Ala–Gly–Ala–Gly–OMe (46%). Further condensation, via azide, with (I) to Bz–Ala–Gly–Ala–Gly–Lys(N$^{\alpha}$Z)–OBzl (43%); N, C deprotection (H_2, Pd/C, 99% yield) to Bz–Ala–Gly–Ala–Gly–Lys(ω-pep) followed by complexation with $CuCO_3 \cdot Cu(OH)_2$ yielded the final recognition metal template $(Bz–Ala–Gly–Ala–Gly–Lys)_2Cu$ [$(AGAGK)_2Cu$] (16) in 66% yields.

EPR spectra of complexes 3–16 (figure 2) in $EtOH/H_2O$ at room temperature,

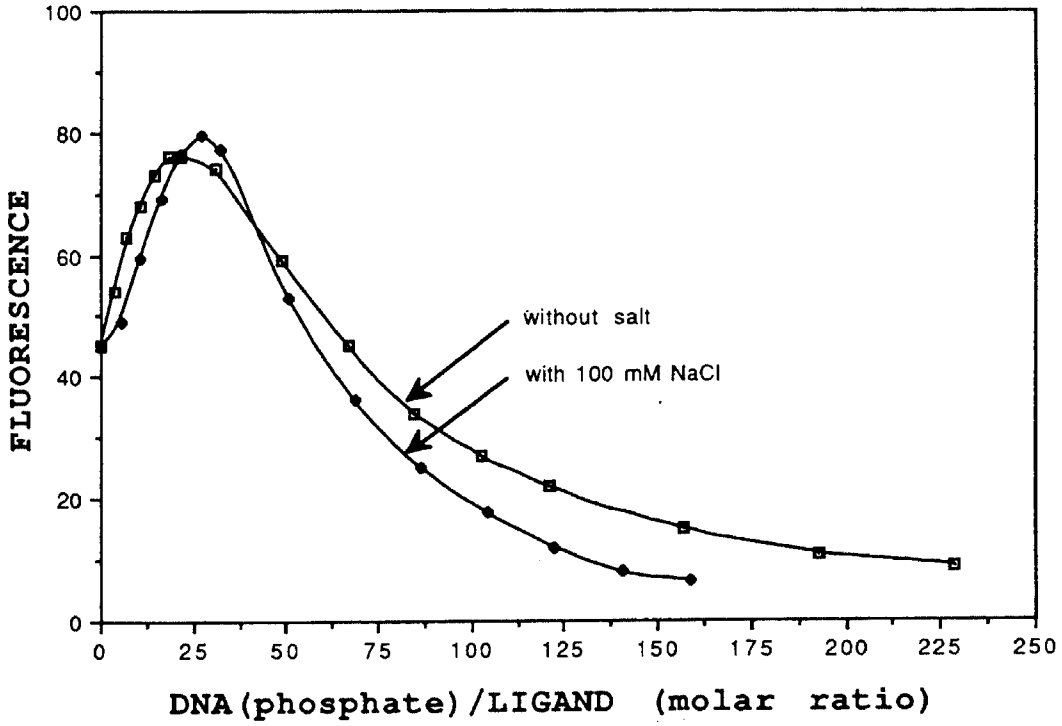


Figure 5. Fluorescence emission spectra (excitation, 280 nm; emission 350 nm) of $(WWK)_2Cu$ (11) with increasing salmon sperm DNA/ligand ratio. 5 μM solution of WWK in water without NaCl or with 0·1 M NaCl in a final volume of 10 μL tris 1 M EDTA, *p*H 7·5 was used for measurements.

exhibited a characteristic Cu(II) four-line pattern which changed to an axial profile with rhombic distortion at liquid nitrogen temperature; $A_{\|}$ 175–180, g_1 2·187–2·296, g_2 2·0023–2·094 and g_3 1·960–2·055.

The UV–Vis spectrum showed, for all compounds, one or more bands in the range 500–800 nm, expected for four coordinated Cu(II) complexes.

The DNA binding and scission profile of the designs constructed from larger peptide units (entries 10–16, figure 2) has been probed by UV, fluorescence quenching and gel electrophoretic studies.

Fluorescence emission studies (figure 5) of $(WWK)_2Cu$ (11) with increasing amounts of salmon sperm DNA demonstrated that at low DNA/ligand (< 20) ratio, $(WWK)_2Cu$ shows enhancement in fluorescence, which changes into a quenching mode with the increasing DNA/ligand ratio, reaching saturation point at ~ 225 (1 ligand bound per ~ 112 bp). This ratio falls to ~ 160 (1 ligand per ~ 80 bp) in the presence of 0·1 mol NaCl.

The concentration dependence of the DNA cleavage in the case of (Bz–Ala–Gly–Trp–Lys)$_2$Cu [$(AGWK)_2Cu$] (12) is presented in figure 6A. The optimum concentration of 200 μM produces the maximum amount of open circular form (form II). Use of higher concentration (400 μM) leads to the cleavage of both the strands as shown by the appearance of linear form (form III).

Time dependent cleavage of (Bz–Ala–Gly–Trp–Lys)$_2$Cu [$(AGWK)_2Cu$] (12) is shown in figure 6B. As could be seen in this figure, a substantial portion of the DNA is cleaved within 20 minutes.

