The EMBO Journal vol.4 no.5 pp.1351-1355, 1985

Evidence for a repeating domain in type I restriction enzymes

Patrick Argos[1]

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

[1]On 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 between 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 σ. Two regions, ~50 residues in length and not indicated in the matrix, were alignable with peaks down to 2.5 σ. The D – B matrix (Figure 2) displays two strongly conserved regions with peaks >4.0 σ (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

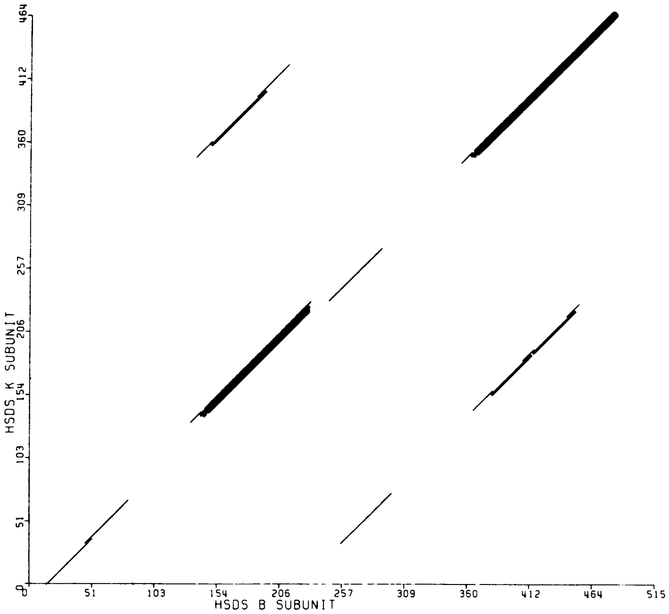

Fig. 1. The structural homology search matrix for the *hsdS* subunits from *E. coli* B (subunit *hsdS* B) and K (subunit *hsdS* K). The search window was 30 residues in length. Line designations selected to indicate the standard deviation (σ) fraction of the search values (S) are 4.1σ ≤ S < 5.0σ (thin line), 5.0σ ≤ S < 5.5σ (thick line), 5.5σ ≤ S < 6.0σ (bars), 6.0σ ≤ S < 15.0σ (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.

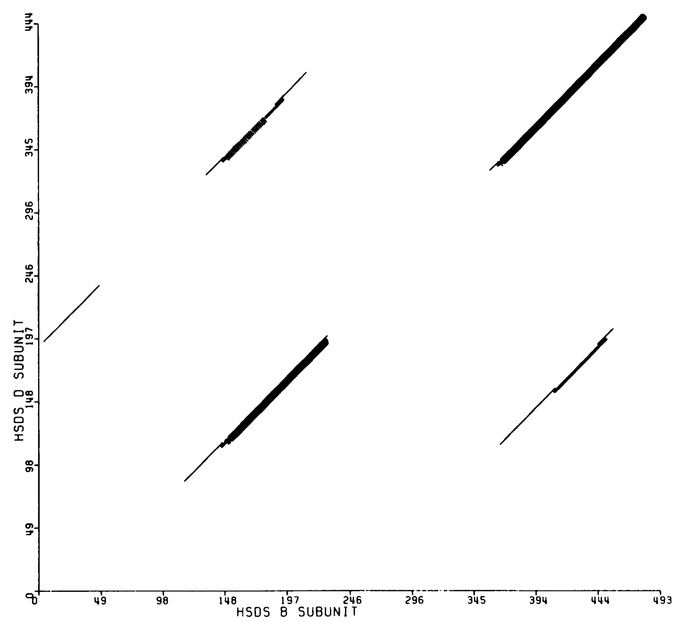

Fig. 2. As Figure 1, except for the *hsdS* subunits from *E. coli* D (subunit *hsdS* D) and B (subunit *hsdS* B).

