Scission of DNA at a preselected sequence using a single-strand-specific chemical nuclease

Chi-hong B Chen[1], Ralf Landgraf[1], Avram D Walts[1], Lisa Chan[1], Petra M Schlonk[2], Thomas C Terwilliger[3] and David S Sigman[1]

Background: We were interested in developing a protocol for cleaving large DNAs specifically. Previous attempts to develop such methods have failed to work because of high levels of nonspecific background scission.

Results: R-loop formation was chosen for sequence-specific targeting, a method of hybridization whereby an RNA displaces a DNA strand of identical sequence in 70% formamide using Watson–Crick base-pairing, leading to a three-stranded structure. R-loops are stabilized in aqueous solution by modifying the bases with chemical reagents. The R-loop was cleaved using a novel nuclease prepared from the Thr48→Cys mutant of the single-strand-specific M-13 gene V protein (GVP), which was alkylated with 5-(iodoacetamido-β-alanyl)1,10-phenanthroline. The cleavage products of the pGEM plasmid were cloned into the pCR 2.1–TOPO vector. Adenovirus 2 DNA (35.8 kb; tenfold larger than the pGEM plasmid) was also cleaved quantitatively at a preselected sequence.

Conclusions: A new method for cleaving duplex DNA at any preselected sequence was developed. The cleavage method relies on the chemical conversion of M-13 GVP into a nuclease, reflecting GVP's specificity for single-stranded DNA. The GVP chimera is the first example of a semisynthetic secondary structure specific nuclease. The chemical nuclease activity of 1,10-phenanthroline-copper is uniquely suited to this technique because it oxidizes the deoxyribose moiety without generating diffusible intermediates, providing clonable DNA fragments. The protocol could be useful in generating large DNA fragments for mapping the contiguity of probes or defining the exon–intron structure of transcription units.

Addresses: [1]Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095–1570, USA. [2]Department of Biochemistry, University of Chicago, Chicago, IL 60637, USA. [3]Life Sciences Division, Los Alamos National Laboratory, Mail Stop M880, Los Alamos, NM 87545, USA.

Correspondence: David S Sigman
E-mail: sigman@mbi.ucla.edu

Key words: chemical nuclease, clonable fragments, gene V protein-1–10-phenanthroline, 1,10-phenanthroline-copper, sequence-dictated DNA scission

Received: **23 February 1998**
Revisions requested: **19 March 1998**
Revisions received: **3 April 1998**
Accepted: **8 April 1998**

Published: **9 May 1998**

Chemistry & Biology May 1998, **5**:283–292
http://biomednet.com/elecref/1074552100500283

Introduction

The starting point in devising a general method for the sequence-specific cleavage of DNA requires the choice of a stringent procedure for sequence recognition. We chose R-loop formation, a method of hybridization in which an RNA displaces a DNA strand of identical sequence in 70% formamide using Watson–Crick base-pairing and leads to a three-stranded structure [1,2]. The dependence of hybridization on the length and structure of the RNA has been studied, revealing that RNAs as short as 44 nucleotides (nts) form R-loops quantitatively if the uracil of the RNA is replaced by 5-allylamine-uracil [3]. For RNAs prepared from the four predominant bases, chain lengths of 100 nts are sufficient for quantitative formation of R-loops [3]. Kinetic studies of the displacement of RNA from R-loops in aqueous solution have shown that the average step time for the reversible exchange of a single nucleotide by DNA is 345 milliseconds/step. The reaction is enhanced by ligands (e.g., magnesium and distamycin) that stabilize the B-DNA structure [3,4].

Initially, the targeted scission of the R-loops was carried out using 1,10-phenanthroline-copper-derivatized oligonucleotides in 70% formamide [5]. More recently, the three-stranded structure was stabilized by reaction in aqueous solution with single-strand-specific chemical modification reagents such as glyoxal or kethoxal. Following hydrolysis of RNA by RNaseH or RNaseA, the backbone of the stabilized single-stranded DNA bubble was cleaved hydrolytically by single-strand-specific nucleases, such as genetically attenuated micrococcal nuclease (MN). These improvements led to the mapping of the exon–intron boundaries of the transcription unit of human isoaspartyl/D-aspartyl *O*-methyl transferase analyzed as a 100 kilobase (kb) P-1 clone [3]. Background scission of large duplex DNAs by MN limited the robustness of this protocol, however.

Because the critical problem with using an R-loop-dependent strategy was the method of cleaving the phosphodiester backbone, we devised a more selective single-stranded nuclease. To this end, we converted the

single-strand-specific DNA-binding protein of M-13, the gene V protein (GVP), into a DNA cleavage reagent. GVP is well suited as the targeting element of this DNA cleavage reagent because it binds with high cooperativity to single-stranded nucleic acids and has low affinity for double-stranded DNA [6,7]. As a result, many molecules of GVP should bind to the single-stranded nucleic-acid bubble generated by R-loop formation, and efficient cleavage at the targeted site might be observed. The three-dimensional structures of GVP determined using both crystallography [6,7] and nuclear magnetic resonance (NMR) [8,9], as well as site-directed mutagenesis [7], served as a guide for the identification of positions for 1,10-phenanthroline (OP) attachment that were close enough to the bound single-stranded DNA for efficient cleavage, yet did not interfere with the protein's cooperative binding to the DNA.

Our laboratory has developed a general method for converting DNA-binding proteins into scission reagents [10–13]. It involves inserting cysteine residues by site-directed mutagenesis in the DNA-binding protein at sequence positions accessible to the bound nucleic acid and then alkylating these positions with an iodoacetamido derivative of 1,10-phenanthroline (e.g., 5-(iodoacetamido-β-alanyl)1,10-phenanthroline, IAAOP). The chemical-nuclease activity of the protein-linked 1,10-phenanthroline-copper is then activated by the addition of copper ions and a reducing agent. Cleavage is observed at the sites of DNA that correspond to the binding specificity of the targeting proteins. In the present protocol, the sites of scission correspond to the sequence of the RNA used to form the R-loop and do not reflect any sequence preference of the OP-derivatized single-strand-specific GVP protein.

Here we report that GVP Thr48→Cys (T48C) derivatized with IAAOP is an efficient single-stranded nuclease. It cleaves the unwound regions of DNA generated by the chemical modification of R-loops at preselected sequences (Figure 1). A key feature of this protocol is that the products generated can be cloned into pCR2.1–TOPO following treatment with Taq polymerase, a DNA polymerase that adds a single nontemplate-directed adenosyl residue on the 3′ end.

Results

Design strategy for the R-loop formation protocol

Figure 1 summarizes the protocol for the double-stranded scission of DNA at a preselected site that leads to clonable products. Generally RNAs of about 100 nts were used to target a specific site for scission via R-loop formation, although RNAs as short as 44 nts can be used if they are synthesized with 5-allylamine-uridines [4]. The method of R-loop formation was the same as those in our studies using high K_m mutations of MN for hydrolysis of the phosphodiester backbone of the single-stranded DNA [14]. Stabilization of R-loops via reaction with glyoxal or kethoxal led to equivalent results. Random cleavage of low melting A–T rich regions is not induced by glyoxal, which lacks the strict specificity of kethoxal for guanines [15,16]. These results demonstrate that the sites of DNA scission are dictated solely by the sequence of the RNA used for R-loop formation.

DNA binding and scission competence of GVP chimeras

The success of this protocol relies on the specificity and cooperative binding of GVP to single-stranded, but not duplex, DNA. In the absence of this positive cooperativity, the GVP–OP chimera would suffer the same restrictions as MN, resulting in random nicks at untargeted sites within high molecular weight DNA. We verified the quantitative and cooperative formation of a single-stranded DNA complex with one of the mutants (GVP T48C) using a gel-retardation assay. Evidence for the cooperative binding of the T48C mutant to single-stranded DNA is provided in Figure 2a, which shows a sharp transition from the largely free DNA at 0.6 μM to a completely bound protein–DNA complex at 1.8 μM. In Figure 2b, the strong preference of GVP T48C–OP for single-stranded DNA is presented. Under conditions when single-stranded DNA lac80 is completely bound, the duplex lacUV5 promoter does not form significant amounts of a protein–DNA complex.