Figure 4B presents cleavage profiles of the tetrapeptide copper complexes, (Bz–Gly–Gly–His–Lys)$_2$Cu [$(GGHK)_2Cu$] (13), and (Leu–Ser–Ser–Lys)$_2$Cu [$(LSSK)_2Cu$]

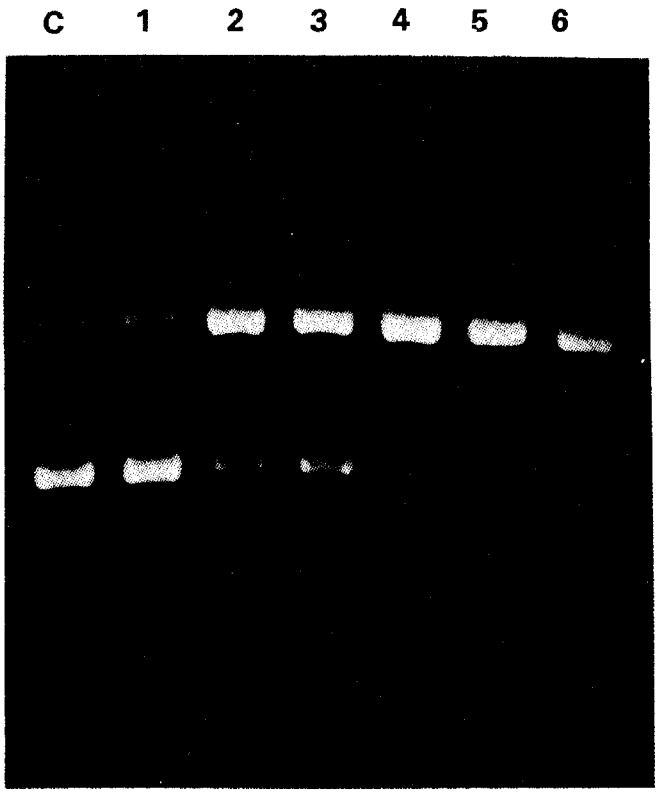


Figure 6A. Concentration dependent DNA (supercoiled plasmid p Blue script 11KS) by $(AGWK)_2Cu$ (12). Experimental conditions are same as in figures 3 and 4. Lane C: DNA alone; Lanes 1–6: DNA in the absence of cleaving agent, DNA + 50 μM, 100 μM, 200 μM, 333 μM and 400 μM of (12) respectively.

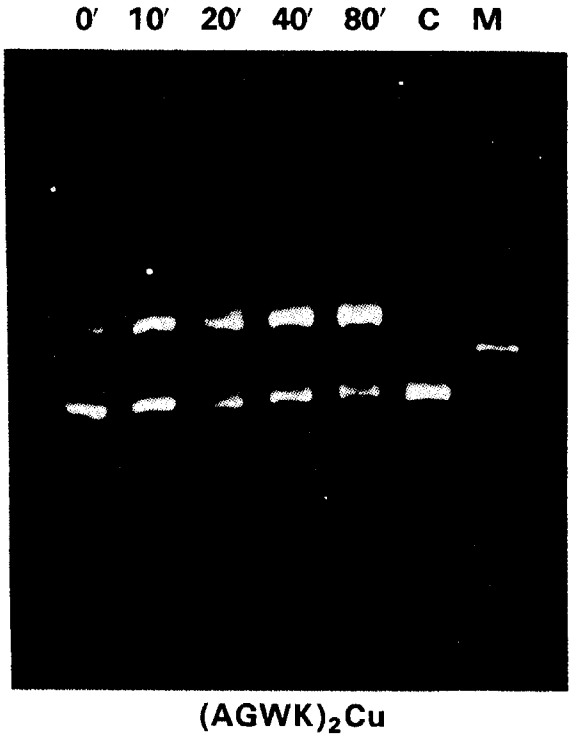


Figure 6B. Time dependent cleavage of supercoiled plasmid, p Blue script 11KS, by $(AGWK)_2Cu$ (12). Under the same reaction conditions as in previous experiments, the reaction was arrested at various time intervals with 100 μM neocuprin and samples analysed on 1% agarose gel. Lane M: Eco RI digested, Lane C: DNA alone, Lanes 0'–80': Reaction arrested at 0', 10', 20', 40', and 80' respectively.

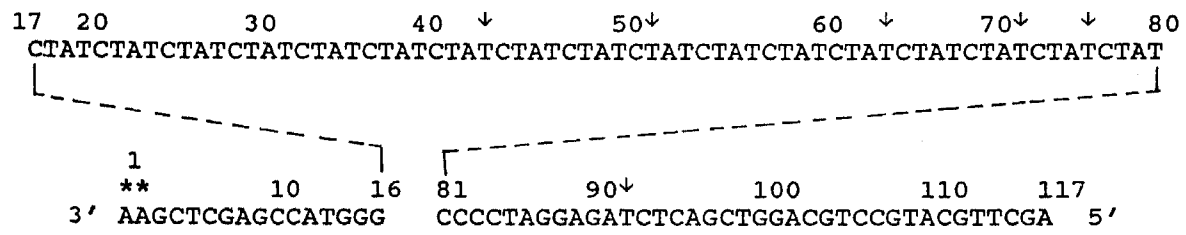


Figure 7. $(AGK)_2Cu$ (10) induced cleavage sites on a ^{32}P 3'-end labelled 117 bp restriction fragment (Eco RI/Hind III) of pUC 18 containing sixteen 3'–CTAT–5' sites. The arrows indicate the preferred sites of attack by the cleaving agent. Approximate size of the fragments resulting from major breaks (as compared to the DNA molecular markers and with undigested DNA) are, from 3' to 5': 44, 52, 64, 72, 76 and 92 corresponding to the cleavage at 5' of the CTAT (3'-5') box. The most preferred site at the trinucleotide level is 3'–ATC–5'. Sequences 3'–GATCT–5' and 3'–TATCT are favoured at the penta nucleotide level.

