The Activity of the *Eco*RV Restriction Endonuclease Is Influenced by Flanking DNA Sequences both inside and outside the DNA–Protein Complex†

John D. Taylor‡ and Stephen E. Halford*
Department of Biochemistry, Centre for Molecular Recognition, University of Bristol, Bristol BS8 1TD, U.K.
Received July 9, 1991; Revised Manuscript Received September 23, 1991

ABSTRACT: The *Eco*RV restriction endonuclease cleaves DNA not only at its recognition sequence but also at most other sequences that differ from the recognition site by one base pair. Compared to the reaction at the recognition site, the reactions at noncognate sites are slow but 1 out of the 12 noncognate sites on the plasmid pAT153 is cleaved more than 50 times faster than any other. The increase in the reaction rate at the preferred noncognate site, relative to other sites, was caused by the DNA sequences in the 4 base pairs from either side of the site. For enhanced activity by *Eco*RV, particular bases were needed immediately adjacent to the site, inside the DNA–protein complex. At these loci, the protein interacts with the phosphate groups in the DNA and the flanking sequence may control the activity of the enzyme by determining the conformation of the DNA, thus aligning the phosphate contacts. But the preferential cleavage also depended on sequences further away from the site, at loci outside the complex. At external positions, beyond the reach of the protein, the *Eco*RV enzyme required flanking sequences that give rise to flexibility in DNA conformation. These may facilitate the distortion of the DNA required for catalysis by *Eco*RV.

In the presence of Mg^{2+} ions, type II restriction endonucleases cleave DNA at fixed locations relative to their recognition sites. For the majority of these enzymes, the recognition site is a unique set of bases, 4–8 bp[1] long (Bennett & Halford, 1989). However, a restriction enzyme can cleave a DNA molecule that contains several copies of its recognition sequence with different rates at each site. This was first seen in the reactions of the *Eco*RI endonuclease at its five sites on phage λ DNA (Thomas & Davis, 1975). Similar observations have since been reported not only for other restriction enzymes such as *Pst*I (Armstrong & Bauer, 1982), *Pae*R7 (Gingeras & Brooks, 1983), and *Hha*I (Drew & Travers, 1985) but also for other types of enzymes that carry out different reactions at specific sites on DNA (for example, the *dam* methyltransferase; Bergerat et al., 1989). In addition, the affinity of genetic regulatory proteins for a given sequence can vary between one site with this sequence and other sites with the same sequence (Gartenberg & Crothers, 1988; Brenowitz et al., 1989).

It has been suggested frequently that differences in the activity of a protein at different copies of the same DNA sequence are due to the bases flanking the target site, but in most systems, this proposal has yet to be tested by systematic alterations to the flanking sequences. With the *Eco*RI restriction enzyme, synthetic oligonucleotides that had A·T bp immediately adjacent to the recognition site were cleaved faster than those with G·C bp (Alves et al., 1984), but this pattern cannot account for the differential activity by *Eco*RI at its sites on λ DNA (Halford et al., 1980). Even so, flanking DNA plays a role in the activity of this enzyme. The *Eco*RI nuclease covers 12 bp when bound to DNA, the 6-bp recognition sequence, and 3 bp on either side (Rosenberg, 1991), and the protein interacts with several phosphates in the flanking DNA (Lu et al., 1981). These interactions are critical for DNA cleavage (Lesser et al., 1990). However, the choice between alternate *Eco*RI sites can also be determined by the locations of the sites in the DNA molecule. On a linear DNA that had two *Eco*RI sites close to one end, both flanked by the same sequences, the inner site was cleaved preferentially to the outer site; when this DNA was circularized, the sites were cleaved equally (Terry et al., 1985). The bias on the linear molecule is due to the protein binding initially to nonspecific DNA followed by linear diffusion along the lattice (Terry et al., 1987).

In addition to cutting DNA at their recognition sites, restriction enzyme also cleave DNA at other sites. The alternative (noncognate) sites differ from the recognition site by 1 bp [for example, Rosenberg and Greene (1982), Halford et al. (1986), and Barany (1988)]. Under standard conditions in vivo, reactions at noncognate sites may be over a million times slower than that at the recognition site, but they can still be detected by using high concentrations of the enzyme (Luke et al., 1987). At noncognate sites, each strand of the DNA is cleaved in a separate reaction while cognate sites are usually cleaved by concerted reactions on both strands (Barany, 1988; Taylor & Halford, 1989; Thielking et al., 1990). Consequently, restriction activity in vivo can be "proofread" by DNA ligase (Taylor et al., 1990). However, in vitro, activity at noncognate sites can be detected readily by using special reaction conditions: high pH and low ionic strength or organic solvents such as DMSO [viz., *Eco*RI (Polisky et al., 1975) and *Eco*RV (Halford et al., 1986)]. In this study, we have analyzed the effect of flanking sequences on the activity of the *Eco*RV restriction endonuclease, not at its cognate site but instead at one noncognate site that is cleaved by this enzyme at an unusually fast rate. At its recognition site, the rate-limiting step in the reaction of the *Eco*RV nuclease is product dissociation (Halford & Goodall, 1988) so it would be possible for flanking sequences that alter the rate of strand scission to have no effect on the overall rate of the reaction.

† This work was supported by the Science and Engineering Research Council.
‡ Present address: Department of Biochemistry, Duke University Medical Center, Durham, NC 27710.

[1] Abbreviations: bp, base pair(s); BME, β-mercaptoethanol; BSA, bovine serum albumin; Cm, chloramphenicol (with superscripts r and s to denote resistance and sensitivity); DMSO, dimethyl sulfoxide; kb, 1000 bp; M_r, relative molecular mass; Tc, tetracycline; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/92/0431-90$03.00/0

The *Eco*RV nuclease has been overproduced and crystallized (Bougueleret et al., 1985; D'Arcy et al., 1985). Structures of the *Eco*RV nuclease have been solved by X-ray crystallography for the free protein, the protein bound to a duplex oligonucleotide containing the *Eco*RV recognition sequence, and the protein bound to nonspecific DNA [Winkler et al., 1991; F. Winkler, personal communication;[2] reviewed by Rosenberg (1991)]. *Eco*RV has no homology to *Eco*RI in either primary or tertiary structures (Bouguelert et al., 1985; Winkler et al., 1991). In addition, the mode of action of *Eco*RV differs radically from that of *Eco*RI. For the *Eco*RI enzyme, a major factor in its specificity for DNA cleavage at the recognition site is its preferential binding to this site compared to any other DNA sequence (Terry et al., 1985, 1987; Thielking et al., 1990; Lesser et al., 1990). In contrast, the *Eco*RV enzyme shows no sequence specificity in binding to DNA: it has the same affinity for all DNA sequences, and all of its specificity for the recognition site in DNA cleavage is due to the catalytic reaction (Taylor & Halford, 1989; Newman et al., 1990; Taylor et al., 1991).