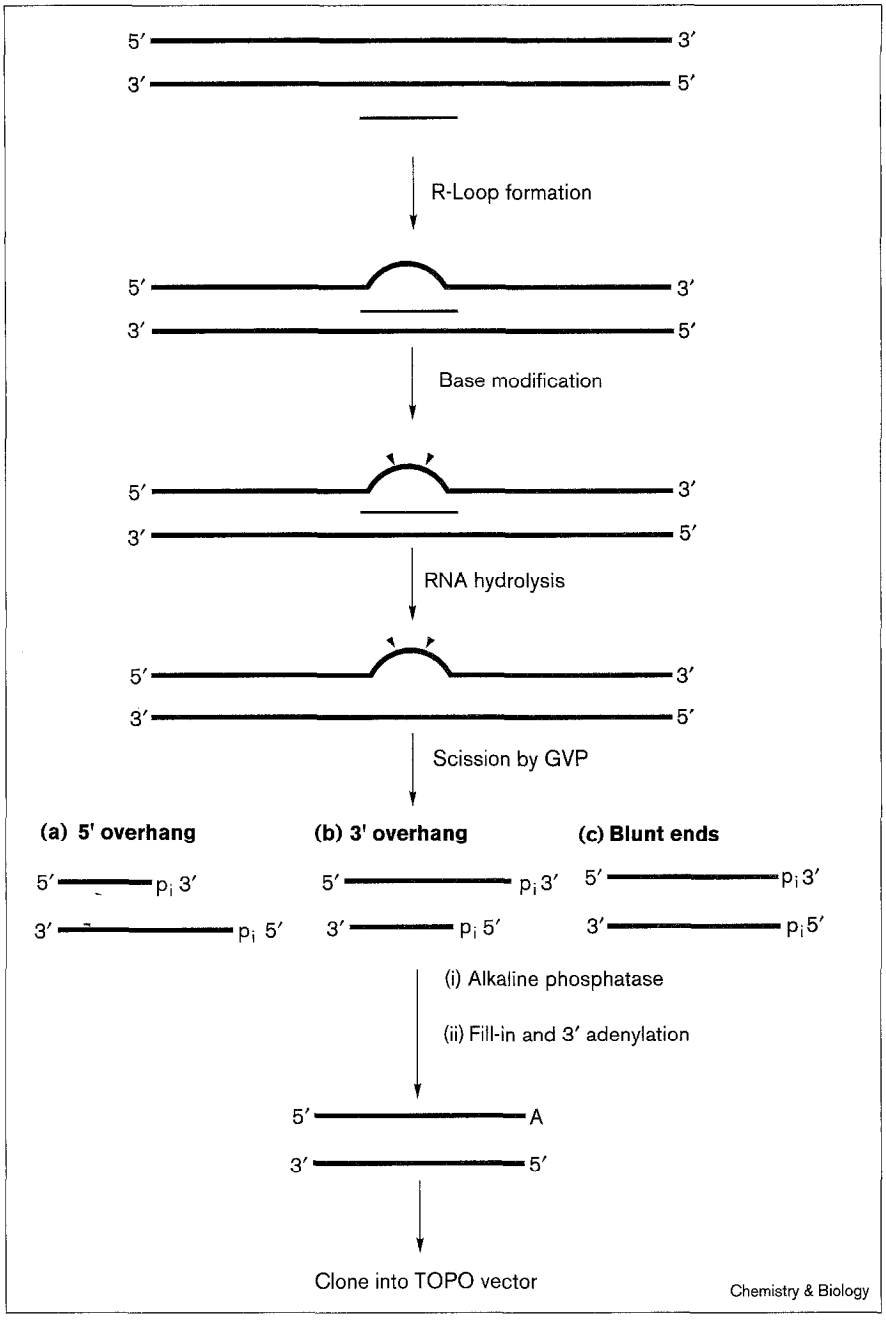
The selection of the candidate amino-acid sequence positions for the site-specific insertion of a cysteine residue into GVP was based on previous studies of the GVP DNA-binding domain using spin-labeled DNA [6], systematic mutagenesis studies of the GVP [7] and the structure of GVP solved to 1.8 Å resolution [6]. The cysteines of GVP mutants, Asp36→Cys (D36C), Gln40→Cys (E40C), Phe13→Cys (F13C), T48C and Val45→Cys (V45C), which varied in their rates of modification with 4,4′-dithiopyridine (Table 1), were reacted with IAAOP. Following exhaustive removal of unreacted IAAOP, the chimeras that retained high-affinity binding to single-stranded DNA were those that had been mutagenized at positions 13, 45 and 48 (Figure 2c). Of these, the only chimera that quantitatively cleaved single-stranded DNA to short oligonucleotides was GVP T48C–OP (Figure 2d). The extent of the substitution of GVP T48C–OP by IAAOP was spectroscopically quantified using an extinction coefficient of 27,000 $M^{-1}cm^{-1}$ at 270 nM for the OP moiety [17] and the Bradford protein assay for GVP. The yield of the alkylation reaction proved to be 89%.

Sequence-specific scission of pGEM–820 using sequence-dictated cleavage and GVP T48C–OP

The competence of the GVP T48C–OP as a single-stranded DNA cleavage agent prompted us to test it as the single-strand-specific nucleolytic agent in the protocol summarized by Figure 1 using the pGEM–820 assay previously employed [4]. pGEM–820 contains promoters for