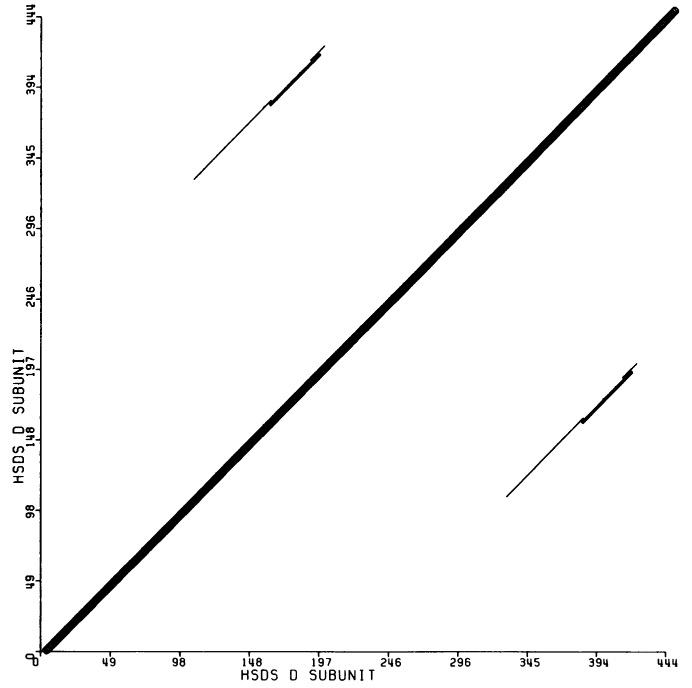

Fig. 3. As Figure 1, except for the *hsdS* subunits from *E. coli* D against itself (subunit *hsdS* D).

of the *hsdS* proteins could be matched with the C-terminal half. The residues in the region of the D – B comparison where no 4.0 σ peaks occurred (residues 47 – 130 in *hsd* B and 238 – 322 in *hsd* D) were easily matched visually as the spans differed in length by only one residue. The final alignment for all three proteins and their repeating domains is shown in Figure 4.

Table I lists the mean correlation coefficients of aligned residues over six physical characteristics for the various protein pairs. The *hsdS* B and K proteins display the strongest relationship at 0.53 for 445 matched residues while *hsdS* D with B or K is at the 0.37 level. The repeating domains correlate at the 0.25 level which is clearly above random (0.00). The comparable correlation for amino acids aligned by superposition of the C_α atoms in the known NAD-binding domain structures of alcohol and glyceraldehyde-3-phosphate dehydrogenases is only 0.12 (Otto *et al.*, 1980).

Table I. Mean correlation coefficients over six physical parameters for aligned *hsdS* residues (Figure 4)

	D	B	K	Dr	Br	Kr
D	X	441	430	210	210	214
B	0.37	X	445	211	211	215
K	0.37	0.53	X	208	208	212
Dr	0.25	0.29	0.27	X	X	X
Br	0.18	0.25	0.28	X	X	X
Kr	0.18	0.21	0.25	X	X	X

The numbers of matched residues for each comparison are given in the upper right. The 'r' designation refers to the repeating domain portions of the D, B, and K proteins (match positions 32 – 267 of Figure 4)

The mean smoothed secondary structural potential plots are illustrated in Figure 5. Figure 4 shows the resulting prediction for residues in each alignment position (helix, β-strand, turn and coil). The predictions suggest an asymmetric distribution in the structural types within each domain. Alignment positions 1 – 174 predict as 16% helix, 33% β-strand, 29% turn, and 22% coil, while sites 175 to 267 show 57% helix, 3% β-strand, 25% turn and 15% coil.

The amino acids within two spans of the *hsdS* D, B and K proteins are nearly identical: alignment positions 173 – 234 and 400 – 494 (Figure 4). There are two segments within the repeating domains that display especially strong conservation: alignment positions 173 – 197 and 212 – 234. The predictions within these latter two spans are largely helical and turn.