Experimental Procedures

Proteins. The *Eco*RV restriction endonuclease was purified, and its concentration was determined as before (Luke et al., 1987; Taylor & Halford, 1989). The molarities cited are for the enzyme dimer of M_r 57 000 (D'Arcy et al., 1985; Luke et al., 1987). Apart from *Eco*RI and T4 DNA ligase (purified here by C. N. Parker and L. R. Evans), all other enzymes for DNA manipulation were obtained from either Gibco BRL or Boehringer Mannheim and were used as advised by the supplier.

Plasmids. Three of the plasmids have been described previously: pACYC184 (Chang & Cohen, 1978), pAO3 (Oka et al., 1979), and pAT153 (Twigg & Sherratt, 1980). The plasmids were used to transform *Escherichia coli* HB101, the transformants were grown in M9 minimal media containing 1 mCi/L [*methyl*-^{3}H]thymidine, and the DNA was purified on CsCl/ethidium bromide gradients (Halford & Goodall, 1988). DNA sequences of these and other plasmids were analyzed with the UWGCG programs (Devereux et al., 1984).

Derivatives of pACYC184 (pACYC184/1 to pACYC184/7) were constructed by using standard procedures (Sambrook et al., 1989) to replace the 0.3-kb *Nco*I–*Eco*RI fragment in pACYC184 with synthetic duplexes. For each derivative, two oligonucleotides were synthesized and these were annealed to produce duplexes that had four-base extensions at both 5′ ends, one matching an *Eco*RI end and the other matching *Nco*I. The sequences of each pair of oligonucleotides are shown below (Figures 1 and 5). The oligonucleotides were made by either L. Hall (this department, using a Du Pont Coder 300 DNA synthesizer) or M. McClean (ICI Diagnostics, Northwich, Cheshire, using an Applied Biosystems machine). Each duplex was ligated to pACYC184 that had been cut previously with *Nco*I and *Eco*RI, and the ligation mix was used to transform *E. coli* HB101 to Tc^r Cm^s (the fragment of pACYC184 replaced by the duplex is from *cat*). Plasmids were isolated from a number of transformants, characterized initially by restriction digests and then by dideoxy sequencing of the double-stranded DNA (Chen & Seeburg, 1985). For sequencing, a primer was synthesized that annealed to pACYC184 with its 3′ end 39 bp downstream of the *Eco*RI site; in all constructs, the sequence between the

[2] The coordinates for the structure of an *Eco*RV–DNA complex are also available from the protein data base at Brookhaven (Ident Code 2RVE).

```
          1704                    1734                    1753
a)  GGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC

b)  AATTCTCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGC
        GAGTCGAGTGAGTTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCGGTAC

c)  AATTCTCAGCTCACTCAAAGGCGGTAATACGGATATCCACAGAATCAGGGC
        GAGTCGAGTGAGTTTCCGCCATTATGCCTATAGGTGTCTTAGTCCCGGTAC
```

FIGURE 1: The preferred noncognate site on pAT153. (a) The sequence shown is the top strand of pAT153 (Twigg & Sherratt, 1980) from 1704 to 1758. Three noncognate sites for *Eco*RV at positions 1704, 1734, and 1753 are underlined. (b) To construct the plasmid pACYC184/1, two oligonucleotides were synthesized with sequences as shown and, after annealing, the resultant duplex was cloned in pACYC184 that had been cut with *Eco*RI and *Nco*I. The insert contains the same sequence as 1708–1752 in pAT153. (c) To construct pACYC184/2, the same procedure was followed except that both oligonucleotides were 1 base different from b; the insert contains the *Eco*RV recognition site, GATATC (underlined), in place of the noncognate site, GTTATC.

*Eco*RI and *Nco*I sites was the same as that synthesized. ^{3}H-Labeled DNA was obtained from the transformants by the procedure given above.

Reaction Conditions. DNA cleavage by *Eco*RV was carried out at 20 °C in either buffer A or buffer C (Taylor & Halford, 1989). Buffer A is 50 mM Tris-HCl/100 mM NaCl/10 mM $MgCl_2$/10 mM BME/100 μg/mL BSA, pH 7.5; this is the standard buffer used for reactions of *Eco*RV at its recognition site, and reactions in this buffer with 0.5 nM *Eco*RV nuclease result in DNA cleavage only at the recognition site (Luke et al., 1987; Halford & Goodall, 1988). Buffer C is 50 mM Tris-HCl/100 mM NaCl/10 mM $MgCl_2$/10 mM BME/100 μg/mL BSA/10% (v/v) DMSO, pH 8.5; reactions in this buffer with 1.2 μM *Eco*RV enzyme result first in the cleavage of the recognition site followed by extensive cleavage of noncognate sites (Taylor & Halford, 1989). Procedures for measuring reaction rates at either cognate or noncognate sites were as given previously (Halford & Goodall, 1988; Taylor & Halford, 1989).

Results

The Preferred Noncognate Site on pAT153. The recognition sequence for the *Eco*RV restriction endonuclease is GATATC (Schildkraut et al., 1984). However, DNA cleavage by the *Eco*RV nuclease can also be detected at a number of additional sites, either in reactions containing DMSO or in standard reactions at high enzyme concentrations (i.e., >1 μM nuclease compared to <1 nM for cutting the recognition site; Halford et al., 1986; Taylor & Halford, 1989). The additional sites include all sequences that differ from the canonical site for *Eco*RV by 1 bp, except when the G is replaced by a pyrimidine or the C by a purine. The plasmid pAT153 contains, in addition to one *Eco*RV recognition site, 12 noncognate sites that can be cleaved by the *Eco*RV restriction enzyme.[3] One of these, GTTATC at position 1734, is cleaved more rapidly than all 11 of the other noncognate sites put together (Taylor & Halford, 1989). In pAT153, GTTATC occurs at three locations in addition to position 1734, but none of these is cleaved as readily as 1734. The nucleotide difference between GTTATC and GATATC therefore cannot be the reason why the noncognate site at 1734 is cut so quickly.

[3] The plasmid pAT1523 has 16 sites where the sequence differs from GATATC by 1 bp, but two of these have a pyrimidine in place of the G. Two further sites have the sequence GAGATC. This sequence includes a *dam* site (GATC; Barras & Marinus, 1989), and these two sites are cleaved by *Eco*RV only if the DNA lacks *dam* methylation (Halford et al., 1986). All of the DNA substrates used in this study were from *dam*$^+$ strains of *E. coli*, so GAGATC sites are also excluded.