Figure 1

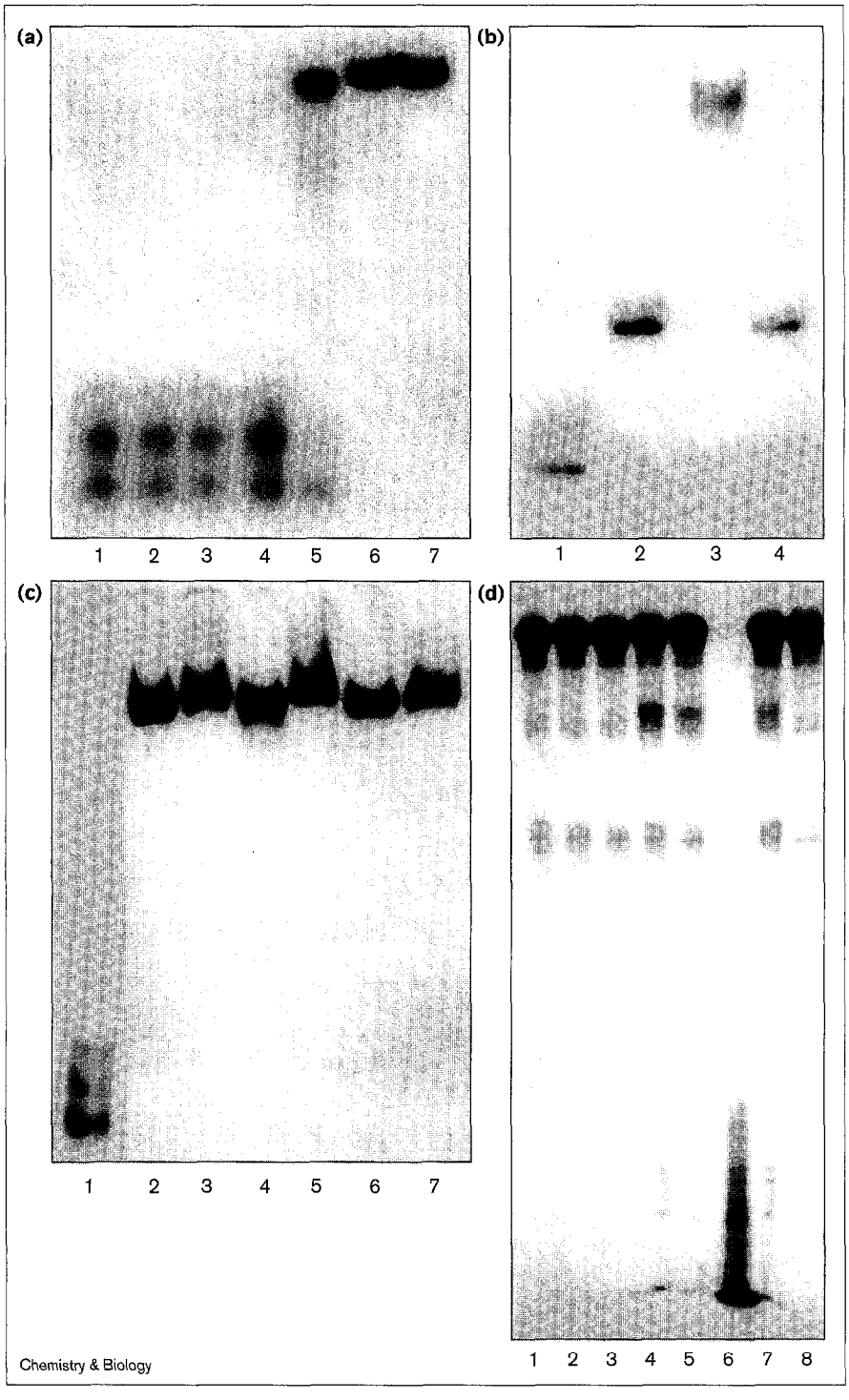
The protocol for sequence-dictated double-stranded cleavage of DNA. A piece of single-stranded RNA (red) is mixed with the double-stranded DNA (green) of interest, generating a R-loop. Specific bases are then modified using kethoxal or glyoxal. Next the RNA is hydrolyzed, leaving the unwound DNA at the site of the R-loop. Addition of GVP T48C–OP leads to one of three DNA scission products, **(a)** 5′ overhang, **(b)** 3′ overhang or **(c)** blunt ends. In order to clone these fragments, the terminal phosphates are removed using alkaline phosphatase, and primer extended using Taq polymerase to generate 5′ overhang products. These products can then be cloned into the pCR2.1–TOPO vector.



both T-7 and SP-6 RNA polymerases, and can be used to express RNA from either strand. In the test system, an RNA that overlapped a *Hind*III restriction site was used for two reasons. First, hydrolysis of the linearized plasmid by *Hind*III provides a convenient molecular weight marker for the anticipated products. Second, formation of R-loops can be assayed independently by their inhibition of *Hind*III digestion [3,4]. R-loops formed within *Bgl*I linearized pGEM–820 were stabilized by chemically modifying the displaced strand using glyoxal. The RNA used for the sequence-specific formation of an R-loop was hydrolyzed by RNaseA. The chemically stabilized single-stranded DNA bubble was then cleaved using GVP T48C–OP (Figure 3). High-yield cleavage precisely at the targeted site was observed. The products were identical in size to those observed when mung bean nuclease and attenuated, high K_m MN were used [3,4].

A stringent test for the specificity of the GVP T48C–OP in cleaving single-stranded regions is provided by the lack of any nonspecific cleavage of pGEM–820 in the absence of R-loop formation (Figure 3a, lane 2), a result

Figure 2



Interaction of GVP–OP chimeras with 5′ ^{32}P single-stranded DNA. **(a)** Concentration dependence of complex formation of 5′ ^{32}P single-stranded DNA with GVP T48C. Concentrations of GVP T48C: lane 1, 0 μM; lane 2, 0.2 μM; lane 3, 0.3 μM; lane 4, 0.6 μM; lane 5, 1.6 μM; lane 6, 3.1 μM; lane 7, 6.2 μM. **(b)** Relative affinities of GVP T48C–OP for single-stranded DNA (lac80) and double-stranded DNA (lacUV-5). Lane 1, lac80 (single-stranded DNA); lane 2, lacUV-5 (double-stranded DNA); lane 3, lac80 + T48C–OP; lane 4, lacUV-5 + T48C–OP. **(c)** Gel-retardation analysis of GVP–OPs with 5′ ^{32}P single-stranded DNA. Lane 1, DNA only; lane 2, GVP F13C–OP 5 μM; lane 3 GVP F13C–OP 10 μM; lane 4, GVP T48C–OP 5 μM; lane 5, GVP T48C–OP 10 μM; lane 6, GVP V45C–OP 5 μM; lane 7, GVP V45C–OP 10 μM. **(d)** Scission of 5′ ^{32}P single-stranded DNA by GVP–OP chimeras. lane 1, buffer + MPA + Cu^{+2} ; lane 2, control using underivatized GVP V45C; lanes 3–7 chimeric GVP's. Lane 3, GVP D36C–OP; lane 4, GVP E40C–OP; lane 5, GVP F13C–OP; lane 6, GVP T48C–OP; lane 7, GVP V45C–OP; lane 8, DNA only. Concentration of different GVPs, 15 μM; MPA, 2.9 mM; Cu^{+2}, 25 μM; Mg^{+2}, 50 μM (where appropriate).

anticipated by the lack of affinity of GVP T48C–OP for duplex DNA (Figure 2b). The absence of background scission by free GVP T48C–OP is consistent with our studies which indicate that binding of the chelate adjacent to its nucleic-acid target is central to its cleavage specificity, and that the reaction does not proceed by diffusible hydroxyl radicals [11,18].

Cloning of the cleaved products

Facile cloning of the reaction products is essential for the utility of the protocol. The method of cloning used must take into consideration the underlying chemical mechanism of the scission reaction, which is initiated by abstraction of the C-1′ hydrogen of the deoxyribose moiety. Removal of this hydrogen leads to four stable

Table 1

Reactivities of cysteine residues in GVP mutants and solvent accessibilities of wild-type residues at these positions.

Residue (mutated to Cys)	Solvent accessibility (%)	k ($M^{-1}sec^{-1}$)
Phe13	38	141
Asp36	31	387
Glu40	86	244
Val45	0	2
Thr48	17	262
Cys33	0	3

products in the scission reaction: 3′ and 5′ phosphorylated termini, free bases and 5-methylene-furanone [19–22]. The source of the C-1 oxygen in the 5-methylene furanone is the solvent water [19].

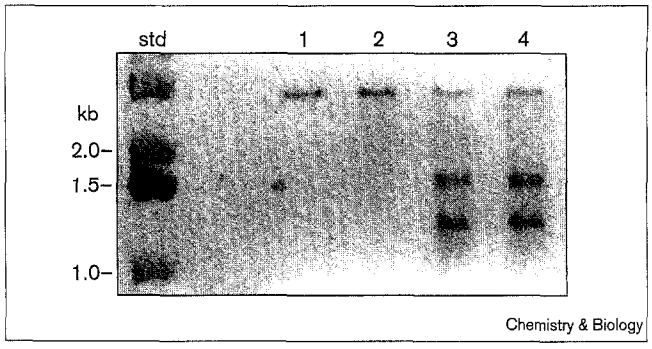
There are two steps in our protocol for cloning the products of sequence-dictated scission (Figure 1). The first is the removal of the 3′ phosphate group using calf intestinal alkaline phosphatase and end-filling with Taq polymerase, which leaves a nontemplate 3′ deoxyadenosine overhang (Figure 1). The second step is the cloning of the products containing an extra adenosyl moiety into the commercially available pCR2.1–TOP0 plasmid, containing a single thymidine overhang, prepared with vaccinia DNA topoisomerase I [23,24].