Discussion

It is suggested that the recognition subunit of type I restriction enzymes may interact with double-stranded B DNA in a fashion similar to that proposed for DNA repressor and activator proteins (Steitz *et al.*, 1982; Ohlendorf *et al.*, 1983; Pabo and Lewis, 1982; Steitz and Weber, 1984). The three-dimensional structures of repressor proteins (Ohlendorf *et al.*, 1983) and the catabolite gene activator protein (CAP) (McKay and Steitz, 1981) show a two-helix motif that is postulated to be involved in B DNA sequence recognition. Since the proteins act as dimers, it has been proposed through model studies that each of the helical motifs penetrates successive major grooves on the same side of the double-stranded B DNA, thereby explaining the rotational symmetry of their recognition sequences. There are ~ 11 nucleotides separating points of intimate contact between DNA and monomers while the furthest interaction points for the dimers are separated by 15 bases. In the repressors, it is suggested that a Gln and Ser at the second helical N terminus provide the most extensive DNA contacts (Ohlendorf *et al.*, 1983; Sauer *et al.*, 1982). Many of the repressor molecules as well as CAP display a two-domain structure with the larger involved in dimer association and the smaller used for DNA interaction (cf. McKay and Steitz, 1981). For example, the three-dimensional structure of CAP (Steitz and Weber, 1984) displays an N-terminal domain of ~ 15 kd consisting largely of β-structure for dimer contact while a C-terminal helical domain near 5 kd is proposed to interact with DNA. A similar suggestion has been recently made for resolvase (Abdel-Meguid *et al.*, 1984).

The *hsdS* restriction subunit may well follow the repressor model. The repeating domains could act as pseudo-dimers providing DNA contact points in successive major grooves on the same DNA side. However, without the exact dimer, the DNA recognition site would not be expected to show rotational symmetry as is observed. The specific recognition sequences for *Eco*B and *Eco*K (Kan *et al.*, 1979; Sommer and Schaller, 1979) are

*Eco*B 5′-T-G-A-N-N-N-N-N-N-N-N-T-G-C-T
*Eco*K 5′-N-A-A-C-N-N-N-N-N-N-G-T-G-C-N

where N can be any nucleotide. The number of nucleotides from the shared A to the common G is 11 while from end-to-end there are 15 nucleotides. The 11 and 15 nucleotide lengths separating putative intimate and end-to-end contacts, respectively, are consistent with the repressor models.

The N-terminal region of each *hsdS* domain predicts mostly as β-strand for roughly a 15-kd segment. The C-terminal portion of each domain is predicted as largely helical; however, this segment is near 10 kd doubling that of some of the repressors, CAP and resolvase. It is possible that the two regions of strong homology in the C-terminal domain segment (alignment positions 173 – 197 and 212 – 236 of Figure 4) would provide two DNA interaction sites: one for the specific recognition sequence to be methylated and the other for restriction. Electron microscopic studies (Rosamond *et al.*, 1979; Yuan *et al.*, 1980) suggest that the DNA loops out as a supercoil structure while it is attached at the distinct recognition and cleavage sites of the *hsd* molecule. The *hsdM* subunit would be necessary along with *hsdS* for methylation after recognition while the *hsdR* subunit would provide energy necessary for restriction and translocation (Endlich and Linn, 1981). It is also noteworthy that both of the strongly-conserved putative helical regions contain about the same number of residues (~ 25) as the repressor two-helix motif and show conserved Glns (alignment positions 181 and 214 in Figure 4) in analogy with the repressors. Furthermore, the second region also exhibits a conserved Gly at position 220 and Ala at 223, also found in the repressors (Ohlendorf *et al.*, 1983). However, no compelling homologies could be found between these two regions and the two-helix motif spans in the repressors or CAP.