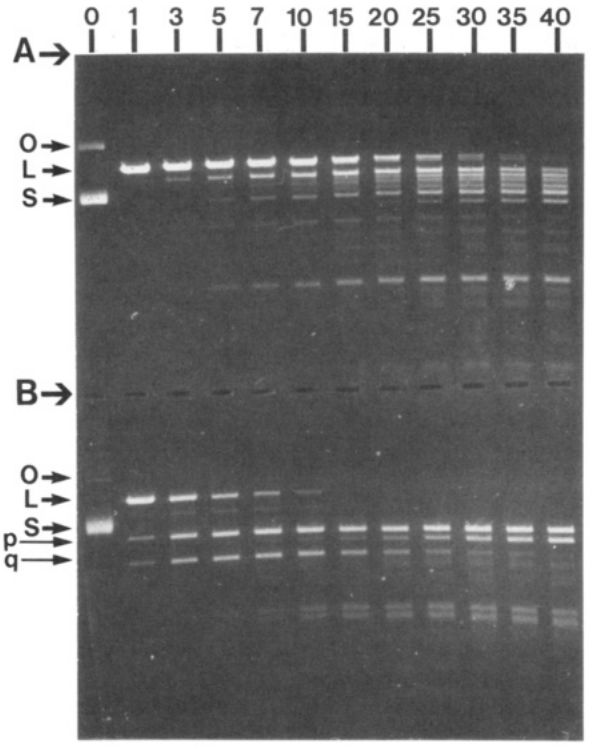

FIGURE 2: Cleavage of noncognate sites. Reactions in buffer C contained 1.2 μM *Eco*RV restriction enzyme and 10 nM DNA (initially, ≥90% supercoiled plasmid) at 20 °C. At timed intervals after the addition of the enzyme, samples were removed from the reactions, the reaction in each was stopped immediately, and the DNA was subsequently analyzed by electrophoresis through 1.2% agarose. The DNA was either pACYC184 (top row of wells, marked A) or pACYC184/1 (bottom row of wells, marked B). For both reactions, the time points (in minutes) at which the samples were removed are given above the gel, and the mobilities of the supercoiled, open-circle, and linear forms of the plasmids are marked on the left of the gel as S, O, and L, respectively. In part B, the initial products from the linear form of pACYC184/1 are marked p and q.

Possible reasons for the enhanced rate at this particular site for *Eco*RV include its flanking DNA sequence and its location within the DNA molecule. The sequence around position 1734 in pAT153 (Figure 1a) contains nothing to suggest that this segment of DNA may have a grossly abnormal conformation: the GC content is close to 50%; it lacks inverted repeats long enough to form a cruciform (Courey & Wang, 1983); it lacks periodic A-tracts to create an intrinsic bend in the DNA (Wu & Crothers, 1984); it lacks runs of >4 bp that are compatible with left-handed helices (McClean & Wells, 1988). However, two noncognate sites for *Eco*RV are located close to the preferred site at 1734; GGTATC at 1704 and GATAAC at 1753 (Figure 1a). No other region of pAT153 contains a density of noncognate sites as high as this section. The *Eco*RV restriction enzyme can translocate from one site to another without dissociating from the DNA (Taylor et al., 1991). If the *Eco*RV enzyme were to pause at noncognate sites during linear diffusion along the lattice, the amount of enzyme in the vicinity of the preferred site would be higher than elsewhere on the DNA; this could enhance the extent of cleavage at this site.

Isolation from the Cluster. To test whether the cluster of noncognate sites was the reason for enhanced cleavage by the *Eco*RV nuclease, the oligonucleotides shown in Figure 1b were synthesized. When annealed together, these produced a duplex with the same sequence as that around 1743 in pAT153 up to, but not including, the two neighboring sites. The duplex was used to construct a plasmid, pACYC184/1, in which this fragment of DNA replaced the 0.3-kb *Eco*RI–*Nco*I segment

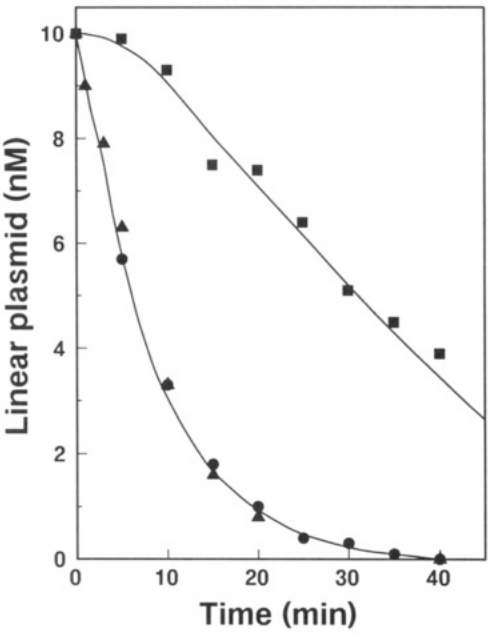

FIGURE 3: Reaction rates at noncognate sites. Reactions at 20 °C contained 1.2 μM *Eco*RV enzyme and 10 nM DNA in buffer C; the DNA was either pACYC184 (■), pACYC184/1 (●), or pAT153 (▲). The substrates were >90% supercoiled DNA and were ^{3}H-labeled. At timed intervals, samples were removed from the reactions, the reaction in each was stopped immediately, and the DNA was subsequently analyzed by electrophoresis through 1.2% agarose as in Figure 2. For each sample, the radioactivity in the segment of the gel containing the full-length linear form of the plasmid was measured by scintillation counting, to determine the concentration of linear DNA. The DNA concentrations given at zero time refer to total DNA; these values correspond to the amounts of full-length linear DNA directly after cutting the supercoiled plasmids at their *Eco*RV recognition sites.

in pACYC184: the new plasmid has no noncognate sites for *Eco*RV within 100 bp of the insert. The reactions of the *Eco*RV restriction enzyme at its noncognate sites were studied on both the vector and the derivative (Figure 2).

The parental plasmid, pACYC184, contains one *Eco*RV recognition site and 13 noncognate sites. When treated with a comparatively high concentration of the nuclease, the supercoiled form of this DNA was converted to the linear form within 1 min, due to the reaction at the recognition site; the linear DNA was subsequently cleaved again, at noncognate sites, to produce a series of smaller fragments (Figure 2A). The linear form of pACYC184 yielded directly a mixture of DNA fragments, each produced by a double-strand break at a different noncognate site. Hence, this plasmid contains several noncognate sites that are cleaved at similar rates. The derivative, pACYC184/1, possesses the same *Eco*RV recognition site as the vector and was also converted from its supercoiled to its linear form within 1 min (Figure 2B). But, in contrast to the parental plasmid, the majority of the linear form of pACYC184/1 was cut first at one particular noncognate site, to yield initially just two products (marked p and q in Figure 2B); later, these were cleaved again to a number of small fragments. The noncognate site on pACYC184/1 that was the first to be cleaved by *Eco*RV was identified by restriction mapping: the site was located on the insert between the *Nco*I and *Eco*RI sites (data not shown).