The successful cloning of these products indicates that the protocol maintains the biological integrity of the cleaved DNA, despite the oxidative chemistry used for the scission of the phosphodiester backbone. The product obtained from the pCR2.1–TOPO clones before and after cleavage with *Eco*RI, whose sites bracket the insertion site, are presented in Figure 4a,b. The cloning of individual scission products permits analysis of the distribution of cleavage sites within the unwound region of DNA generated by the formation of the R-loops. Five pCR2.1–TOPO clones have been retrieved and sequenced. Each clone provides only one of the products generated from the linearized target DNA. The termini obtained indicate no definitive preference for the site of cleavage within the unwound region (Figure 4c).

Scission of adenovirus 2 DNA using GVP T48C–OP

Because background scission by MN and mung bean nuclease is a limiting factor in the sequence-dictated cleavage of high molecular weight DNAs, we tested the utility of GVP T48C–OP as a single-stranded nuclease in the targeted scission of adenovirus 2 DNA. With a molecular weight of 35.9 kB, this DNA approaches the size of DNAs that could be profitably analyzed using sequence-dictated cleavage. It is small enough, however, to test the applicability of the sequence-dictated cleavage protocol using GVP T48C–OP in the absence of experimental problems associated with the handling of much larger DNAs.

Figure 3



Cleavage of pGEM–820 DNA using GVP T48C–OP. pGEM–820 (1 pmol) was linearized with *Bgl*I. R-loops were prepared using 50 pmol of an RNA that overlapped a unique *Hind*III site. Following chemical modification of the displaced strand with glyoxal or kethoxal, the RNA was hydrolyzed and incubated with GVP T48C–OP. Lane 1, control with no protein or RNA; lane 2, no RNA but containing GVP T48C–OP, Cu^{2+}, MPA; lane 3, 2 h cleavage at room temperature; lane 4, 3 h cleavage at room temperature.

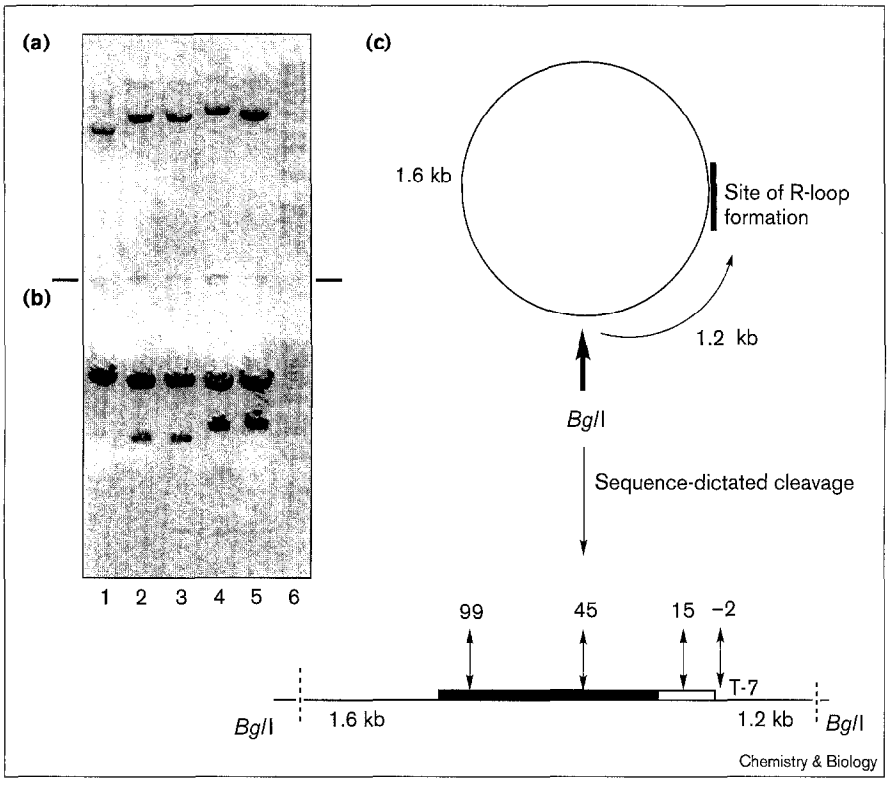
The map of the 35,937 base pair adenovirus 2 DNA is summarized in Figure 5 [25,26]. We chose to cleave adenovirus 2 DNA at a position that overlapped a restriction site for *Nde*I. Because R-loops inhibit restriction-enzyme cleavage, their formation could therefore be assayed independently of scission by GVP T48C–OP, if necessary [4]. RNA used for targeting the scission was generated from adenovirus 2 DNA by PCR using one primer that incorporated the promoter of T-7 RNA polymerase. The anticipated size of the cleavage products would be approximately 20.6 and 15.3 kb.

The single-stranded bubble essential for sequence-dictated cleavage was prepared in solution as described above in our studies with pGEM–820. RNA complementary to the lower strand was synthesized and used to prepare an R-loop in 70% formamide. The DNA strand that was displaced upon R-loop formation, in this case the upper strand, was modified by glyoxal. The RNA used for the sequence-selective hybridization was then hydrolyzed by RNaseA so that the upper strand corresponding to the sequence of the RNA was glyoxylated and the lower strand, which had been hybridized by the RNA, was single-stranded but not chemically modified. GVP T48C–OP was then added and cleavage was activated with cupric ion and MPA. Scission of the adenovirus 2 DNA at the targeted sequence with high yield was observed (Figure 5b). In contrast to previous work [27], Southern blotting was not essential to visualize the cleavage; nonspecific scission was so slight that the products could be observed on ethidium-bromide-stained gels.

Discussion

Because R-loop formation exploits the universal code of Watson–Crick base pairing, R-loops can be used to target

Figure 4

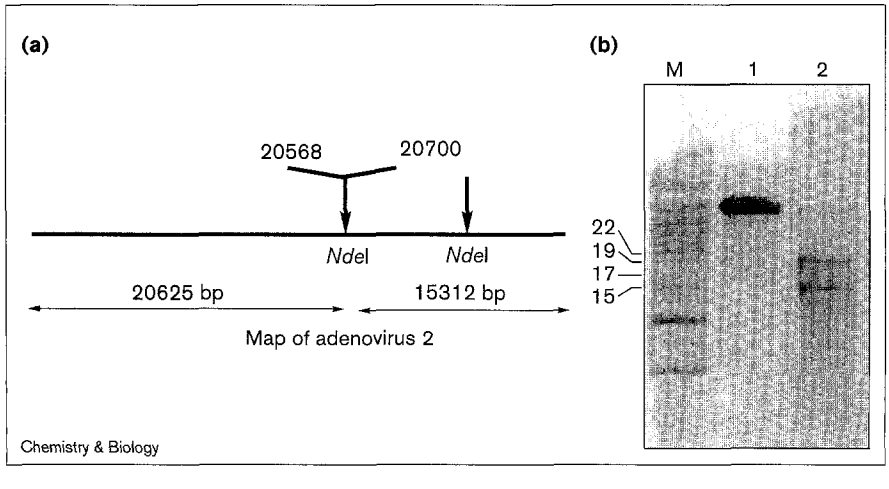


Cloning of cleaved products in pCR2.1–TOP0. **(a)** pCR2.1–TOPO plasmids containing cloned fragments of pGEM–820 DNA. Cleaved fragments of pGEM–820 were purified from agarose gel, dephosphorylated with alkaline phosphatase and then end-filled using Taq polymerase. The products were mixed with pCR 2.1–TOPO and then used to transform competent cells provided by the manufacturer. The lanes correspond to distinct colonies, except for lane 6, which corresponds to size markers. Plasmids were isolated from mini-preps, and then subjected to non-denaturing electrophoresis (0.8% agarose). **(b)** *Eco*RI digest fragments of pCR2.1–TOP0 plasmids. Plasmids of (a) were digested using *Eco*RI and the fragments were separated by nondenaturing electrophoresis (0.8% agarose). **(c)** Sequence summary of cloned inserts. Map of pGEM–820 with indicated *Bgl*I site and R-loop target. Vertical double-headed arrows correspond to sites of cleavage within the DNA insert downstream from the T-7 promoter also used for preparing the targeting RNA. Products were sequenced using the M13 reverse and forward primers. Open bar represents RNA not derived from lac.