Since the putative *hsdS* structure would be a pseudo-dimer, the specificity at one end of the DNA recognition site would be provided by one of the two repeating domains while the remaining domain would allow recognition at the other end of the nucleic acid specificity site. A recent genetic experiment of Fuller-Pace *et al.* (1984) supports this contention. Their work involved the recombination of a genetic segment that codes for roughly one-half of the *hsdS* protein from *Salmonella typhimurium* with a genetic span from *Salmonella potsdam* coding for the other half of its *hsdS* protein. The hybrid protein acquired the DNA specificity at one end of the recognition from one of the *Salmonella* strains, while the specificity at the other end of the

```
      1       10        20        30        40        50
D                                (1)MSAGKLPVDWKTVELGEL--
B  (1)MSFNSTSKELIEQNINGLLSIHDSWLRISMDSVANITNGFAFKSSEFNN-
K                 (1)MSAGKLPEGWVIAPVSTVTTLIRGVTYKKEQAINY
Dr                              (223)QVADIASKLK-SPLDYPNT
Br                              (253)GVVQPGDDIK-DGIELIRV
Kr                              (238)GLSSKPNESG-VGHPILRI
                                        +  :   +    ::+:
      cctttttaaaaaatttcccccccccbbbbbbcbbbttttttccccccbbb

              60        70        80        90      100
D     IKLSTGKLDANAADNDGQYPFFTCAESVSQINSW-----AFDTSAVLLAGNGS
B     -RKDGVPLIRIRDVLKGNTSTYYSGQIPEGYWVY-----PEDLIVGMDGDFNA
K     LKDDYLPLIRANNIQNGKFDTTDLVFVPKNLVKESQKISPEDIVIAMSSGSKS
Dr    I-HLAPNHIESWTGKASG-YQTILEDGVTSAKHEF--YTGQ--IIYSKIRPYL
BR    C-DINDGEVDLNHLRKIS-KEIDLQYKRSKVRKND--ILVT--IVGAIGRIGI
Kr    S-SVRAGHVDQNDIRFLECSESELNRHKLQDGDLL--FTRYNGSLEFVGVCGL
      :     :++   :    :    +     +      :   : :+ +    :
      btttttaaaaaattttttbbbbbcccctttaaaaabbbtttbbbbbttttb

          110      120       130      140       150
D     FSIKKYTGKFN---AYQRTYVIEPILIKTE--FLYWLLR--GNIKKITENGR-
B     TIWCSEPALLN-----QRVCKIEVQEDKYNKRFFYHALP--GYLSAINANTS-
K     VVGKSAHQHLPFECSFGAFCGVLRPEKLIFSGFIAHFTKSSLYRNKISSLSA-
Dr    CK--VTIATFDGMCS---ADMYPINSKIDTHFLFRWML-TNTFTDWASNAED-
Br    VR--EDINVNIARAV---ARISPEYKIIVPMFLHIWLS-SPVMQTWLVQSSK-
Kr    LK-KLQHQNLLYPDK---LIRARLTKDALPEYIEIFFS-SPSARNAMMNCVKT
      +     : +        +: +     ::  +++++  :  + :+ :
      bbbcccccccccccccbbbbbbaaaaaacccbbbbbbttttaaaaatttt

        160      170       180      190       200      21
D     GSTIPYIRKGDITDISVALPSPSEQTLIAEKLDTLLAQVESTKARLEQIPQIL