The rate for the decline in the concentration of full-length linear DNA, caused by *Eco*RV reactions at noncognate sites, was measured for both pACYC184 and pACYC184/1 (Figure 3). The disappearance of linear DNA requires both strands to be cut at the same noncognate site, and the lag phase observed in these reactions is due to the two strands being cut

in sequential reactions. Reaction rates were measured from the subsequent exponential phase as in Taylor and Halford (1989). The rate for the parental plasmid was about 5 times slower than that for the derivative. On pACYC184, the decline in the amount of full-length linear DNA is due to the sum of *Eco*RV reactions at several noncognate sites, and it is impossible to determine the rate at any one site. But on pACYC184/1, the decline is caused by the reaction at one noncognate site (Figure 2). If we assume that all 13 noncognate sites on pACYC184 are cut at similar rates, the reaction at the preferred site on pACYC184/1 is 65 times faster than that at each site on pACYC184.

Similar experiments were also carried out with pAT153. This DNA was also linearized rapidly by the reaction of *Eco*RV at its cognate site; the rate for the subsequent decline in the concentration of the linear form, due in this case to the reaction at the noncognate site at position 1734, was indistinguishable from the rate with pACYC184/1 (Figure 3). These experiments eliminate the possibility that the enhanced rate of cleavage of the noncognate site at 1734 in pAT153 is due to the cluster of neighboring sites (Figure 1a). Instead, it is an intrinsic feature of the section of DNA shown in Figure 1b, regardless of where the section is placed.

The Recognition Site Flanked by the Same Sequence. Do the flanking sequences that cause enhanced *Eco*RV activity at a noncognate site also increase activity at the recognition site for *Eco*RV? To answer this question, the same strategy as above was used to construct a second derivative of pACYC184. For pACYC184/2, the duplex that was inserted between the *Eco*RI and *Nco*I sites on the vector was made from the two oligonucleotides shown in Figure 1c. They differ from the pair used to construct pACYC184/1 by one base in each strand, in order to replace GTTATC by GATATC. The construct thus contains two copies of the *Eco*RV recognition sequence; one on the vector and another on the insert. In order to measure the reaction rates at both *Eco*RV sites on pACYC184/2, the circular DNA was first converted to its linear form by digestion with *Ava*I (the plasmid has only one *Ava*I site). The linear DNA (3.9 kb) is cleaved by *Eco*RV initially to two fragments, but the reaction at the site in the vector yields products of 2.6 and 1.3 kb while that at the insert gives fragments of 2.9 and 1.0 kb; subsequent cleavage at both *Eco*RV sites generates three fragments of 1.6, 1.3, and 1.0 kb. Hence, each set of products could be separated from the other and from the substrate by electrophoresis through agarose.

The fraction of the *Ava*I-linearized pACYC184/2 that was cleaved by *Eco*RV first at the canonical site in the *Eco*RI–*Nco*I insert was measured, as was also the fraction cleaved first at the canonical site in the vector. Both sets of initial products were formed at the same rate (Figure 4). Hence, the relative rates of *Eco*RV reactions at two copies of its cognate site, GATATC, are unaffected by adjacent nucleotides, even though the same flanking sequences cause one copy of the noncognate site, GTTATC, to be cleaved faster than other copies. The reaction of the *Eco*RV restriction enzyme at its recognition site in pAT153 has been characterized previously (Halford & Goodall, 1988), and the site in pACYC184 is located within an identical sequence for >150 bp in both directions. On both plasmids, the *Eco*RV site is within the gene for Tc^r. The sequence around this site has no resemblance to that in Figure 1c.

Nucleotides That Enhance Activity at the Noncognate Site. The DNA sequences of several plasmids related to pAT153 were examined to see if any contained a sequence that was similar, but not identical, to that around position 1734 in pAT153. [The plasmid pAT153 (Twigg & Sherratt, 1980) was derived from pBR322, which in turn was derived from pMB1 (Bolivar et al., 1977); pMB1 is a naturally occurring plasmid that is homologous to colE1.] One plasmid that met this requirement was pAO3, a deletion mutant of colE1 (Oka et al., 1979). On pAO3, there are seven noncognate sites for the *Eco*RV restriction enzyme, one of which is GTTATC surrounded by a sequence similar to that at 1734 in pAT153 (Figure 5a,b). When the activity of the *Eco*RV nuclease at

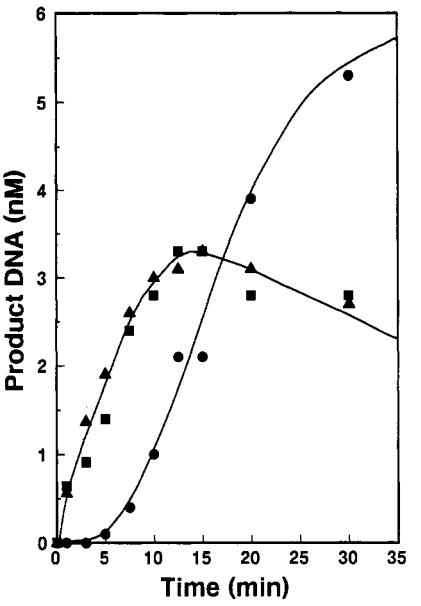

FIGURE 4: Effect of flanking sequences at canonical sites. The reaction at 20 °C in buffer A contained 0.5 nM *Eco*RV restriction endonuclease and 10 nM ^{3}H-labeled pACYC184/2 that had been cut previously with *Ava*I. Samples were withdrawn at timed intervals and the reaction in each sample was terminated immediately. The DNA was subsequently analyzed by electrophoresis through 1.2% agarose, and the concentrations of the DNA products were determined by scintillation counting on slices from the gel: product from cutting the *Eco*RV site in the *Eco*RI–*Nco*I insert (■); product from cutting the *Eco*RV site in the vector DNA (▲); and product after cutting both *Eco*RV sites (●).

```
    1712                    1734                    1753
a)  CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC   pAT153
        •   **      *  *
b)  CTCACaCAAAaaCGGTAAcACaGTTATCCACAGAATCAGGGGATAAC   pAO3
    311                 333                     357

c)                AATtCGGTTATCCAagcttc                pACYC184/3
                    GCCAATAGGTtcgaagGTaC

d)              aaTtcTACGGTTATCCACAagcTtc              pACYC184/4
                  gATCGCAATAGGTGTTcgAaggtaC

e)              aaTtcTACaGTTATCCACAagcTtc              pACYC184/5
                  gATCtCAATAGGTGTTcgAaggtaC

f)              aaTtccACGGTTATCCACAagcTtc              pACYC184/6
                  ggTCGCAATAGGTGTTcgAaggtaC

g)              aaTtcaACGGTTATCCACAagcTtc              pACYC184/7
                  gtTCGCAATAGGTGTTcgAaggtaC
```