(14). As can be seen from the figure, appearance of considerable amount of linear form with (14) demonstrates the cleavage of both the strands of DNA.

Figure 4C shows the concentration dependent cleavage by pentapeptide complex $(GGGGK)_2Cu$ (15). Full conversion of supercoiled form into linear and open circular forms is clearly seen in lane 2 with 50 μM concentration of (15). With still higher concentration (100 μM) only the linear form of DNA is visible, as shown in lane 3.

Having established the DNA binding as well as the scission profile of the design, it was considered important to determine sequence specificity, if any, shown by their specially crafted designs.

Sequence specificity experiments were carried out with $(AGK)_2Cu$ (10) on a ^{32}P 3'-end. labelled 117 base pair restriction fragments (EcoRI/Hind III) of pUC18 containing sixteen 3'-CTAT-5' sites. Analysis of fragments resulting from the major breaks showed almost exclusive attacks of (10) at thymidylate residues, in particular at the 5'T of CTAT(3' -5') box. It was interesting to note that whilst the most preferred site of attack for (10) is at T of 3' ATC-5' region at the trinucleotide level, cleavage studies at low concentration have shown that at the pentanucleotide level, the lone sequence 3'–GATCT–5' is more favoured than the more frequently occurring 3'–TATCT–5' segments (figure 7).

Experimental section

Melting points are uncorrected. Optical rotations were measured with an automatic JASCO polarimeter; concentrations are given in grams/100 ml. Infrared spectra were recorded with a Perkin-Elmer 580/1600 FT spectrometers either as neat liquids or KBr pellets and prominent peaks are expressed in cm^{-1}. ^1H NMR spectra were determined on Bruker WM 400, WP 80 and Hitachi R 600 spectrometers. The chemical shifts are recorded in δ, ppm with TMS at 0·00 as internal reference. FAB mass was obtained on a JEOL SX-120/DA-6000 instrument using m-nitrobenzyl alcohol as the matrix. Elemental analysis was carried out with automatic C, H, N analyzer. Silica gel G (Merck) was used for TLC and column chromatography was done on silica gel (acme, 100–200 mesh) columns, which were generally made from a slurry in hexane or benzene. Reactions were monitored wherever possible by TLC.

Synthesis of peptides

General

All amino acids used were of L-configuration and all peptides were synthesized by solution phase method (Bodanszky 1988) using either DCC/HOBt mediated coupling or azide coupling via hydrazide or diphenyl phosphoryl azide. The crude peptides were, in most cases, directly crystallized from EtOAc/hexane mixture or purified on a short column of silica gel using benzene/EtOAc as eluents.

