Design of a Simple and Flexible Dimeric Peptide Model for DNA Recognition and Scission

Darshan Ranganathan,*[a] Bhisma Kumar Patel[b] and Rakesh K. Mishra[c]

[a] *Regional Research Laboratory (CSIR), Trivandrum 695 019, India*
[b] *Department of Chemistry, Indian Institute of Technology, Kanpur 208 016, India*
[c] *Centre for Cellular and Molecular Biology, Hyderabad 500 007, India*

Models, containing symmetric dimeric peptide units as the recognition systems, constructed by attaching di, tri and tetra peptide units at the ω-NH_2 end of the duplex termini in $(Lys)_2Cu$, bind to DNA in a sequence specific manner and effect scission at specific sites.

We have recently reported[1] a chemical nuclease model crafted from L-lysine wherein the DNA-cleaving centre was placed at the core of a duplex recognition termini harbouring purines and quinazolines separated from the centre by a four methylene spacer.

Many naturally occurring sequence-specific DNA-binding proteins bind to DNA, most effectively, as dimer.† This communication reports that closely related designs containing symmetric dimeric peptide units as recognition elements, related by a twofold axis of symmetry and having the cleaving elements at the core, are effective in bringing about sequence-specific DNA scission. The present model (Scheme 1) can be considered as being composed of two sub-units, each carrying a peptide recognition module attached *via* a linker to the central metal template, which acts as the dimerization element. Additionally, the redox active nature of the metal template enables it to act as a cleaving centre for the DNA double helix. The design is simple, flexible and in profile, akin to restriction enzymes.‡

Scheme 1 presents a profile of chemical nucleases **1–14**, which have been assembled§ by attachment of the appropriate recognition termini comprising of peptide units,¶ so selected to create diverse environments, and covering almost the whole spectrum of coded amino acids.‖

The binding capability of these models was probed by UV absorption and fluorescence quenching studies. Their binding to calf thymus DNA is associated with a decrease in extinction coefficient, and a *ca.* 10–15 nm red shift. Fluorescence emission studies (Fig. 1) of $(WWK)_2Cu$ **9** with increasing amounts of salmon sperm DNA demonstrate 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 *ca.*

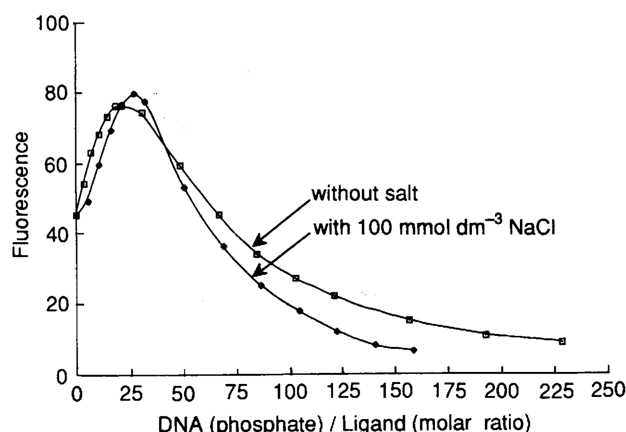
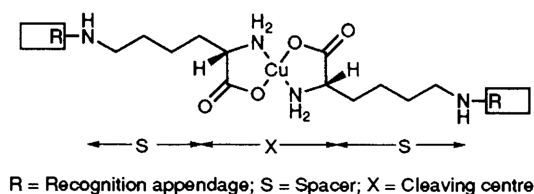


Fig. 1 Fluorescence emission spectra (excitation, 280 nm; emission 350 nm) of $(WWK)_2Cu$ **9** with increasing salmon sperm DNA/ligand ratio. 5 μmol dm^{-3} solution of WWK in water without NaCl or with 0.1 mol dm^{-3} NaCl in a final volume of 10 μl tris 1 mol dm^{-3} EDTA, pH 7.5 was used for measurements.