any region of the genome for double-stranded scission [28]. No unique sequences at a particular locus, such as a restriction site [29] or polypurine/polypyridimine sequences susceptible to triple-helix formation [30], are required for directing the scission. Our prior reliance on single-strand-specific hydrolytic nucleases, such as mung bean nuclease or high K_m mutants of MN for scission of the single-strand DNA bubble, resulted in significant nonspecific cleavage of duplex DNA [14]. Although the method worked successfully with small DNAs, variable background cleavage of high molecular weight DNAs made the method very sensitive to small changes in experimental conditions and limited its effectiveness.

Crucial for the success of the protocol reported in this study was the improvement phosphodiester backbone scission method. Our solution has been to reengineer the single-stranded-binding protein of M-13 (GVP), and convert it

Figure 5



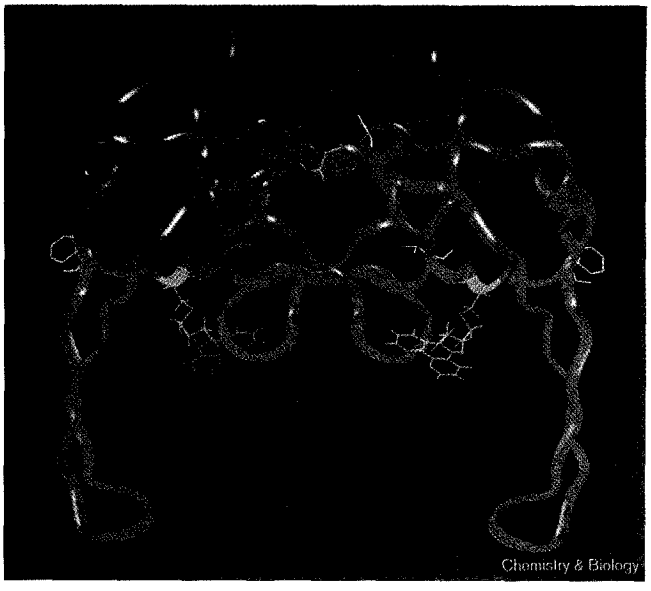
Scission of adenovirus 2 DNA by sequence-dictated cleavage. **(a)**. Adenovirus map used for sequence-dictated cleavage. **(b)** Isolation of products. A 132 nt RNA fragment was used to form an R-loop overlapping an *Nde*I restriction site at position 20624. The R-loop was stabilized by modification with glyoxal. Subsequent to removal of the RNA, the unwound DNA was incubated with GVP T48C–OP for 30 min in Tris buffer, pH 8.0. The cleavage reaction was initiated by adding $CuSO_4$ and MPA. Reaction time was 2 h at room temperature; electrophoresis was carried out for 16 h, 1% agarose 0.1 to 2.5 s switching at 200 volts. Lane M, size markers; lane 1, all reaction components were added except GVP T48C–OP; Lane 2, products of sequence-dictated double-stranded DNA cleavage.

into a nuclease, using a procedure that we have now applied to a series of DNA-binding proteins [13,31,32]. The GVP protein is the first secondary-structure-specific protein converted into a nuclease. By converting GVP into a scission reagent, we have exploited the cooperative binding of the protein to single-stranded regions of DNA [6], and generated a nucleolytic agent that is exceptionally single-strand specific. Because the protein–protein interactions of GVP greatly favor its interaction with the single-stranded regions generated by R-loop formation, little background scission is observed. The development of a single-strand-specific nuclease based on GVP requires a chemical nuclease activity like 1,10-phenathroline copper. This redox-active coordination complex does not generate diffusible intermediates like ferrous–EDTA, which reacts via hydroxyl radicals. Instead, modified GVP cleaves DNA via a copper–oxo intermediate that must be bound to the target DNA. The practical consequence of this intrinsic difference in reactivity is that it is not necessary to suppress background scission of target DNA by adding hydroxyl radical traps, such as salmon sperm or calf thymus DNA, that would block subsequent manipulations of the cleaved DNA.

All of the GVP mutants contain Cys33, which is present in the wild-type protein and not accessible to modification reagents. Amino-acid residues at five sequence positions have been converted to cysteines, including one that has been characterized as buried (Val45), one near the postulated oligonucleotide-binding site (Thr48), and three surface sites distal from the bound oligonucleotide (Phe13, Asp36 and Gln30). All these sequence positions are tolerant to a range of amino-acid substitutions [7]. The threonine residue at position 48 of GVP can be changed into a cysteine or a valine residue by mutagenesis and retain full activity. If a glycine or proline residue replaces the threonine residue, however, the protein is no longer active. Of the five mutants studied, GVP T48C was the only one that yielded a scission-competent chimera when alkylated by IAAOP. Our studies support the previous designation of this amino-acid position as near the bound oligonucleotide [8,9,33] (Figure 6). A similar chemical approach could be used to determine if other neighboring residues (e.g., Arg16, Arg21, Gln22, Lys24, Glu30, Lys46, Glu51, Gln72 and Arg80) are near the single-stranded DNA-binding site [7]. The lack of observable R-loop-independent background scission by GVP T48C–OP, as well as the stability of the duplex products of scission, underscores the specificity of this cleavage reaction for single-stranded regions.

The GVP T48C–OP scission of the single-stranded bubble generated by the chemical modification of the R-loop results in the loss of at least 95% of the parent band. This high yield of cleavage is remarkable for targeted scission reactions [12]. Although the Trp-repressor–OP chimera cleaves tryptophan-regulated operators

Figure 6



A ribbon diagram of the crystal structure of the GVP indicating the sites of cysteine substitution. Color key: Phe13, white; Asp36, yellow; Gln40, purple; Val45, blue; Thr48, green (drawn with OP attached) Thr48 remains represented in the structure.

with 70–80% efficiency [11,34], scission yields of the Fis–OP chimera rarely exceed 30%. The Fis–OP constructs react with all the known sites of Fis binding and have been used to identify previously unknown sites of this accessory protein [31,32,35].

The most likely explanation for the high cleavage efficiency of the GVP T48C–OP protein is that it is a consequence of the cooperative binding of the protein to the single-stranded bubble generated by R-loop formation. Detailed models for wild-type GVP binding to single-stranded DNA indicate that dimers assemble on a DNA strand in a left-handed helical array with a repeat distance of about 15 Å [6]. This distance can be spanned by four nucleotides in single-stranded DNA in a partially extended conformation. One possible explanation for the virtually quantitative yield of scission is that approximately 25 dimers or 50 tethered 1:1 1,10-phenanthroline-copper chelates would be present on a single-stranded DNA created within an R-loop by a 100 nt RNA. Even if the cleavage efficiency of any single chimera is only 5–10%, the likelihood of quantitative cleavage at these preselected sites of R-loop formation would be high.