Selected data

9-Benzyl purinyl(6–N^{ω})–N^{α}Z–Lys–OH: m.p. 97–98°C; ^{1}H NMR (80 MHz; $(CD_3)_2SO$) δ 1·56 (6H, *br*), 3·64 (2H, *br*), 4·45 (1H, *m*), 5·07 (2H, *s*), 5·42 (2H, *s*), 7·04–7·82 (12H, *m*), 8·17 (1H, *s*), 8·26 (1H, *s*); IR: ν_{max}(KBr) cm^{-1}: 3337, 3288, 2934, 1687, 1628, 1589, 1531; MS (*m/z*) 489 $(M+H)^+$; 9-Benzylpurinyl(6–N^{ω})–Lys–OH: m.p. 245–247°C; ^{1}H NMR (80 MHz, $(CD_3)_2SO$) δ 1·53 (6H, *br*), 3·53 (2H, *br*), 4·46 (1H, *m*) 5·35 (2H, *s*), 7·29 (5H, *s*), 7·82–8·76 (3H, *br*), 8·23 (1H, *s*), 8·32 (1H, *s*); IR: ν_{max}(KBr) cm^{-1}: 3419, 3055, 1672, 1619, MS (*m/z*) 354 $(M)^+$, 355 $(M+H)^+$; $(Pur–Lys)_2Cu$ (1): m.p. 215–217°C; IR: ν_{max}(KBr) cm^{-1}: 3239, 2930, 1685, 1618; MS (*m/z*) 770 $(M)^+$, 355 $[1/2(M–Cu)+H]^+$; Quinazolyl(4–N^{ω})–N^{α}Z–Lys–OMe: m.p. 113–115°C; ^{1}H NMR (80 MHz, $CDCl_3$) δ 1·68 (6H, *m*), 3·43–3·84 (5H, *s* + *br*), 4·39 (1H, *m*), 5·07 (2H, *s*), 5·73 (1H, *d*, $J = 7·5$ Hz), 6·60 (1H, *m*), 7·04–8·14 (9H, *s* + *m*), 8·65 (1H, *s*); IR: ν_{max}(KBr) cm^{-1}: 3333, 3175, 2949, 1748, 1701, 1581; MS (*m/z*) 422 $(M)^+$; Quinazolyl(4–N^{ω})–Lys–OMe: m.p. 130–132°C; ^{1}H NMR (80 MHz, $CDCl_3$ + $(CD_3)_2SO$) δ 1·65 (6H, *m*), 3·50 (2H, *m*), 3·70 (3H, *s*), 4·34 (1H, *m*), 6·50 (3H, *br*), 7·10–8·00 (4H, *m*), 8·50 (1H, *s*); IR: ν_{max}(KBr) cm^{-1}: 3272, 2951, 1748, 1618, 1586, 1546; Quinazolyl(4–N^{ω})–Lys–OH: ^{1}H NMR (80 MHz, $(CD_3)_2SO$) δ 1·62 (6H, *m*), 3·62 (2H, *m*), 4·34 (1H, *m*), 7·35–8·07 (5H, *m*), 8·61 (2H, *br*), 8·71 (1H, *s*), 9·60 (1H, *br*); MS (*m/z*) 275 $(M+H)^+$; $(Qu–Lys)_2Cu$ (2): m.p. 208–210°C; IR: ν_{max}(KBr) cm^{-1}: 3329, 3280, 3237, 2941, 2856, 1615, 1507, 1469; MS (*m/z*) 613 $(M+H)^+$; Z–Ser–Lys(N^{α}Z)–OBzl: m.p. 70–72°C; $[\alpha]_D^{27}$: −17·5 (*c* 0·52, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3304, 3062, 3031, 2934, 1734, 1688, 1645, 1538; ^{1}H NMR (80 MHz, $CDCl_3$) δ 1·00–1·90 (6H, *m*), 3·10 (2H, *br*), 3·65 (2H, *br*), 4·15 (2H, *br*), 5·10 (6H, *brs*), 5·67 (1H, *brd*), 6·01 (1H, *brd*), 6·80 (1H, *brd*), 7·15 (15H, *brs*); Ser–Lys–OH (SK): m.p. 51–52°C; IR: ν_{max}(KBr) cm^{-1}: 3398, 3238, 2938, 1653, 1565; $(SK)_2Cu$ (3): m.p. 101°C; IR: ν_{max}(KBr) cm^{-1}: 3433, 2934, 1653, 1617, 1575; UV–Vis: λ_{max}(H_2O) nm (ε, L mol^{-1} cm^{-1}): 631 (89); EPR (H_2O:EtOH, −196°C): $A_{\parallel}$ 175, g_1 2·221, g_2 2·067, g_3 2·026; Boc–Asn–Lys(N^{α}Z)–OBzl: m.p. 154–156°C; $[\alpha]_D^{26}$: −30·79 (*c* 0·31, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3402, 3326, 3297, 3030, 2978, 2950, 1732, 1686, 1656, 1527; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 1·09–2·00 (15H, *s* + *m*), 2·59 (2H, *br*), 3·28 (2H, *m*), 4.29 (2H, *m*), 5·14, 5.21 (2H, 2H, *s*, *s*), 6·43 (2H, *m*), 7·59 (13H, *s* + *m*); MS (*m/z*) 585 $(M+H)^+$; Boc–Asn–Lys(ω-pep) (NK): m.p. 202–204°C; IR: ν_{max}(KBr) cm^{-1}: 3458, 2967, 2937, 1681, 1657, 1588, 1520; ^{1}H NMR (80 MHz, D_2O) δ 1·25–2·00 (15H, *br*), 2·59 (2H, *br*), 3·28 (2H, *m*), 4·29 (2H, *m*); MS (*m/z*) 361 $(M+H)^+$; $(NK)_2Cu$ (4): m.p. 185–186°C; IR: ν_{max}(KBr) cm^{-1}: 3400, 3330, 2933, 1684, 1655, 1525; UV–Vis: λ_{max}(H_2O) nm (ε, L mol^{-1} cm^{-1}): 615 (36); EPR (H_2O:EtOH, −196°C): $A_{\parallel}$ 175,

