

Evidence for a repeating domain in type I restriction enzymes

Patrick Argos¹

European Molecular Biology Laboratory, Biological Structures Division,
Postfach 10.2209, 6900 Heidelberg, FRG

¹On sabbatical leave from Purdue University, Department of Biological
Sciences, West Lafayette, IN 47907, USA

Communicated by T.Bickle

The primary structures of the recognition subunit (*hsdS*) in type I restriction enzymes from three isolates of *Escherichia coli* were compared and aligned by use of amino acid physical properties. A repeating domain was found in each of the subunits suggesting a pseudo-dimeric structure. Secondary structure predictions delineated two helical regions in each domain which suggested that the recognition subunits may act in a fashion similar to that proposed for repressor and activator molecules; namely, interaction with double-stranded DNA through helices and in two successive major grooves on the same DNA side. One helical motif could provide the specific recognition site and the other, the restriction site.

Key words: type I restriction enzymes/structure analysis/*Escherichia coli*

Introduction

Type I restriction enzymes of *Escherichia coli* and other bacteria are complex multifunctional molecules consisting of three subunit proteins coded for by chromosomally located genes (for reviews, see Endlich and Linn, 1981; Yuan, 1981; Bickle, 1982). The *hsdS* gene product is responsible for recognition of a specific DNA sequence while the *hsdM* protein coupled with that from *hsdS* allows methylase activity at the DNA recognition site resulting in 6-methyladenine. The *hsdR* gene product along with the other two is essential for endonuclease activity; however, the site of restriction cleavage occurs randomly 0.4–7.0 kb from the recognition sequence (Bickle *et al.*, 1978). Magnesium, S-adenosylmethionine and ATP are required as co-factors; ATP is hydrolyzed during and after restriction and DNA translocation. The type II restriction endonucleases often used as reagents for site-specific DNA cleavage, contrast with the type I enzymes by virtue of their one-subunit composition and cleavage within their recognition site.

Gough and Murray (1983) have recently determined the nucleotide sequence of the *hsdS* gene for three bacterial systems: K, B and D. The *hsdS* K, B and D proteins contain between 444 and 474 amino acids and display two strongly conserved spans of ~40 and 90 residues in length near the middle and C-terminal regions, respectively.

In the present work the entire sequences of the three proteins were aligned using comparisons derived from amino acid physical properties. A likely repeating domain in each of the *hsdS* proteins was found with implications for interaction of the recognition domains with DNA. It must be emphasized that the repeat is within the same *hsdS* gene and differs from the previous work of Gough and Murray (1983) who found limited similarities bet-

ween the *hsdS* genes in the three bacterial systems. A protein helical structure similar to that adopted by gene repressors and activators (Steitz *et al.*, 1982; Ohlendorf *et al.*, 1983) is proposed.

Results

All pairwise comparisons of the *hsdS* subunits from the D, B and K *E. coli* systems were effected using the search procedures described in Materials and methods. The results for the *hsdS* B–K and *hsdS* D–B comparisons are shown in Figures 1 and 2. It is clear from the B–K matrix that nearly the entire sequences are alignable using peak values $>4.0\sigma$. Two regions, ~50 residues in length and not indicated in the matrix, were alignable with peaks down to 2.5σ . The D–B matrix (Figure 2) displays two strongly conserved regions with peaks $>4.0\sigma$ (residues 104–240 and 345–474 of *hsdS* B with 74–210 and 315–444, respectively, for *hsdS* D). Since the stagger relationship of the two sets was exactly 30, the remaining regions were matched using the same stagger. All the pairwise search matrices were consistent in the regions suggested to be structurally homologous.

The symmetric *hsdS* D–D matrix is shown in Figure 3. The strong peaks on either side of the main diagonal point to a repeating domain in the protein. The D–B and B–K matrices show a similar phenomenon such that the entire N-terminal half

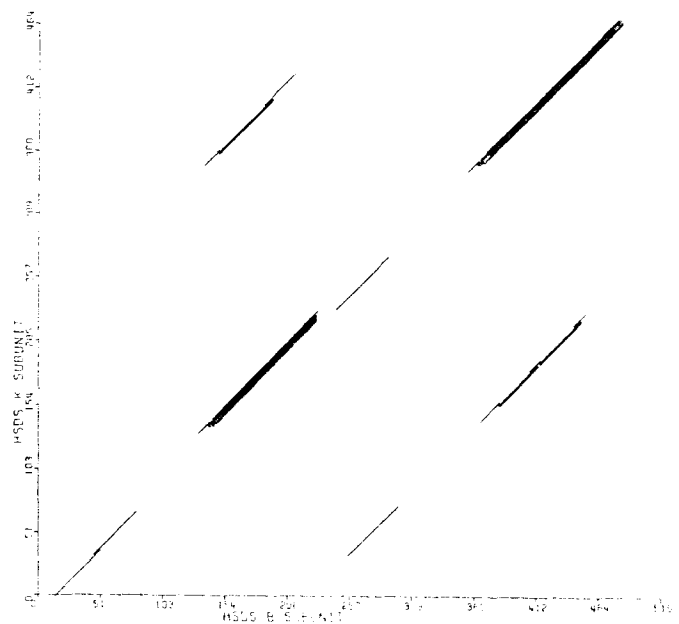


Fig. 1. The structural homology search matrix for the *hsdS* subunits from *E. coli* B (subunit *hsdS* B) and K (subunit *hsdS* K). The search window was 30 residues in length. Line designations selected to indicate the standard deviation (σ) fraction of the search values (S) are $4.1\sigma \leq S < 5.0\sigma$ (thin line), $5.0\sigma \leq S < 5.5\sigma$ (thick line), $5.5\sigma \leq S < 6.0\sigma$ (bars), $6.0\sigma \leq S < 15.0\sigma$ (overlapping circles). The symbols were placed over the entire 30-residue probe segment. The symbol corresponding to the higher fraction was chosen where overlap was possible.