Our ability to clone the fragments of the oxidative cleavage reaction is important in two contexts. Most significantly, we have learned that it will be possible to retrieve the DNA products for use in a range of applications. The upper limit on the size of the fragments that can be cloned using the protocol outlined here has not been determined

and must be established by further experimentation. DNAs at least as large as 10 kb can presently be inserted with greater than 70% cloning efficiency, however. The cleavage of nucleic acids by hydrolysis, although a challenging chemical problem, is not necessary for DNA manipulations because oxidative scission provides clonable products [36–38]. In addition, the cloning of the fragments has permitted an examination of the cleavage results at a sequence level of resolution. Although an extensive number of clones derived from products with 5′ overhangs have not been sequenced, our results suggest that the cleavage sites are distributed within the unwound DNA. Although the cleavage efficiency of GVP T48C-OP on isolated single-stranded DNA is high, no method is available for estimating the effectiveness of the cleavage reaction within the unwound DNA. Because the targeted scission is effectively quantitative, improvement of the cleavage efficiency of the GVP chimera, which might lead to enhanced background scission, is not warranted.

Cleavage of DNA at any preselected sequence has a variety of possible applications in the analysis of genomic DNA. These include mapping sites of chromosomal translocations [39,40], measuring the distance between two marker sequences in units of base pairs [41], and generating large DNA fragments for mapping the contiguity of probes using pulse-field gel electrophoresis [42]. In addition, this type of cleavage would provide a method for isolating a DNA fragment from a YAC or genomic DNA greater than 30kb, which is the maximum length accessible by PCR, to facilitate the search for a genetic locus, to characterize a transcription unit by exon trapping [43–45] or to use as a starting point for Genome Mismatch Scanning [46].

Significance

Using chemical and molecular biological techniques, we have developed a robust protocol that permits the scission of DNA at any preselected sequence targeted by the formation of R-loops. The methodology relies on the efficient and single-strand-specific scission activity of GVP T48C–OP, a modified version of the virus M-13 gene V protein (GVP). The reaction preference of GVP T48C–OP for extended regions of single-stranded DNA relative to single-stranded nicks depends on its cooperative binding to unwound DNA. Products are clonable in *Escherichia coli* following end-filling with Taq polymerase. The new protocol could prove useful in measuring the distance between two marker sequences in units of base pairs, generating large DNA fragments for mapping the contiguity of probes or defining the exon–intron structure of transcription units.

Materials and methods

Synthesis of RNAs

lac RNA was synthesized using pGEM–820, which contains the 5′-region of the *E. coli* lacZ gene from position −19 upstream to position +63 downstream of a T7 promoter as described earlier [47]. pGEM–820 was first restriction-enzyme digested with *Eco*RI and then transcribed using the Promega Riboprobe system.

For adenovirus 2 RNA, two deoxyoligonucleotides were synthesized as primers for the PCR reaction using a Pharmacia DNA synthesizer. One corresponded to sequence positions 20568 to 20587 of adenovirus 2 DNA (5′-GGTGCCCCAAAAGTTTTTTG-3′) and was linked synthetically 3′ to the T-7 polymerase promoter (5′-TAATACGACTCACTATA-3′). The second corresponded to sequence positions 20681 to 20700 of adenovirus 2 DNA (5′-CCCGTCAACTCTAAGATCGT-3′). The resulting 132 nt DNA fragment was used as the template for *in vitro* transcription to prepare RNA from sequence position 20568 to 20700 of adenovirus 2. The DNA fragment obtained contained a *Nde*I site at position 20624, and was purified on a 2% agarose gel. *In vitro* transcription was performed using the Promega Riboprobe system as described above.

GVP Mutants

The five GVP mutants, D36C, E40C, F13C, T48C and V45C were constructed by oligonucleotide-directed mutagenesis of the plasmid pTT18, which encodes wild-type GVP as described earlier [48]. The proteins were purified essentially the same way as the wild-type protein except that 1–5 mM β–mercaptoethanol or 1–5 mM dithiothreitol (DTT) was included in all buffers [49]. All mutants retained a nonreactive cysteine residue at position 33.

Derivatization of GVP with IAAOP. The reactivity of the cysteines of the GVP mutants was assayed using 4,4′-dithiopyridine (PDS) and correlated with solvent accessibility (Table 1) [49,50]. Each protein (4 μM) was reacted with 80 μM PDS in 50 mM MOPS, 100 mM NaCl, 1 mM EDTA, pH 7.0. The initial rate of the reaction was monitored by detecting the absorbance of the product 4-thiopyridine at 324 nM using an extinction coefficient of $1.98 \times 10^4\ cm^{-1}M^{-1}$ [51]. The reaction rate was assumed to be the product of the second-order reaction constant (k) and the concentrations of protein and PDS. Although wild-type GVP has a single cysteine at position 33, GVP reacts with PDS at the same rate ($k_{app} = k(PDS) = 3\ M^{-1}s^{-1}$) as a mutant in which Cys33 has been replaced by an alanine ($k_{app} = 5\ M^{-1}s^{-1}$). These findings indicate that Cys33 is solvent inaccessible.

Prior to the addition of IAAOP (20 mM in DMF), the cysteine mutants were reduced with dithiothreitol (DTT) on ice for two hours. A typical reaction mixture contained 60 μM protein and 6 mM IAAOP in 50 mM MOPS/100 mM NaCl/ 1 mM EDTA/pH 7.0/1 mM DTT and was carried out at 4°C overnight. The excess IAAOP was removed by either passing samples through a G-25 spin column twice or using a Centricon column with at least five volumes of dialysate (50 mM Tris, pH 8.0,100 mM, NaCl).

Preparation of the lacZ *DNA.* A synthetic 80 base-pair fragment of the *E. coli* lacZ gene (template strand) was used as a source of single-stranded DNA. This was 5′ labeled using γ^{32}P-ATP and T4 polynucleotide kinase for 30 min at 37°C [52].

GVP gel-retardation assay. ^{32}P-*lacZ* single-stranded DNA (1 μl) was mixed with approximately 300 pmol protein (either derivatized or not) in a 50 mM Tris (pH 7.9), 13 mM NaCl buffer that contained 25 ng polydeoxyinosine–polydeoxycytosine per reaction mixture [53] in a final reaction mixture of 20 μl. Binding of GVP to the DNA was carried out at room temperature for 30 min. Glycerol (5 μl of 80%) was subsequently added to each sample and the entire 25 μl mixture was run on a 6% nondenaturing polyacrylamide gel containing 2% glycerol, using a running buffer composed of Tris–boric acid buffer (.045 M).

GVP–OP cleavage of single-stranded DNA. The same reaction mixture used for the gel retardation assay was used for the cleavage of DNA except for the addition of 1 μl of 1 mM $MgCl_2$, 1 μl of 500 μM $CuSO_4$, and 1 μl of 58 mM 3-mercaptopropionic acid (MPA). The cleavage reaction was allowed to proceed for 1 h at room temperature after which it was quenched by the addition of 1 μl of 40 mM 2,9-dimethyl-1,10-phenanthroline and the products analyzed on a sequencing gel.

Sequence-dictated cleavage: R-loop formation

Glyoxal stabilization. R-loops were prepared by mixing 2 μg of DNA (*Bgl*I restriction digested pGEM–820 or adenovirus 2 DNA) and 50–100 pmol of RNA (T7 *lacZ* RNA or adenovirus 2 RNA respectively) in 70% deionized formamide, 83 mM PIPES (pH 7.8), 400 mM NaCl and 20 mM EDTA. The total volume was 300 μl. The R-loop reaction mixture was brought to 55°C and gradually cooled down to 45°C over a period of 4 h in a thermocycler [5]. The resulting R-loop mixture was cooled to 4°C before glyoxal was added to a final concentration of 1 M. After reaction at 12°C for 2 h, the modified DNA was recovered by ethanol precipitation with the addition of glycogen to aid in the precipitation.