FIGURE 5: Alterations in flanking sequences at the noncognate site. (a) The sequences shown is from 1712 to 1758 in pAT153. The noncognate site at 1734 is underlined. (b) The sequence shown is from 311 to 362 in pAO3 (Oka et al., 1979). A noncognate site at 333 is underlined, and the differences between this sequence and that above are marked with asterisks. (c–g) Derivatives of pACYC184 were constructed by annealing pairs of synthetic oligonucleotides to make duplexes and ligating each duplex to pACYC184 that had been cut previously with *Eco*RI and *Nco*I: for pACYC184/3, duplex c; for pACYC184/4, duplex d; for pACYC184/5, duplex e; for pACYC184/6, duplex f; for pACYC184/7, duplex g. In addition to a GTTATC site (underlined), each duplex also contains a *Hin*dIII site (AAGCTT) to facilitate the identification of clones carrying the requisite plasmid. The sequences are aligned on the GTTATC site; bases at positions where the nucleotide differs from pAT153 are given in lower case while all matches are in upper case.

its noncognate sites on pAO3 was measured as in Figure 2, the initial cleavage of the DNA generated a number of different products due to reactions at several sites (data not shown). Like pACYC184 (Figure 2A), none of the alternate sites on pAO3 was cleaved more readily than any other and none were cleaved as fast as the preferred noncognate site on pAT153. The reduced activity on pAO3, relative to pAT153, is likely to be due to one or more of the bases that differ between the two sequences.

The differences in the sequences from pAT153 and pAO3 (Figure 5a,b) were used as a guide to the construction of further plasmids, like pACYC184/1 but with different oligonucleotide inserts (Figure 5c–g). To find out the length of flanking DNA needed to give enhanced susceptibility to *Eco*RV, one derivative of pACYC184 was made with an *Eco*RI–*Nco*I insert that retained 10 bp from the pAT153 site, GTTATC, and 2 bp on either side (Figure 5c, pACYC184/3). Another construct kept 4 bp from either side of the GTTATC site (Figure 5d, pACYC184/4). When tested as substrates for the *Eco*RV restriction enzyme under the same conditions as in Figure 2, both pACYC184/3 and pACYC184/4 were first linearized by the reaction at the recognition site, and the linear forms were then cleaved at one or more noncognate sites (data not shown). With pACYC184/3, the initial reactions at noncognate sites gave the same set of multiple products as pACYC184 itself (as in Figure 2A). Thus, the decline in the concentration of the full-length linear form of pACYC184/3 (Figure 6) is due to *Eco*RV reactions at several noncognate sites; the rate of the decline was the same as that of pACYC184. In contrast, the *Eco*RV reactions as noncognate sites on pACYC184/4 produced initially the same two fragments seen with pACYC184/1 (p and q; Figure 2B). Hence, pACYC184/4 possesses one noncognate site that is preferred to any other, and the reaction at that site resulted in a decline in the concentration of linear DNA (Figure 6) that was as rapid as pACYC184/1. The retention of 4 bp from either side of the GTTATC site is therefore sufficient for enhanced activity by *Eco*RV while the retention of 2 bp from either side is not.

Given the orientation of the sequences in Figure 5, the 4 bp on the right of the target site are the same in both pAT153 and pAO3. Hence, by itself, the right-hand flanking sequence, CACA, cannot be responsible for the enhanced reaction rate by *Eco*RV. The left-hand flanking sequences differ between the two plasmids at two positions: TACG in pAT153 and CACA in pAO3. To examine these differences, three more derivatives of pACYC184 were made with the *Eco*RI–*Nco*I inserts shown in Figure 5e–g. All three constructs kept the CACA sequence on the right-hand side of GTTATC but altered the TACG sequence on the left to either TACA (pACYC184/5), CACG (pACYC184/6), or AACG (pACYC184/7). The reactions of the *Eco*RV restriction endonuclease on the plasmid constructs were measured as above (Figure 6). With TACA as the left-hand flanking sequence instead of TACG, the construct behaved like pACYC184 rather than pACYC184/1: it lacked a preferred noncognate site and the sum of *Eco*RV reactions at all sites was comparatively slow (Figure 6). With either CACG or AACG on the left, the site on the *Eco*RI–*Nco*I insert was cleaved more rapidly than all other noncognate sites on the DNA (data not shown) though the rate for cleaving this site varied between plasmids (Figure 6). With CACG (pACYC184/6), the site was cleaved at the same rate as the wild-type site flanked by TACG (pACYC184/4), but with AACG (pACYC184/7), the rate was reduced. The rate on pACYC184/7 was about 2 times slower than that on pACYC184/6, and the same 2-fold difference was also observed under a variety of experimental conditions, covering a range of enzyme and $MgCl_2$ concentrations (data not shown).

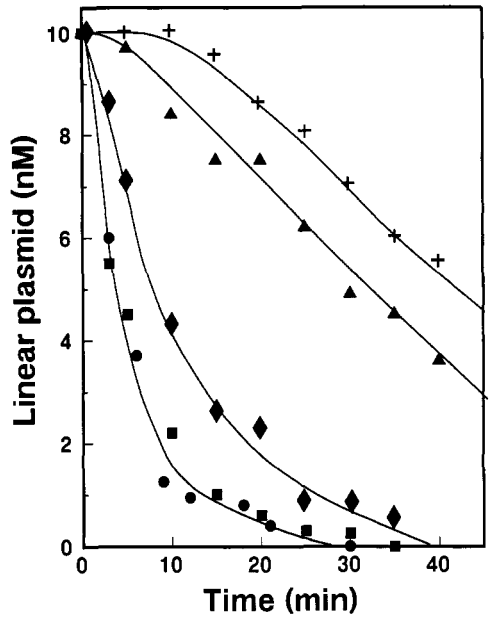

FIGURE 6: Reaction rates at noncognate sites with altered flanking sequences. Reactions at 20 °C in buffer C contained 1.2 μM *Eco*RV restriction enzyme and 10 nM ^{3}H-labeled DNA from one of the following plasmids (initially in their supercoiled states): pACYC184/3 (duplex c in Figure 5), +; pACYC184/4 (duplex d), ●; pACYC184/5 (duplex e), ▲; pACYC184/6 (duplex f), ■; pACYC184/7 (duplex g), ♦. At timed intervals, samples were removed from the reactions, the reaction in each was stopped immediately, and the DNA was subsequently analyzed by electrophoresis through agarose. The concentration of the full-length linear form of the plasmid was measured at each time point by scintillation counting on slices from the gels. For each plasmid, the linear form was generated by cutting the supercoiled DNA at the *Eco*RV recognition site within the first minute of the reaction. The DNA concentrations given at zero time are the total DNA in the reaction mixtures; these correspond to the concentrations of the full-length linear form directly after the reaction at the recognition site.