R = Recognition appendage; S = Spacer; X = Cleaving centre

R = H; Code: K–Cu–K

Scheme 1

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

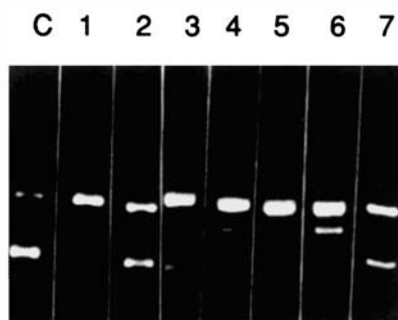


Fig. 2 Cleavage of supercoiled plasmid p Blue script 11 KS DNA with $(SK)_2Cu$, $(WK)_2Cu$, $(CitK)_2Cu$, $(RK)_2Cu$, $(GGHK)_2Cu$, $(LSSK)_2Cu$ and $(AGAGK)_2Cu$ cleaving agents

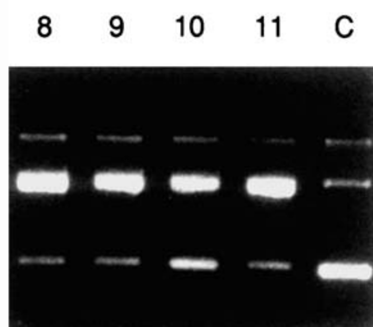


Fig. 3 Cleavage of pBR 322 supercoiled plasmid DNA with $(NK)_2Cu$, $(QK)_2Cu$, $(YK)_2Cu$, and $(AGK)_2Cu$

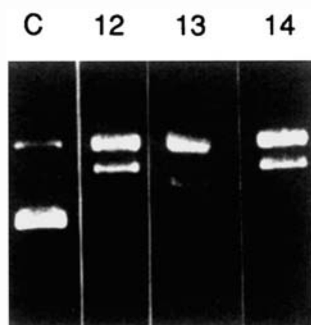


Fig. 4 Cleavage of p Blue script 11 KS supercoiled DNA with $(WWK)_2Cu$, $(AGWK)_2Cu$ and $(GGGGK)_2Cu$

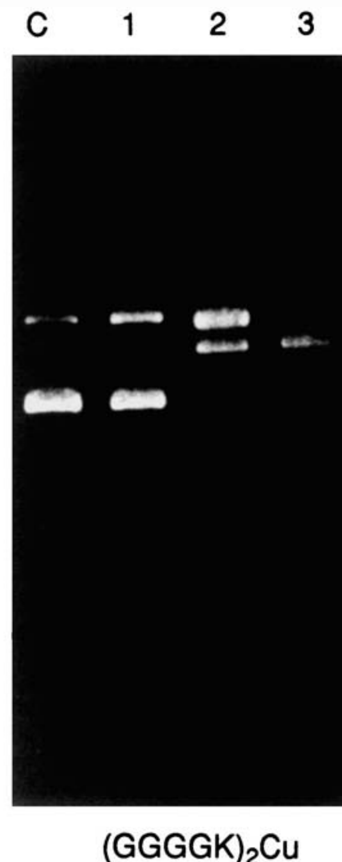


Fig. 5 Concentration-dependent cleavage of p-Blue script 11 KS supercoiled DNA with $(GGGGK)_2Cu$ **13**

225 (1 ligand bound per *ca* 112 base pairs). This ratio falls to *ca*. 160 (1 ligand per *ca*. 80 base pairs) in the presence of 0.1 mol dm^{-3} NaCl.**

Fig. 2, 3 and 4 show the cleavage results of compounds **1–14** with supercoiled p Blue script 11 KS and pBR 322 DNA at pH 7.5 at 37 °C. In all instances, 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 cuts in the case of **6**, **12** and **14**, arising from scission of both the strands is clearly visible in the gel picture. Fig. 5 shows the concentration dependence of DNA cleavage with $(GGGGK)_2Cu$ **14**. As could be seen clearly in the gel picture, increase of concentration produces linear cuts in the supercoiled form.

Sequence specificity experiments were carried out with $(AGK)_2Cu$ **8** on a ^{32}P 3′-end labelled 117 base pair restriction fragment (Eco RI/Hind III) of pUC 18 containing sixteen 3′-CTAT-5′ sites. Analysis of fragments resulting from the major breaks showed almost exclusive attacks of **8** 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 **8** 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′ (a part of the inverted repeat -GAGATCTC-) is favoured (fragment 92) over the more frequently occurring 3′-TATCT-5′ segment (Fig. 6).†† Sequence specificity experiments with other models are in progress.