Kethoxal stabilization. The formation of R-loops were carried out as described previously. After the formation, 30 μl of kethoxal solution (37.0 mg/ml in 20% EtOH) was added to the R-loops mixture at 4°C. The final concentration of kethoxal was 3.7 mg/ml. This reaction then proceeded at 4°C for 2 h before ethanol precipitation. The pellet was resuspended in 50 μl of Tris 100 mM, pH 6.3. RNaseA (3 μl) was then added. After 30 min at 37°C and desalting using a G-50 spin column, the pH of the solution was adjusted to 8 by adding 2 μl of 1 M Tris–HCl (pH 8). Sodium borate was also added to a final concentration of 10 mM to stabilize the kethoxal–DNA adduct.

Scission of R-loops

Before activation of the scission reaction, the glyoxal- or kethoxal-modified R-loop and GVP T48C–OP were incubated at room temperature for 30 min in 100 mM Tris (pH 8.0)/100 mM NaCl. The cleavage reaction was initiated by adding 2 μl of 0.5 mM $CuSO_4$ and 1 μL of 58 mM MPA. The final volume was 20 μl. The reaction was carried out for 2 h at room temperature. Reaction products for the *lac* system were analyzed with 0.8% agarose whereas pulsed field gel electrophoresis was used (1% agarose gel, 0.1 to 2.5 switching at 200 volts for 16 h) for the analysis of the products of adenovirus 2 DNA cleavage. Both systems were visualized by ethidium bromide staining.

Cloning of cleaved fragments

Two products of R-loop-directed cleavage reactions of pGEM–820 were isolated from a 0.8% agarose gel. The slower moving fragment corresponded to a molecular size of 1.5 kb, whereas the faster fragment was approximately 1.1 kb. The products isolated from the cleavage reaction were then dephosphorylated using calf intestinal phosphatase at 37°C for 1 h. After two phenol/chloroform extractions, the mixture was desalted by passing it through G-50 spin columns and concentrated to small volumes on a speed vac.

To clone the fragment, the cleavage products were reacted with all four dNTPs (5 mM each) and Taq polymerase (1 unit) in a total volume of 50 μl of PCR buffer. The reaction was carried out at 72°C for 1 h to ensure the anomalous nontemplate-directed addition of a single adenosine at the 3′ end, characteristic of Taq polymerase.

After phenol/chloroform extraction and desalting by G-50 spin columns, the eluant was concentrated to dryness and redissolved in 4 μl H_2O. After gently mixing with 1 μl of pCR2.1–TOPO vector comprising a single 3′-thymidine overhang (Invitrogen), the mixture was incubated at room temperature for 5 min and then placed on ice. The transformation reaction was carried out following the protocol (One-Shot Transformation Reaction) provided by Invitrogen. Ten white colonies were picked and cultured overnight in LB medium containing ampicillin. Plasmid DNAs were obtained using standard mini-prep procedures. The plasmids were restriction digested with *Eco*RI and analyzed on 0.8% agarose gels.

Sequencing of cloned inserts

The inserts were sequenced on an ABI automated sequencer using the universal primers M13 Reverse 5′-GTC CTT TGT CGA TAC TG-3′ and M13 Forward 5′-GTA AAA CGA CGG CCA G-3′.

Acknowledgements

This research was supported by USPHS HG-00255 and USPHS GM-21199. We wish to thank Young-Moon Cho for the synthesis of IAAOP.

References

1. Thomas, M., White, R.L. and Davis, R.W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc. Natl Acad. Sci. USA* **73**, 2294-2298.
2. Rosbash, M., Blank, D., Fahrner, K., Hereford, L., Ricciardi, R., Roberts, B., Ruby, S. & Woolford, J. (1979). R-looping and structural gene identification of recombinant DNA. *Methods Enzymol.* **68**, 454-469.
3. Landgraf, R., Chen, C.B. & Sigman, D.S. (1995). Double-stranded scission of DNA directed through sequence-specific R-loop formation. *Nucleic Acids Res.* **23**, 3524-3530.
4. Landgraf, R., Chen, C.B. & Sigman, D.S. (1995). R-loop stability as a function of RNA structure and size. *Nucleic Acids Res.* **23**, 3516-3523.
5. Chen, C.-h.B., Gorin, M.B. & Sigman, D.S. (1993). Sequence-specific scission of DNA by the chemical nuclease activity of 1,10-phenanthroline-copper(I) targeted by RNA. *Proc. Natl Acad. Sci. USA* **90**, 4206-4210.
6. Skinner, M.M. *et al.*, & Terwilliger, T.C. (1994). Structure of the gene V protein of bacteriophage F1 determined by multiwavelength X-ray diffraction on the selenomethionyl protein. *Proc. Natl Acad. Sci. USA* **91**, 2071-2075.
7. Terwilliger, T.C., Zabin, H.B., Horvath, M.P., Sandberg, W. S. & Schlunk, P.M. (1994). *In vivo* characterization of mutants of the bacteriophage F1 gene V protein isolated by saturation mutagenesis. *J. Mol. Biol.* **236**, 556-571.
8. Folkers, P.J.M., Stassen, A.P.M., Vanduynhoven, J.P.M., Harmsen, B.J.M., Konings, R.N.H. & Hilbers, C..W. (1991). Characterization of wild-type and mutant M13 gene-V proteins by means of ^{1}H–NMR. *Eur. J. Biochem.* **200**, 139-148.
9. Folkers, P.J.M., Vanduynhoven, J.P.M., Jonker, A.J., Harmsen, B.J.M., Konings, R.N.H. & Hilbers, C.W. (1991). Sequence-specific ^{1}H–NMR assignment and secondary structure of the Tyr41→His mutant of the single-stranded DNA-binding protein, gene-V protein, encoded by the filamentous bacteriophage-M13. *Eur. J. Biochem.* **202**, 349-360.
10. Bruice, T.W., Wise, J., Rosser, D.S.E. & Sigman, D.S. (1991). Conversion of Lambda Phage Cro into an operator-specific nuclease. *J. Am. Chem. Soc.* **113**, 5446-5447.
11. Landgraf, R., Pan, C., Sutton, C., Pearson, L. & Sigman, D.S. (1996). Engineering of DNA-binding proteins into site-specific cutters: reactivity of Trp-repressor-1,10-phenanthroline chimera. *Protein Eng.* **9**, 603-610.
12. Sigman, D.S., Bruice, T.W., Mazumder, A. & Sutton, C.L. (1993). Targeted chemical nucleases. *Accounts Chem. Res.* **26**, 98-104.
13. Pan, C.Q., Landgraf, R. & Sigman, D.S. (1994). DNA-binding proteins as site-specific nucleases. *Mol. Micro.* **12**, 335-342.
14. Landgraf, R., Chen, C.-h.B. & Sigman, D.S. (1994). Oligonucleotide-directed nucleic acid scission by micrococcal nuclease. *Biochemistry* **33**, 10607-10615.
15. Noller, H.F. & Chaires, J.B. (1972). Functional modification of 16s ribosomal RNA by kethoxal. *Proc. Natl Acad. Sci. USA* **69**, 3115-3118.
16. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P. & Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 9109-9128.
17. Sigman, D.S. (1967). Interactions of substrates, inhibitors and coenzymes at the active site of horse liver alcohol dehydrogenase. *J. Biol. Chem.* **242**, 3815-3824.
18. Sigman, D.S. (1996). Site-specific oxidative scission of nucleic acids and proteins. Series C-Vol. 479, (Meunier, B., ed.), pp119-132, Kluwer Academic Publishers. Dordrecht/Boston/London.
19. Meijler, M., Zelenko, O. & Sigman, D.S. (1997). Chemical mechanism of DNA scission by 1,10-phenanthroline-copper. Carbonyl oxygen of 5-methylene furanone is derived from water. *J. Am. Chem. Soc.* **119**, 1135-1136.
20. Sigman, D.S. (1986). Nuclease activity of 1,10-phenanthroline-copper ion. *Accounts Chem. Res.* **19**, 180-186.
21. Goyne, T.E. & Sigman, D.S. (1987). Nuclease activity of 1,10-phenanthroline-copper ion. Chemistry of deoxyribose oxidation. *J. Am. Chem. Soc.* **109**, 2846-2848.
22. Pope, L.M., Reich, K.A., Graham, D.R. & Sigman, D.S. (1982). Products of DNA cleavage by the 1,10-phenanthroline-copper complex. inhibitors of *E. coli* DNA polymerase I. *J. Biol. Chem.* **257**, 12121-12128.
23. Shuman, S. (1994). Novel approach to molecular cloning and polynucleotide synthesis using Vaccinia DNA topoisomerase. *J. Biol. Chem.* **269**, 32678-84.