Discussion

The reaction of the *Eco*RV restriction endonuclease at 1 out of 12 noncognate sites on pAT153 is faster than the sum of the reactions at the other 11 sites (Taylor & Halford, 1989). This site, GTTATC at position 1734, is also cleaved faster than all 13 noncognate sites in pACYC184 (Figure 3) and all 7 sites on pAO3 (data not shown). The site retained its susceptibility to *Eco*RV when a 45-bp segment of DNA from pAT153, encompassing position 1734 but no other noncognate sites, was cloned in pACYC184 (Figure 2B). Hence, it seems that the enhanced rate is due solely to the nucleotide sequences flanking the site. However, the same flanking sequences caused no increase in activity when placed around the canonical site for *Eco*RV, GATATC (Figure 4). As noted above, this may be due to the fact that the rate-limiting step for *Eco*RV at its recognition site is product dissociation (Halford & Goodall, 1988). If any step in the mechanism before product release had been accelerated by this flanking sequence, it would have made no difference to k_{cat}.

The activity of the *Eco*RI restriction enzyme at sites that differ from its recognition site by 1 bp varies with the nature of the substitution: certain substitutions cause much greater reductions in activity than others (Rosenberg & Greene, 1982; Thielking et al., 1990; Lesser et al., 1990). The *Eco*RV re-

Table I: Sequences Flanking GTTATC Sites

Plasmid (Position)[a]	Sequence[b]	Relative Rate[c]	Flanking Character[d] Left / Right
pAT153 (1734)	TAATACGGTTATCCACAAGA	++	+G / C+
pACYC184/1	TAATACGGTTATCCACAAGA	++	+G / C+
pACYC184/4	TTCTACGGTTATCCACAAGC	++	+G / C+
pACYC184/6	TTCCACGGTTATCCACAAGC	++	+G / C+
pACYC184/7	TTCAACGGTTATCCACAAGC	•	-G / C+
pACYC184/3	GAATTCGGTTATCCAAGCTT	-	-G / C-
pACYC184/5	TTCTACAGTTATCCACAAGC	-	+A / C+
pAO3 (333)	TAACACAGTTATCCACAAGC	-	+A / C+
pAT153 (1753)	TTCCTGCGTTATCCCCTGAT	-	+C / C-
pAT153 (2672)	TATCGTAGTTATCTACACGA	-	+A / T-
pAT153 (3060)	CCGCAGTGTTATCACTCATG	-	+T / A+
pACYC184 (247)	TTTTTGAGTTATCGAGATTT	-	-A / G+

[a] The plasmids pAT153, pAO3, and pACYC184 contain the sequence GTTATC at the positions indicated: all copies of this sequence are listed. For the derivatives of pACYC184, the sequence shown is that on the *Eco*RI–*Nco*I insert. [b] Asterisks mark the bases that differ from the sequence around position 1734 on pAT153. [c] Relative rates are defined as follows: ++, the site is cleaved faster than all other noncognate sites on the plasmid, at the same rate as position 1734 on pAT153; +, the site is cleaved faster than all other sites on the plasmid, but at a rate slower than position 1734 on pAT153; -, the site is not cleaved faster than other sites on the plasmid. [d] Flanking characters are specified by + or - to denote the flexibility of the DNA, assessed by the algorithm of Satchwell and Travers (1988), and by the nucleotide immediately adjacent to the GTTATC sequence.

striction enzyme behaves similarly in this respect (Newman et al., 1990; A. Pingoud, personal communication). Therefore, the effect of flanking sequences on the reactivity of the *Eco*RV nuclease at a noncognate site can only be characterized by comparing different copies of the same sequence. The sequences around all of the GTTATC sites on the plasmids used here, and their relative rates for cleavage by *Eco*RV, are shown in Table I. The table includes both the GTTATC sites constructed here and the intrinsic sites in the vector DNA. Among the intrinsic sites, 1734 in pAT153 is unique in that none of the others is cleaved by *Eco*RV as readily as this one. If a plasmid lacks a noncognate site that is cleaved faster than all the other sites, it is impossible to measure the reaction rate at an individual site from the decline in the concentration of full-length linear DNA. However, one interpretation of the kinetic data given here is that certain copies of GTTATC are cleaved 65 times faster than other copies.

Nucleotides inside the DNA–Protein Complex. Position 333 in the plasmid pAO3 is surrounded by a sequence similar to that at 1734 in pAT153 (Figure 5a,b), yet the site in pAO3 is cleaved slowly. One or more of the bases that differ between the two sequences must be responsible for the change in *Eco*RV activity. Since a derivative of pACYC184 carrying GTTATC and 4 bp from either side of the 1734 site showed the enhanced reaction rate (pACYC184/4; Figure 6), the key sequence can be narrowed down to the 4 bp on the left of the GTTATC site: TACG in pAT153 and CACA in pAO3. When the TACG flanking sequence was replaced with TACA (pACYC184/5), the site was no longer cleaved at an elevated rate (Figure 6). Hence, one factor that determines *Eco*RV activity is the nucleotide immediately adjacent to the site. All of the GTTATC sites that were cleaved rapidly are bounded at their 5′ and 3′ ends by G and C, respectively (Table I). The pAO3 site is flanked by A and C.

In the crystal structures of the *Eco*RV restriction endonuclease bound to duplex oligonucleotides (Winkler et al., 1991), the protein engulfs 10 bp of DNA: i.e., the 6-bp recognition sequence (GATATC) and 2 bp on either side [F. Winkler, personal communication;[2] reviewed by Rosenberg (1991)]. All of the bases that are contacted directly by the protein are within the central 6-bp sequence, but the protein also contacts virtually all of the phosphates throughout the 10-bp segment. It is now well-established that the interaction of a protein with a specific DNA sequence can be modulated by alterations to the sequence at loci where the only part of the DNA contacted by the protein is the phosphodiester backbone [reviewed by Travers (1989, 1991) and by Harrison and Aggarwal (1990)]. Even though the *Eco*RV nuclease makes no contact with the base on the nucleotide next to the target site, the protein still makes use of the phosphates on the adjacent nucleotides. Flanking sequences could modulate these contacts with the phosphates by determining the precise conformation of the DNA in the DNA–protein complex.