The present work has shown the viability of using peptide units as duplex recognition appendages in $(Lys)_2Cu$ based nuclease model. The present design is flexible and would permit the attachment of a variety of recognition ligands at the ω-amino end—for example, peptide units derived from the well-identified DNA-binding structural motifs such as helix–turn–helix,[2] leucine zipper,[3] Zn fingers[4] and helix–loop–helix[5]—to generate chemical nucleases tuned to recognition of targetted sequences by binding to DNA as symmetric dimers[6] related by a twofold axis of symmetry and effect scission in a predictable manner.

We thank Professor S. Ranganathan, IIT, Kanpur, for his constant advice and encouragement and Professor D. Balasubramanian, Dr D. Chatterji, CCMB, Hyderabad for helpful discussions and for their interest in this work. Experimental help from Miss Bindu Diana Paul, a Summer trainee at CCMB

17 20 30 40 ↓ 50↓ 60 ↓ 70↓ ↓ 80
CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT

1
** 10 16 81 90↓ 100 110 117
3′ AAGCTCGAGCCATGGG CCCCTAGGAGATCTCAGCTGGACGTCCGTACGTTCGA 5′

Fig. 6 $(AGK)_2Cu$ **8** induced cleavage sites on a ^{32}P 3′-end labelled 117 base pairs 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 pentanucleotide level.

is gratefully acknowledged. The work was supported by a grant from DST, New Delhi.

Received, 6th May 1993; Com. 3/02608F

Footnotes

† Dimerization prior to or during the process of DNA recognition has emerged as a unifying property of most of the sequence-specific DNA binding proteins. Constructs designated as dimerization elements, varied in profile, permit dimer formation. For example, in regulatory proteins, such as Met J repressor, the two antiparallel β-sheet strands make up the dimerization element; in yeast activator protein GAL 4, two amphipathic α-helices pack together to form the dimerization 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 dimerization can be understood to reflect the fact that if dimer acts as a single structural element, the binding should double the DNA contact area and thereby square the affinity constant.

‡ Most restriction enzymes exhibit a twofold rotational symmetry in their recognition sequences and so recognize palindromic sequences in DNA.

§ The three steps of the strategy employed for the construction of models **1–14** (Scheme 1) are as follows. (*i*) the preparation of the recognition systems represented by peptide units, for example, dipeptides **8** and **9** (R = A–G and W–W, respectively), tripeptides **10**, **11** and **12** (R = A–G–W, G–G–H and L–S–S, respectively) and tetrapeptides **13** and **14** (R = G–G–G–G and A–G–A–G, respectively), using a combination of DCC/HOBt and azide coupling procedures in solution phase. (*ii*) The coupling of these recognition appendages (the single amino acid residues **1–7** and the peptide units **8–14** to the ω-NH_2 end of N^{α}Z-Lys-CO_2Bzl (Z = benzyloxycarbonyl, Bzl = benzyl) giving rise to spacer linked recognition modules. (*iii*) The complexation with Cu^{II} (basic cupric carbonate/aqueous methanol/heat/ *ca.* 10 min) of the *N*-, *C*-deprotected recognition systems, brings about the dimerization of these units leading to Cu-bridged symmetric duplexes (**1–14**, Scheme 1).

All N, C protected peptides listed in Scheme 1 were purified by column chromatography on Silica gel using $CHCl_3$–MeOH as eluent and were shown to be homogenous (TLC). The complete absence of diastereoisomeric mixtures, that may arise by racemisation, was established by their clean ^1H NMR spectra at 400 MHz.

In their IR spectra, these complexes exhibited characteristics bands in the region 3450–3300, 2950–2920, 1680–1650 and 1570–1520 cm^{-1}.

A characteristic four-line pattern was seen, at room temp., in the EPR spectra (aq. ethanol) of these complexes (**1–14**, Scheme 1), which changes to an axial spectrum with rhombic distortion, at liquid nitrogen temperature: A_{II} = 175–180; g_{II} = 2.187–2.296; g_2 = 2.0023–2.094 and g_3 = 1.960–2.055.

For all compounds, one or more bands were observed in the range 500–850 nm expected for four-coordinated Cu^{II} complexes.