B     SVTVKHLSSRTLQDTLLPLPPLAEQKIIAEKLDTLLAQVDSTKARLEQIPQIL
K     GANINNIKPASFDLINIPIPPLAEQKIIAEKLDTLLAQVDSTKARFEQIPQIL
Dr    RTVLPKINQKDLSEIPVPTPPLPEQHEIVRRVEQLFAYADTIEKQVNNALARV
Br    EVARKTLNLKDLKNAFVPLPSIEEQHEIVRRVEQLFAYADSIEKQVNNALARV
Kr    TSGQKGISGRDIKSQVVLLPPVKEQAEIVRRVEQLFAYADYIEKQVNNALARV
         :  *  :*+ :: *****+ ** *+ +++ *+* +**   ++++  +
                         *   ** *      * *

      tbbbbbbttttccbbbbbttttaaaaaaaaaaaaaacccaaaaaaccaaa

      0       220       230       240       250      260
D     KRFRQAVLTFAMNGELTKEWRSQNNNPAFFPAEKNSLKQFRNKELPSIPNNWS
B     KRFRQAVLAAAVTGRLTKEDKDFITKKVELDNYKILIPEDWSETILNNIINTQ
K     KRFRQAVLGGAVNGKLTEKWRNFEPQHSVFKKLNFESIL--------------
Dr    NNLTQSILAKAFRGELTAQWRAENPDLISGENSAAALLEKIKAERAASGGKKA
Br    NNLTQSILAKAFRGELTAQWRAENPDLISGENSAAALLEKIKAERAASGGKKA
Kr    NNLTQSILAKAFRGELTAQWRAENPDLISGENSAAALLEKIKAERAASGGKKA
        + * +** *+ **** +** +**+:+ :* +:*:* :  * :  :
          *  *  *  * **

      aaaabbbaaaaatttaaaaaaattttcccttttcaaaaaaaaaatttttt

           270      280      290       300      310
D     WMRFDQVADIASKLKSPLDYPNTIHLAPNHIESWTGKASG-YQTILEDGVTSA
B     RPLCYGVVQPGDDIKDGIELIRVCDINDGEVDLNHLRKIS-KEIDLQYKRSKV
K     TELRNGLSSKPNESGVGHPILRISSVRAGHVDQNDIRFLECSESELNRHKLQD
Dr    SRKKS(444)
Br    SRKKF(474)
Kr    SRKKF(464)
      : ::  +        ++ +     +   ++    +    +  *+
      ccccc

        320      330       340      350      360       3
D     KHEFYTGQ--IIYSKIRPYLCK-VTIATFDGMCSADMYPINSKIDTHFLFRWM
B     RKNDILVT--IVGAIGRIGIVR-EDINVNIARAVARISPEYKIIVPMFLHIWL
K     GDLLFTRYNGSLEFVGVCGLLKKLQHQNLLYPDKLIRARLTKDALPEYIEIFF
          +     +       +++   +       +       +   +++ ++

      70      380       390       400       410      420
D     LTNTFTDWASNAES-RTVLPKINQKDLSEIPVPTPPLPEQHEIVRRVEQLFAY
B     SSPVMQTWLVQSSK-EVARKTLNLKDLKNAFVPLPSIEEQHEIVRRVEQLFAY
K     SSPSARNAMMNCVKTTSGQKGISGRDIKSQVVLLPPVKEQAEIVRRVEQLFAY
       +  + ++ +         +  +*+   *  * + **+***********

           430       440       450       460       470
D     ADTIEKQVNNALARVNNLTQSILAKAFRGELTAQWRAENPDLISGENSAAALL
B     ADSIEKQVNNALARVNNLTQSILAKAFRGELTAQWRAENPDLISGENSAAALL
K     ADYIEKQVNNALARVNNLTQSILAKAFRGELTAQWRAENPDLISGENSAAALL
      ** ***********************************************

         480      490 494
D     EKIKAERAASGGKKASRKKS(444)
B     EKIKAERAASGGKKASRKKF(474)
K     EKIKAERAASGGKKASRKKS(464)
      *******************
```