The conformation of the DNA in its complex with *Eco*RV may in turn determine the activation energy barrier for DNA cleavage. This restriction endonuclease binds all DNA sequences with the same equilibrium constant (Taylor et al., 1991). The complexes with specific and nonspecific DNA are isoenergetic, but they are not isosteric. In particular, the structure of the cognate DNA is distorted significantly from

that of B-DNA (F. Winkler, personal communication). In the specific complex, all of the energy that is potentially available from interactions with the bases in the recognition sequence appears to be dissipated in deforming the structure of the complex (Taylor et al., 1991). The purpose of the deformation of the cognate DNA is presumably to reduce the activation energy barrier for DNA scission (Newman et al., 1990).

Nucleotides outside the DNA–Protein Complex. The nucleotides immediately adjacent to the target sequence cannot, however, be the only reason why certain copies of GTTATC are cut by *Eco*RV at faster rates than other copies. In pACYC184/3 and pACYC184/4 (Figure 5c,d), the 10-bp segments that fill the active site of the *Eco*RV restriction enzyme have the same sequence, CGGTTATCCA, yet only the site on pACYC184/4 is cleaved rapidly (Figure 6). Hence, the activity of the *Eco*RV nuclease is influenced by flanking sequences beyond the reach of the protein. Outside the 10-bp segments, the protein cannot contact directly either the bases or the phosphodiester backbone (Winkler et al., 1991).

All of the GTTATC sites that are cleaved quickly by *Eco*RV are flanked by alternating pyrimidine/purine sequences for ≥4 bp on both sides (Table I; the possible significance of this was pointed out to us by A. Travers, personal communication). Moreover, when TACG on the left of the site in pACYC184/4 was replaced by AACG (pACYC184/7), *Eco*RV activity was reduced, even though this T → A transversion is well clear of the surface of the protein. In contrast, the replacement of TACG by CACG (pACYC184/6) made no difference to *Eco*RV activity (Figure 6). Hence, it seems that the identity of the base 4 bp away from the target site is not critical but that the alternation between pyrimidines and purines is favorable. DNA molecules with alternating sequences of this type can readily take up difference conformations, depending on the local sequence context (Travers, 1989). In contrast, molecules with nonalternating sequences, particularly homopolymeric A_n or T_n, have much less conformational freedom (Travers, 1991). Consequently, loci within DNA–protein complexes where the DNA has to be deformed can display a strong bias toward "flexible" DNA sequences over "inflexible" sequences. For example, in intact (H1-containing) nucleosomes, the most deformed section of the DNA contains a high incidence of alternating sequences, and this can be correlated to the physical flexibility of the DNA (Satchwell & Travers, 1989).

The algorithm used by Satchwell and Travers (1989), to identify flexible DNA in nucleosomes, was applied to the flanking sequences around all of the GTTATC sites shown in Table I. (The algorithm yields either a positive or a negative number from any given length of DNA sequence; a positive number is taken to indicate that the segment has a flexible conformation which can be readily distorted, while a negative number suggests the converse.) All of the sites that were cleaved rapidly not only contained G and C at their 5′ and 3′ edges but were also surrounded on both sides by sequences that scored positively for DNA flexibility. Though pACYC184/3 had the same sequence as pACYC184/4 over 10 bp, the former was embedded in a sequence that scored negatively for flexibility while the latter had a positive score. In addition, the replacement of TACG with CACG as the left-hand flanking sequence retained a positive score for flexibility while AACG yielded a negative score (Table I). By itself, DNA flexibility cannot enhance the activity of the *Eco*RV restriction enzyme. Table I contains three examples of GTTATC sites, flanked on both sides by flexible DNA, that are cleaved slowly: all three lack G and C immediately next to the site. However, the deformation of the DNA required for catalysts by *Eco*RV may be facilitated by flanking sequences that confer flexibility to DNA structure.

Added in Proof

In the crystal structure of the *Eco*RV nuclease bound to its recognition sequence, the DNA is a 10-bp duplex. Recent model-building studies, involving the addition of B-DNA to the ends of the 10-bp segment (F. K. Winkler, personal communication), suggest that the protein may be in direct contact with the phosphates over a longer section of DNA than had been anticipated. In particular, Lys102 and Lys104 could interact with phosphates outside the 10-bp segment.

Acknowledgments

We are indebted to Fritz Winkler for the crystal structures of *Eco*RV and to Andrew Travers for both the idea of flexible DNA and the algorithm for its identification. We thank Bernard Connolly, Len Hall, Mick McClean, Alfred Pingoud, Christian Vermote, and Barry Vipond for materials, for discussions, and for unpublished data.

Registry No. *Eco*RV restriction endonuclease, 83589-02-0.

References

Alves, J., Pingoud, A., Haupt, W., Langowski, J., Peters, F., Maass, G., & Wolff, C. (1984) *Eur. J. Biochem. 140*, 83–92.

Armstrong, K. A., & Bauer, W. R. (1982) *Nucleic Acids Res. 10*, 993–1007.

Barany, F. (1988) *Gene 56*, 149–165.

Barras, F., & Marinus, M. G. (1989) *Trends Genet. 5*, 139–143.

Bennett, S. P., & Halford, S. E. (1989) *Curr. Top. Cell. Regul. 30*, 57–104.

Bergerat, A., Kriebardis, A., & Guschlbauer, W. (1989) *J. Biol. Chem. 264*, 4064–4070.

Bolivar, F., Rodriguez, R. L., Betlach, M. C., & Boyer, H. W. (1977) *Gene 2*, 75–93.

Bougueleret, L., Tenchini, M. L., Botterman, J., & Zabeau, M. (1985) *Nucleic Acids Res. 13*, 3823–3839.

Brenowitz, M., Senear, M., & Ackers, G. K. (1989) *Nucleic Acids Res. 17*, 3747–3755.

Chang, A. C. Y., & Cohen, S. N. (1978) *J. Bacteriol. 134*, 1141–1156.

Chen, E. J., & Seeburg, P. H. (1985) *DNA 4*, 165–170.

Courey, A. J., & Wang, J. C. (1983) *Cell 33*, 817–829.

D'Arcy, A., Brown, R. S., Zabeau, M., Van Resandt, R. W., & Winkler, F. K. (1985) *J. Biol. Chem. 252*, 1987–1990.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.

Drew, H. R., & Travers, A. A. (1985) *Nucleic Acids Res. 13*, 4445–4467.

Gartenberg, M. R., & Crothers, D. M. (1988) *Nature (London) 333*, 824–829.