g_1 2·187, g_2 2·0023, g_3 1·960; MS (*m/z*) 782 $(M)^+$; Z–Gln–Lys(N^{α}Z)–OBzl: m.p. 148–150°C; $[\alpha]_D^{28}$: −16·26 (*c* 0·75, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3424, 3317, 3181, 2938, 1736, 1698, 1653, 1542; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 0·93–2·35 (10H, *brm*), 3·12 (2H, *m*), 4·04 (2H, *m*), 5·03, 5·09 (4H, 2H, *s, s*), 6·10 (1H, *br*), 6·70 (1H, *d*, $J = 7{\cdot}5$ Hz), 6·95 (1H, *d*, $J = 5{\cdot}0$ Hz), 7·07–7·73 (17H, *m*); MS (*m/z*) 633 $(M + H)^+$, 655 $(M + Na)^+$; Gln–Lys(ω-pep) (QK): m.p. gummy; IR: ν_{max}(KBr) cm^{-1}: 3342, 2945, 1702, 1670, 1560; MS (*m/z*) 274 $(M)^+$; $(QK)_2Cu$ (5): m.p. 215–217°C; IR: ν_{max}(KBr) cm^{-1}: 3415, 3291, 2930, 2856, 1682, 1617; UV–Vis: $\lambda_{max}(H_2O)$ nm (ε, L $mol^{-1}cm^{-1}$): 620 (29); Boc–Tyr–Lys(N^{α}Z)–OBzl: m.p. 127–129°C; $[\alpha]_D^{28}$: −7·17 (*c* 0·85, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3340, 2968, 1730, 1685, 1646, 1519; ^{1}H NMR (80 MHz), $CDCl_3 + (CD_3)_2SO$ δ 0·78–1·90 (15H, *s + m*), 2·93 (4H, *m*), 4·25 (2H, *m*), 5·17, 5·20 (2H, 2H, *s, s*), 5·60 (2H, *m*), 6·92 (4H, *dd*, $J = 12{\cdot}5$ Hz, 10 Hz), 7·42 (11H, *brs*); MS (*m/z*) 634 $(M + H)^+$; Boc–Tyr–Lys(ω-pep) (YK): m.p. 203–204°C; IR: ν_{max}(KBr) cm^{-1}: 3310, 3053, 2947, 1645, 1520; ^{1}H NMR (80 MHz, D_2O) δ 1·00–1·87 (15H, *s + br*), 2·93 (4H, *m*), 4·25 (2H, *m*), 6·29 (1H, *d*, $J = 7{\cdot}5$ Hz), 6·76 (4H, *m*), 7·67 (1H, *br*); MS (*m/z*) 410 $(M + H)^+$; $(YK)_2Cu$ (6): m.p. 193–195°C; IR: ν_{max}(KBr) cm^{-1}: 3337, 2933, 2864, 1686, 1650, 1516; UV–Vis: λ_{max}(DMSO:H_2O) nm (ε, L mol^{-1} cm^{-1}): 629 (59); MS (*m/z*) 881 $(M + H)^+$; Bz–Trp–Lys(N^{α}Z)–OBzl: m.p. 138–140°C; IR: ν_{max}(KBr) cm^{-1}: 3420, 3340, 3080, 2950, 2880, 1722, 1682, 1627, 1570, 1530; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 0·87–2·00 (6H, *brm*), 3·07 (2H, *m*), 3·28 (2H, *m*) 4·17 (1H,*m*) 4·79 (1H, *m*), 5·07, 5·10 (2H, 2H, *s, s*), 6·54 (1H, *d*, $J = 7{\cdot}5$ Hz), 6·76–8·00 (22H, *s + m*), 9·73 (1H, *brs*); MS (*m/z*) 661 $(M + H)^+$; Bz–Trp–Lys(ω-pep) (WK): m.p. 168°C; IR: ν_{max}(KBr) cm^{-1}: 3372, 3180, 3000, 2950, 1680, 1660, 1521, 1510; MS (*m/z*) 437 $(M + H)^+$, 459 $(M + Na)^+$; $(WK)_2Cu$ (7): m.p. 176–178°C; IR: ν_{max}(KBr) cm^{-1}: 3319, 2928, 1638, 1526; UV–Vis: $\lambda_{max}(H_2O)$ nm (ε, L mol^{-1} cm^{-1}): 674 (101); Bz–Cit–Lys(N^{α}Z)–OBzl: m.p. 120–121°C; IR: ν_{max}(KBr) cm^{-1}: 3449, 3290, 3061, 2928, 2853, 1724, 1690, 1650, 1627, 1534; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 1·00–2·00 (10H, *brm*), 3·18 (4H, *m*), 4·17 (1H, *m*), 4·57 (1H, *m*), 5·07, 5·14 (2H, 2H, *s, s*), 5·98 (1H, *m*), 6·92–8·34 (20H, *s + m*); Bz–Cit–Lys(ω-pep) (CitK): m.p. 139–140°C; IR: ν_{max}(KBr) cm^1: 3325, 3062, 2928, 2850, 1626, 1575, 1536; $(CitK)_2Cu$ (8): m.p. 171–173°C; IR: ν_{max}(KBr) cm^{-1}: 3351, 2930, 2860, 1640, 1610, 1575, UV–Vis: λ_{max}(DMSO:H_2O) nm (ε, L mol^{-1} cm^{-1}) 630 (57); EPR (H_2O:EtOH, −196°C): $A_{\parallel}$ 175, g_1 2·251, g_2 2·057, g_3 2·018; Boc–Arg(N^GNO_2)–Lys(N^{α}Z)–OBzl: m.p. 142–143°C; IR: ν_{max}(KBr) cm^{-1}: 3328, 2928, 2854, 1733, 1703, 1655, 1534, 1454, 1391, 1367, 1258; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 1·00–2·00 (19H, *s + m*), 3·12 (4H, *m*), 3·98 (2H, *m*), 5·01, 5·10 (2H, 2H, *s, s*), 6·29 (2H, *d*, $J = 7{\cdot}5$ Hz), 7·00–8·09 (14H, *s + m*); MS (*m/z*) 672 $(M + H)^+$, 694 $(M + Na)^+$; Boc–Arg(N^GNO_2)–Lys(ω-pep) (RK): m.p. 110–111°C; $[\alpha]_D^{27}$: −5·9 (*c* 0·66, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3354, 2931, 1659, 1528, 1450, 1395, 1250; $(RK)_2Cu$ (9): m.p. 160–162°C; IR: ν_{max}(KBr) cm^{-1}: 3322, 2976, 1662, 1451, 1395, 1367, 1249; UV–Vis: $\lambda_{max}(H_2O)$ nm (ε, L mol^{-1} cm^{-1}): 643 (57); EPR (H_2O:EtOH, −196°C): $A_{\parallel}$ 178, g_1 2·296, g_2 2·094, g_3 2·055; Bz–Ala–Gly–OMe: m.p. gummy; IR: ν_{max}(neat) cm^{-1}: 3320, 3065, 2982, 2951, 1732, 1682, 1650, 1542; Bz–Ala–Gly–Lys(N^{α}Z)–OMe: m.p. 151–153°C; IR: ν_{max}(KBr) cm^{-1}: 3320, 3061, 2952, 2867, 1742, 1684, 1638, 1536; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 1·50 (9H, *d + m*), 3·18 (2H, *m*), 3·65 (5H, *s + m*), 3·87 (1H, *m*), 4·14 (1H, *m*), 5·03 (2H, *s*), 6·48 (1H, *d*, $J = 7{\cdot}5$ Hz), 7·17–8·35 (13H, *m*); MS (*m/z*) 527 $(M + H)^+$; Bz–Ala–Gly–Lys(ω-pep) (AGK): m.p. gummy; IR: ν_{max}(KBr) cm^{-1}: 3313, 1703, 1644, 1537, MS (*m/z*) 379 $(M + H)^+$; $(AGK)_2Cu$ (10): m.p. 202–203°C; IR: ν_{max}(KBr) cm^{-1}: 3219, 1624, 1559, UV–Vis: $\lambda_{max}(H_2O)$ nm