The dimeric nature of the complexes was strongly supported by their elemental (C, H and N) analysis and FAB MS. *Selected data* for Cu complexes **1–14** (Scheme 1): entry: mp °C decomp.; FAB-MS **1**: 101–102; **2**: 185–186; 783 $(MH)^+$; **3**: 215–216; **4**: 193–195; 881 $(MH)^+$; **5**: 176–178; **6** 171–173; **7** 160–162; 447 $[(M\text{-}Cu)/2 + H]^+$; **8** 202–203; 818 $(M)^+$; **9** 218–220; 1342 $(M^+ + 2H_2O)$; **10** 172–174; 1190 (M^+); **11** 196–198; **12** 160–161; 928 (M^+); **13** 190–192; 1054 $(M^+ + 2H_2O)$; **14** 103–104.

¶ All *N*-, *C*-protected peptides listed in Scheme 1 gave satisfactory spectral (^1H NMR, IR and FAB MS) and analytical data.

|| 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 decrease in specificity of binding (low, DNA/ligand ratio) may be due to the enhanced rigidity induced in DNA with salt, as a result of which the ligand settles down to a larger number of less preferred sites.

†† A non-specific Cu^{II} catalysed oxidative cleavage was seen when a ^{32}P 3′-end labelled 117 base pair restriction fragment (Eco RI/Hind III) of pUC 18, was digested with either $(Pur\text{-}Lys)_2Cu$ or $(Qu\text{-}Lys)_2Cu$ (ref. 1).

References

1 D. Ranganthan, B. K. Patel and R. K. Mishra, *J. Chem. Soc., Chem. Commun.*, 1993, 337.

2 C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.*, 1984, **53**, 293; S. C. Harrison and A. K. Aggarwal, *Annu. Rev. Biochem.*, 1990, **59**, 933; W. J. Gehring, M. Muller, M. Affolter, A. Percival-Smith, M. B. Lleter, Y. Q. Quian, G. Otting and K. Wuthrich, *Trends Genet.*, 1990, **6**, 363; C. R. Kissinger, B. Liu, E. Martin-Blanco, T. B. Kornberg and C. O. Pabo, *Cell*, 1990, **63**, 579.

3 W. H. Landshultz, P. F. Johnson and S. L. Mcknight, *Science*, 1988, **240**, 1759; C. R. Vinson, P. B. Sigler and S. L. Mcknight, *Science*, 1989, **246**, 911; K. T. O'Neil, R. H. Hoess and W. F. Degrado, *Science*, 1990, **249**, 774; E. K. O'Shea, R. Rutkowski and P. S. Kim, *Science*, 1989, **243**, 538; M. Oakly and P. B. Dervan, *Science*, 1990, **248**, 847; R. V. Talanian, C. J. Mcknight and P. S. Kim, *Science*, 1990, **249**, 769.

4 A. Klug and D. Rhodes, *Trends Biochem.*, 1987, **12**, 464; B. L. Vallee, J. E. Coleman and D. S. Auld, *Proc. Natl. Acad. Sci., USA*, 1991, **88**, 999; N. P. Pavletich and C. O. Pabo, *Science*, 1991, **252**, 809; R. Marmorstein, M. Carey, M. Ptashne and S. C. Harrison, *Nature*, 1992, **356**, 408; P. J. Kraulis, A. R. C. Raine, P. L. Gadhavi and E. D. Lane, *Nature*, 1992, **356**, 448; J. P. Baleja, R. Marmorstein, S. C. Harrison and G. Wagner, *Nature*, 1992, **356**, 450.

5 G. C. Prendergest and E. B. Ziff, *Nature*, 1989, **341**, 392; C. Murre, P. S. McCow and D. Baltimore, *Cell*, 1989, **56**, 777; A. B. Lasser, J. N. Buskin, D. Lockson, R. L. Davis, S. Apone, S. D. Hauschka and H. Wintraub, *Cell*, 1989, **58**, 823; R. L. Davis, P.-F. Cheng, A. B. Lasser and H. Wintraub, *Cell*, 1990, **60**, 773; X.-H. Sun and D. Baltimore, *Cell*, 1991, **64**, 459; S. J. Anthony-Cahill, P. A. Benfield, R. Fairman, Z. R. Wasserman, S. L. Brenner, W. F. Stafford, III, C. Altenbach, W. Hubbell and W. F. Degrado, *Science*, 1992, **255**, 979.

6 K. Hottori, M. Miyata and H. Yamamoto, *J. Am. Chem. Soc.*, 1993, **115**, 1151.