Gingeras, T. R., & Brooks, J. E. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 402–406.

Halford, S. E., & Goodall, A. J. (1988) *Biochemistry 27*, 1771–1777.

Halford, S. E., Johnson, N. P., & Grinsted, J. (1980) *Biochem. J. 191*, 581–592.

Halford, S. E., Lovelady, B. M., & McCallum, S. A. (1986) *Gene 41*, 173–181.

Harrison, S. C., & Aggarwal, A. K. (1990) *Annu. Rev. Biochem. 59*, 933–969.

Lesser, D. R., Kurpiewski, M. R., & Jen-Jacobson, L. (1990) *Science (Washington, D.C.) 250*, 776-786.
Lu, A.-L., Jack, W. E., & Modrich, P. (1989) *J. Biol. Chem. 256*, 13200-13206.
Luke, P. A., McCallum, S. A., & Halford, S. E. (1987) *Gene Amplif. Anal. 5*, 183-205.
McClean, M. J., & Wells, R. D. (1988) *Biochem. Biophys. Acta 950*, 243-254.
Newman, P. C., Williams, M. D., Cosstick, R., Seela, F., & Connolly, B. A. (1990) *Biochemistry, 29*, 9902-9910.
Oka, A., Nomura, N., Morita, M., Sugisaki, H., Sugimoto, K., & Takanami, M. (1979) *Mol. Gen. Genet. 172*, 151-159.
Polisky, B., Greene, P., Garfin, D. E., McCarthy, B. J., Goodman, H. M., & Boyer, H. W. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3310-3314.
Rosenberg, J. M. (1991) *Curr. Opinion Struct. Biol. 1*, 104-113.
Rosenberg, J. M., & Greene, P. (1982) *DNA 1*, 117-123.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Satchwell, S. C., & Travers, A. A. (1989) *EMBO J. 8*, 229-238.
Schildkraut, I., Banner, C. D., Rhodes, C. S., & Parekh, S. (1984) *Gene 27*, 327-329.
Taylor, J. D., & Halford, S. E. (1989) *Biochemistry 28*, 6198-6207.
Taylor, J. D., Goodall, A. J., Vermote, C. L., & Halford, S. E. (1990) *Biochemistry 29*, 10727-10733.
Taylor, J. D., Badcoe, I. G., Clarke, A. R., & Halford, S. E. (1991) *Biochemistry 30*, 8743-8753.
Terry, B. J., Jack, W. E., & Modrich, P. (1985) *J. Biol. Chem. 260*, 13130-13137.
Terry, B. J., Jack, W. E., & Modrich, P. (1987) *Gene Amplif. Anal. 5*, 103-118.
Thielking, V., Alves, J., Fliess, A., Maass, G., & Pingoud, A. (1990) *Biochemistry 29*, 4682-4691.
Thomas, M., & Davis, R. W. (1975) *J. Mol. Biol. 91*, 315-328.
Travers, A. A. (1989) *Annu. Rev. Biochem. 58*, 427-452.
Travers, A. A. (1991) *Curr. Opinion Struct. Biol. 1*, 114-122.
Twigg, A. J., & Sherratt, D. J. (1980) *Nature (London) 283*, 216-218.
Winkler, F. K., D'Arcy, A., Blocker, H., Frank, R., & van Boom, J. H. (1991) *J. Mol. Biol. 217*, 235-238.
Wu, H.-M., & Crothers, D. M. (1984) *Nature (London) 308*, 509-513.

Recognition of the Oligosaccharide and Protein Moieties of Glycoproteins by the UDP-Glc:Glycoprotein Glucosyltransferase†

Marcelo C. Sousa,‡ Miguel A. Ferrero-Garcia,§ and Armando J. Parodi*,‖
Instituto de Investigaciones Bioquimicas, Fundación Campomar, Antonio Machado 151, 1405 Buenos Aires, Argentina
Received May 9, 1991; Revised Manuscript Received August 12, 1991

ABSTRACT: It was found, in cell-free assays, that the $Man_8GlcNAc_2$ and $Man_7GlcNAc_2$ isomers having the mannose unit to which the glucose is added were glucosylated by the rat liver glucosyltransferase at 50 and 15%, respectively, of the rate of $Man_9GlcNAc_2$ glucosylation. This indicates that processing by endoplasmic reticulum mannosidases decreases the extent of glycoprotein glucosylation. All five different glycoproteins tested (bovine and porcine thyroglobulins, phytohemagglutinin, soybean agglutinin, and bovine pancreas ribonuclease B) were found to be poorly glucosylated or not glucosylated unless they were subjected to treatments that modified their native conformations. The effect of denaturation was not to expose the oligosaccharides but to make protein determinants, required for enzymatic activity, accessible to the glucosyltransferase because (a) cleavage of denatured glycoproteins by unspecific (Pronase) or specific (trypsin) proteases abolished their glucose acceptor capacities almost completely except when the tryptic peptides were held together by disulfide bonds and (b) high mannose oligosaccharides in native glycoproteins, although poorly glucosylated or not glucosylated, were accessible to macromolecular probes as concanavalin A-Sepharose, endo-β-*N*-acetylglucosaminidase H, and jack bean α-mannosidase. In addition, denatured, endo-β-*N*-acetylglucosaminidase H deglycosylated glycoproteins were found to be potent inhibitors of the glucosylation of denatured glycoproteins. It is suggested that in vivo only unfolded, partially folded, and malfolded glycoproteins are glucosylated and that glucosylation stops upon adoption of the correct conformation, a process that hides the protein determinants (possibly hydrophobic amino acids) from the glucosyltransferase.

The N-glycosylation of protein is initiated by the transfer of an oligosaccharide having in most cases the composition $Glc_3Man_7GlcNAc_2$ from a dolichol-P-P derivative to asparagine residues in nascent or recently terminated polypeptide chains. The transfer reaction occurs in the rough endoplasmic reticulum and is immediately followed by removal of the glucose units catalyzed by two specific glucosidases: glucosidase I, which removes the more external unit, and glucosidase II, which excises both remaining units. Several of the peripheral α(1,2)-linked mannose residues may be then removed by two α-mannosidases located, the same as for both glucosidases, in the lumen of the endoplasmic reticulum. Further

† This work was supported by United States Public Health Service Grant 1 RO1 GM44500-01, by the National Research Council (Argentina), by the University of Buenos Aires, and by the Swedish Agency for Research Cooperation with Developing Countries.
* To whom correspondence should be addressed.
‡ Fellow of the National Research Council (Argentina).
§ On leave from the University of León, Spain.
‖ Career Investigator of the National Research Council (Argentina).

0006-2960/92/0431-97$03.00/0