(ε, L mol^{-1} cm^{-1}): 759 (117); MS (*m/z*) 818 $(M+H)^+$; Bz–Trp–Trp–OMe: m.p. 188–190°C; IR: ν_{max}(KBr) cm^{-1}: 3380, 3305, 3048, 2927, 2849, 1743, 1694, 1651, 1538; ^{1}H NMR (80 MHz, $CDCl_3$) δ 2·90 (4H, *m*), 3·56 (3H, *s*), 4·62 (2H, *m*), 6·12–8·00 (19H, *m*); Bz–Trp–Trp–Lys(N$^\alpha$Z)–OBzl: m.p. 110–112°C; $[\alpha]_D^{24}$: −25·11 (*c* 0·86, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3315, 3058, 2930, 1740, 1519; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 1·45 (4H, *m*), 1·62 (2H, *br*), 3·03 (6H, *m*), 4·34 (3H, *m*), 4·90, 5·09 (2H, 2H, *s* + *m*), 6·10 (1H, *d*, $J = 7·5$ Hz), 6·48–7·82 (28H, *m*), 9·93, 10·10, (1H, 1H, *s*, *s*); Bz–Trp–Trp–Lys(ω-pep) (WWK): m.p. gummy; IR: ν_{max}(neat) cm^{-1}: 3258, 2924, 2853, 1653, 1558, 1458; $(WWK)_2Cu$ (11): m.p. 218–220°C; IR: ν_{max}(KBr) cm^{-1}: 3391, 3261, 2923, 1647, 1607; UV–Vis: λ_{max}(DMSO:H_2O) nm (ε, L mol^{-1} cm^{-1}) 663 (21); Bz–Ala–Gly–Trp–OMe: m.p. 163–164°C; IR: ν_{max}(KBr) cm^{-1}: 3578, 3399, 3290, 3054, 1746, 1668, 1578, 1534; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 1·40 (3H, *d*), 3·15 (2H, *m*), 3·62 (3H, *s*), 3·18 (2H, *m*), 4·36 (2H, *m*), 6·85–8·42 (13H, *m*), 10·39 (1H, *s*); MS (*m/z*) 451 $(M+H)^+$; Bz–Ala–Gly–Trp–Lys(N$^\alpha$Z)–OBzl: m.p. 87–89°C; $[\alpha]_D^{26}$: −7·87° (*c* 0·94, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3312, 3060, 2933, 2863, 1714, 1646, 1532; ^{1}H NMR (80 MHz, $CDCl_3$) δ 0·71–1·75 (9H, *brm*), 2·57–3·34 (2H, *br*), 3·62 (2H, *br*), 4·07 (3H, *m*), 4·56 (2H, *m*), 4·95, 5·04 (2H, 2H, *s*, *s*), 6·07 (1H, *br*), 6·73–8·14 (24H, *m*), 9·15 (1H, *brs*); MS (*m/z*) 789 $(M+H)^+$; Bz–Ala–Gly–Trp–Lys (ω-pep) (AGWK): m.p. 175–177°C; IR: ν_{max}(KBr) cm^{-1}: 3286, 3052, 2932, 1645, 1533; $(AGWK)_2Cu$ (12): m.p. 172–174°C; IR: ν_{max}(KBr) cm^{-1}: 3310, 2930, 1645, 1532; UV–Vis: λ_{max}(DMSO:H_2O) nm (ε, L mol^{-1} cm^{-1}) 624 (38); EPR (MeOH, −196°C) $A_\parallel$ 180, g_1 2·246, g_2 2·057, g_3 2·017; Bz–Gly–Gly–His–Lys(N$^\alpha$Z)–OBzl: m.p. 108–109°C; $[\alpha]_D^{24}$: −16·32 (*c* 0·79, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3363, 3033, 2950, 1733, 1690, 1620, 1519; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 1·53 (6H, *br*), 2·79 (2H, *br*), 2·81–3·84 (6H, *br*), 4·18 (2H, *m*), 5·07, 5·14 (2H, 2H, *s*, *s*), 6·60 (1H, *d*, $J = 7·5$ Hz), 6·85–8·18 (21H, *m*); Bz–Gly–Gly–His–Lys(ω-pep) (GGHK): m.p. 237–239°C; IR: ν_{max}(KBr) cm^{-1}: 3423, 3065, 2931, 1690, 1647, 1605, 1576; $(GGHK)_2Cu$ (13): m.p. 196–198°C; IR: ν_{max}(KBr) cm^{-1}: 3376, 3034, 2922, 1658, 1602; UV–Vis: λ_{max}(DMSO:H_2O) nm (ε, L mol^{-1} cm^{-1}): 615 (54); EPR (H_2O:EtOH, −196°C): $A_\parallel$ 175, g_1 2·254, g_2 2·057, g_3 2·008; Z–Leu–Ser–Ser–Lys(N$^\alpha$Z)–OBzl: m.p.122–123°C; IR: ν_{max}(KBr) cm^{-1}: 3284, 2958, 1742, 1719, 1690, 1667, 1532; ^{1}H NMR (80 MHz, $CDCl_3$) δ 0·90 (6H, *d*, $J = 5·0$ Hz), 1·05–2·00 (9H, *m*), 3·09 (2H, *m*), 3·87–4·68 (8H, *m*), 5·04 (6H, *brs*), 5·51 (1H, *br*), 5·89 (1H, *d*, $J = 7·5$ Hz), 6·71–7·76 (16H, *m*), 8·14 (2H, *m*); MS (*m/z*) 791 $(M)^+$, 700 $(M^+ - C_6H_5CH_2)$; Leu–Ser–Ser–Lys(ω–pep) (LSSK): m.p. gummy; IR: ν_{max}(KBr) cm^{-1}: 3224, 2962, 1684, 1556; $(LSSK)_2Cu$ (14): m.p. 160–161°C; IR: ν_{max}(KBr) cm^{-1}: 3218, 2953, 1641, 1599, UV–Vis: λ_{max}(H_2O) nm (ε, L mol^{-1} cm^{-1}): 720 (88); MS (*m/z*) 928 $(M)^+$; Bz–Gly–Gly–Gly–Gly–Lys(N$^\alpha$Z)–OBzl: m.p. 151–152°C; $[\alpha]_D^{24}$: −12·64 (*c* 0·67, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3340, 3100, 2957, 2882, 1732, 1700, 1660, 1640, 1548; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$) δ 1·09–2·00 (6H, *m*), 3·09 (2H, *m*), 3·82, 3·95 (4H, 4H, *d*, *d*, $J = 5·0$ Hz, 5·0 Hz), 4·15 (1H, *m*), 5·06, 5·13 (2H, 2H, *s*, *s*), 7·00 (1H, *d*, $J = 7·5$ Hz), 7·10–8·10 (19H, *m*), 8·59 (1H, *t*); MS (*m/z*) 611 $(MH\text{-}C_6H_5CH_2)^+$; Bz–Gly–Gly–Gly–Gly–Lys(ω–pep) (GGGGK): m.p. 130–132°C; IR: ν_{max}(KBr) cm^{-1}: 3307, 3069, 2935, 1651, 1541; $(GGGGK)_2Cu$ (15): m.p. 190–192°C; IR: ν_{max}(KBr) cm^{-1}: 3445, 3296, 2926, 2869, 1652, 1541; UV–Vis: λ_{max}(DMSO:H_2O) nm (ε, L mol^{-1} cm^{-1}): 641 (23); EPR (H_2O:EtOH, −196°C): $A_\parallel$ 180, g_1 2·252, g_2 2·061, g_3 2·020; Bz–Ala–Gly–Ala–Gly– Lys(N$^\alpha$Z)–OBzl: m.p. gummy; $[\alpha]_D^{24}$: −13·76 (*c* 0·77, MeOH); IR: ν_{max}(KBr) cm^{-1}: 3302, 3064, 2935, 1714, 1649, 1536; ^{1}H NMR (80 MHz, $CDCl_3 + (CD_3)_2SO$)