24. Mead, D.A., Pey, N.K., Herrnstadt, C., Marcil, R.A. & Smith, L.M. (1991). A universal method for the direct cloning of PCR-amplified nucleic acid. *Bio-Technology* **9**, 657-663.
25. Akusjarvi, G., Pettersson, U. and Roberts, R.J. (1986). Structure and function of the adenovirus-2 genome. (Doerfler, W., ed.) pp 52-96. Martinus Nijhoff Publishing. Boston/Dordrecht/Lancaster.
26. Roberts, R.J., Akusjarvi, G., Alestrom, P., Gelinas, R.E., Gingerase, T.R., Sciaky, D. & Petterson, U. (1986). A consensus sequence for the adenovirus-2 genome. (Doerfler, W., ed.) 1-52. Martinus Nijhoff Publishing. Boston/Dordrecht/Lancaster.
27. Chen, C.-h.B., Mazumder, A., Constant, J.-F. & Sigman, D.S. (1993). Nuclease activity of 1,10-phenanthroline-copper. New conjugates with low molecular weight targeting ligands. *Bioconj. Chem.* **4**, 69-77.
28. Sigman, D.S., Chen, C.-h.B. and Gorin, M.B. (1993). Sequence-specific scission of DNA by RNAs linked to a chemical nuclease. *Nature* **363**, 474-475.
29. Koob, M. & Szybalski, W. (1990). Cleaving yeast and *Escherichia coli* genomes at a single site. *Science* **250**, 271-273.
30. Strobel, S.A. & Dervan, P.B. (1991). Triple-helix protection in achilles heel protocol. *Nature* **350**, 172-174.
31. Pan, C.Q., Johnson, R.C. & Sigman, D.S. (1996). Identification of new Fis-binding sites by DNA scission with Fis-1,10-Phenanthroline-copper(I) chimeras. *Biochemistry* **35**, 4326-4333.
32. Pan, C.Q., Finkel, S.E., Cramton, S.E., Feng, J.A., Sigman, D.S. & Johnson, R.C. (1996). Variable structures of Fis–DNA complexes determined by flanking DNA–protein contacts. *J. Mol. Biol.* **264**, 675-695.
33. Folkers, P.J.M., Vanduynhoven, J.P.M., Vanlieshout, H.T.M., Harmsen, B.J.M., Vanboom, J.H., *et al.* & Hilbers, C.W. (1993). Exploring the DNA-binding domain of gene-V protein encoded by bacteriophage M13 with the aid of spin-labeled oligonucleotides in combination with ^{1}H–NMR. *Biochemistry* **32**, 9407-9416.
34. Sutton, C.L., Mazumder, A., Chen, C.H.B. & Sigman, D.S. (1993). Transforming the *Escherichia coli* Trp-repressor into a site-specific nuclease. *Biochemistry* **32**, 4225-4230.
35. Pan, C.Q., Feng, J., Finkel, S.E., Landgraf, R., Johnson, R. & Sigman, D.S. (1994). Structure of the *Escherichia coli* Fis protein–DNA complex probed by protein conjugated with 1,10-phenanthroline copper(I) complex. *Proc. Natl Acad. Sci. USA* **91**, 1721-1725.
36. Chin, J. (1991). Developing artificial hydrolytic metalloenzymes by a unified mechanistic approach. *Accounts. Chem. Res.* **24**, 145-152.
37. Sigman, D.S., Mazumder, A. & Perrin, D.M. (1993). Chemical nucleases. *Chem. Rev.* **93**, 2295-2316.
38. Hall, J., Husken, D. & Haner, R. (1996). Towards artificial ribonucleases: the sequence-specific cleavage of RNA in a duplex. *Nucleic Acids Res.* **24**, 3522-3526.
39. Denny, C.T., Shah, N.P., Odgen, S., Willman, C., McConnell, T., Crist, W., Carroll, A. & Witte, O.N. (1989). Localization of preferential sites of rearrangement with the BCR gene in Philadelphia chromosome-positive acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **86**, 4254-4258.
40. Rabbits, T.H. (1991). CLL chromosomal translocations. *Cell.* **67**, 641-648.
41. Smith, C.L., Econome, J.G., Schutt, A., Klco, S. & Cantor, C.R. (1987). A physical map of the *Escherichia coli* K12 genome. *Science* **236**, 1448-1453.
42. Smith, C.L. & Cantor, C.R. (1989). Pulsed-field gel electrophoresis and the technology of large DNA molecules. IRL Press Inc. McLean, VA.
43. Datson, N.A., Vandevosse, E., Dauwerse, H.G., Bout, M., Vanommen, G.J.B. & Dendunnen, J.T. (1996). Scanning for genes in large genomic regions – cosmid-based exon trapping of multiple exons in a single product. *Nucleic Acids Res.* **24**, 1105-1111.
44. Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A. & Housman, D.E. (1991). Exon amplification - a strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl Acad. Sci. USA* **88**, 4005-4009.
45. Church, D.M., Stotler, C.J., Rutter, J.L., Murrell, J.R., Trofatter, J.A. & Buckler, A.J. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nat. Genet.* **6**, 98-105.
46. Nelson, S.F., McCusker, J.H., Sander, M.A., Kee, Y., Modrich, P. & Brown, P.O. (1993). Genomic-mismatch scanning: a new approach to the genetic linkage mapping. *Nat. Genet.* **4**, 11-18.
47. Murakawa, G.J., Chen, C.-h.B., Kuwabara, M.D., Nierlich, D. & Sigman, D.S. (1989). Scission of RNA by the chemical nuclease 1,10-phenanthroline-copper ion. preference for single-stranded loops. *Nucleic Acids Res.* **17**, 5361-5369.
48. Terwilliger, T.C. (1988). Simple and highly efficient site-specific mutagenesis by ligation of an oligodeoxyribonucleotide into gapped heteroduplex DNA in which the template strand contains deoxyuridine. *Gene* **69**, 317-24.
49. Liang, H. & Terwilliger, T.C. (1991). Reversible denaturation of the gene V-protein of bacteriophage-F1. *Biochemistry* **30**, 2772-2782.
50. Shrake, A. & Rupley, J.A. (1973). Environment and exposure to solvent of protein atoms. lysozyme and insulin. *J. Mol. Biol.* **79**, 351-371.
51. Richards, F.M. (1977). Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151-176.
52. Mazumder, A., Chen, C.-h.B., Gaynor, R.B. & Sigman, D.S. (1992). 1,10-phenanthroline-copper, a footprinting reagent for single-stranded regions of RNAs. *Biochem. Biophys. Res. Commun.* **187**, 1503-1509.
53. Kuwabara, M.D. & Sigman, D.S. (1987). Footprinting DNA-protein complexes in situ following gel retardation assays using 1,10-phenanthroline-copper Ion: *Escherichia coli* RNA polymerase-*lac* promoter complexes. *Biochemistry* **26**, 7234-7238.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.