δ 1·00–2·10 (12H, *m*), 3·12 (2H, *m*), 4·04 (4H, *brd*), 4·29 (2H, *m*), 4·76 (1H, *m*), 5·07, 5·14 (2H, 2H, *s*, *s*), 6·73 (2H, *br*), 7·01–8·10 (19H, *m*); MS (*m/z*) 639 $(MH\text{-}C_6H_5CH_2)^+$; Bz–Ala–Gly–Ala–Gly–Lys(ω-pep) (AGAGK): m.p. 98–100°C; IR: ν_{max}(KBr) cm^{-1}: 3316, 3063, 2929, 2860, 1652, 1579, 1539; $(AGAGK)_2Cu$ (16): m.p. 103–104°C; IR: ν_{max}(KBr) cm^{-1}: 3402, 3302, 2928, 1629, 1542; UV–Vis: λ_{max}(DMSO:H_2O) nm (ε, $L\,mol^{-1}\,cm^{-1}$): 639 (41); EPR (H_2O:EtOH, −196°C): $A_{\parallel}$ 175, g_1 2·255, g_2 2·057, g_3 2·020.

Summary

The present work describes the crafting of a design for sequence specific chemical nuclease model, wherein pairs of symmetrically placed DNA recognition elements are linked by a tetramethylene spacer to the core cleavage motif, which is crafted from the α amino carboxylate unit as the copper complex. This unique design crafted from L-lysine is simple and flexible and has striking similarities with DNA recognition protein such as GAL-4, wherein the recognition takes place at the termini of a symmetric duplex, assembled by a central dimerisation element linked by a spacer to the end units.

In the present design, the recognition elements thus far studied encompass a spectrum of structures ranging from quinazoline and purine residues to specially crafted peptide segments, that may have potential to form secondary structures. These models, synthesized by conventional peptide synthesis procedures, have been fully characterized by analytical and spectral (UV, IR, EPR, NMR, MS) means. The DNA binding and scission profile of all the synthetic models described in the present work have been probed by UV, fluoresence quenching and gel electrophoretic studies. The DNA cleaving properties, using supercoiled pBR 322 and p-blue script11KS plasmid DNA have shown that in all cases, covalently closed (form I) DNA is converted largely into open circular (form II) and that the cleavage is facilitated with increasing concentration, resulting, in cases, the generation of linear form (form III) arising from the scission of both the strands.

Sequence specific experiments with one of the models $(Bz\text{–}Ala\text{–}Gly\text{–}Lys)_2Cu$ (10) has strongly indicated the possibility of palindromic recognition in this model.

In conclusion, the present work has shown that the design of chemical nucleases having the cleaving domain flanked by symmetric dimeric recognition element is feasible, and thus by a judicious selection may lead to composites with a higher order of sequence specificity, a property very desirable in diverse genomic studies.

Acknowledgements

We are most grateful to Professor S Ranganathan for his encouragement and keen interest in this work and to Professor D Balasubramanian and Dr D Chatterji (Centre for Cellular and Molecular Biology, Hyderabad) for helpful discussions. Financial support was provided by the Department of Science and Technology, New Delhi.

References

Anthony-Cahill S J, Benfield P A, Fairman R, Wasserman Z R, Brenner S L, Stafford W F III, Altenbach C, Hubbell W and Degrado W F 1992 *Science* **255** 979

Armarego W L F 1961 *J. Appl. Chem.* **11** 70

Baleja J P, Marmorstein R, Harrison S C and Wagner G 1992 *Nature (London)* **356** 450

Bodanszky M 1988 *Peptide chemistry, A practical textbook* (New York: Springer-Verlag)

Corey D R, Pei D and Schultz P G 1990 *Biochemistry* **28** 8277

Costopanagiotis A A, Handford B O and Weinstein B 1968 *J. Org. Chem.* **33** 1261

Davis R L, Cheng P -F, Lasser A B and Wintraub H 1990 *Cell* **60** 773

Dervan P B 1986 *Science* **232** 464

Dervan P B 1992 *Nature (London)* **359** 87

Distefano M D, Shin J A and Dervan P B 1991 *J. Am. Chem. Soc.* **113** 5901

Gehring W J, Muller M, Affolter M, Percival-Smith A, Lleter M B, Quian Y Q, Otting G and Wuthrich K 1990 *Trends Genet.* **6** 323

Harrison S C and Aggrawal A K 1990 *Annu. Rev. Biochem.* **59** 933

Keçk M V and Lippard S J 1992 *J. Am. Chem. Soc.* **114** 3386

Kissinger C R, Liu B, Martin-Blanco E, Kornberg T B and Pabo C O 1990 *Cell* **63** 579

Klug A and Rhodes D 1987 *Trends Biochem.* **12** 464

Koh J S and Dervan P B 1992 *J. Am. Chem. Soc.* **114** 1470

Kraulis P J, Raine A R C, Gadhavi P L and Lane E D 1992 *Nature (London)* **356** 448

Landshultz W H, Johnson P F and Mcknight S L 1988 *Science* **240** 1759

Lasser A B, Buskin J N, Lockson D, Davis R L, Apone S, Hauschka S D and Wintræub H 1989 *Cell* **58** 823

Mack D P and Dervan P M 1990 *J. Am. Chem. Soc.* **112** 4604

Mack D P, Iverson B L and Dervan P B 1988 *J. Am. Chem. Soc.* **110** 7572

Marmorstein R, Carey M, Plastine M and Harrison S C 1992 *Nature (London)* **356** 408

Montgomery J A and Temple C 1961 *J. Am. Chem. Soc.* **83** 630

Murre C, McCow P S and Baltimore D 1989 *Cell* **56** 777

Nagai K, Carter B J, Xu J and Hecht S M 1991 *J. Am. Chem. Soc.* **113** 5099

Oakly M and Dervan P B 1990 *Science* **248** 847

O'Neil K T, Hoess R H and Degrado W F 1990 *Science* **249** 774

O' Shea E K, Rutkowski R and Kim P S 1989 *Science* **243** 538

Pabo C O and Sauer R T 1984 *Annu. Rev. Biochem.* **53** 293

Pavletich N P and Pabo C O 1991 *Science* **252** 809

Pei D and Schultz P G 1990 *J. Am. Chem. Soc.* **112** 4579

Prendergest G C and Ziff E B 1989 *Nature (London)* **341** 392

Pyle A M, Morii T and Barton J K 1990 *J. Am. Chem. Soc.* **112** 9432

Sigman D S 1990a *Biochemistry* **29** 9097

Sigman D S and Chen C B 1990b *Annu. Rev. Biochem.* **59** 207

Sitlani A, Long E C, Pyle A M and Barton J K 1992 *J. Am. Chem. Soc.* **114** 2303

Sluka J P, Griffin J H, Mack J P and Dervan P B 1990 *J. Am. Chem. Soc.* **112** 6369

Strobel S A and Dervan P B 1990 *Science* **249** 73

Sun X -H and Baltimore D 1991 *Cell* **64** 459

Talanian R V, Mcknight C J and Kim P S 1990 *Science* **249** 769

Uhlmann E and Peyman A 1990 *Chem. Rev.* **90** 543

Vallee B L, Coleman J E and Auld D S 1991 *Proc. Natl. Acad. Sci. USA* **88** 999

Vinson C R, Sigler P B and Mcknight S L 1989 *Science* **246** 911