Fig. 4. The amino acid alignment of the *hsdS* D, B and K proteins along with their repeating domains (Dr, Br and Kr). A (**) symbol indicates an amino acid match in all six sequences while a (*) refers to residue identity when only three sequences are aligned or to four and five residue identity when six sequences are compared. A (+) symbol refers to conserved residues in three proteins when three sequences are aligned, or in five and six proteins when six sequences are aligned according to the following groups: (K,R); (S,T); (P,G); (Q,N,E,D); and (H,Y,W,F,I,L,V,M,C,A) with the latter large group corresponding to hydrophobic amino acids. A (:) indicates four of six residues are conserved. The regions predicted to be in a given secondary structural state (a, helix; b, strand; t, turn; c, coil) are appropriately annotated. The numbering scheme counts all positions, including gaps, in the alignment of all the sequences. The sequence number of the first and last residues for a particular protein are given in parentheses.

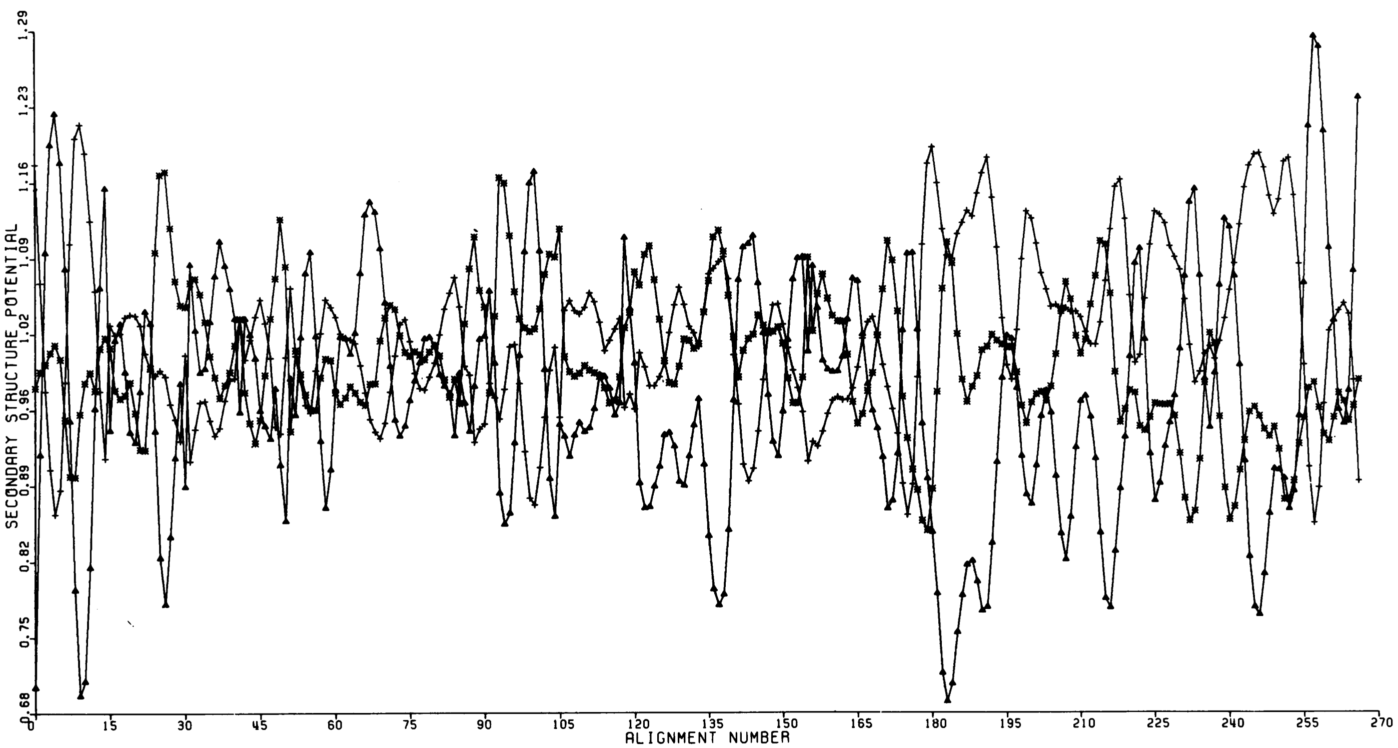

Fig. 5. The mean secondary structural potential plots for the aligned sequences of Figure 4. The alignment number refers to the alignment position annotated in Figure 4. Results are given for only the first 267 positions as the predictions for the C-terminal positions would be similar. The plot symbols indicate the helix (+), β-strand (*), and turn (△) potential curves.

DNA site was that of the other *Salmonella* species. The recombination site occurred roughly near the mid-point of the *hsdS* primary sequence probably in the vicinity of alignment positions 230 – 250 of Figure 4 (Fuller-Pace *et al.*, 1984). This latter stretch occurs just C-terminal to the strongly conserved middle region when considering the entire *hsdS* primary structures. Within the repeating domains the best conservation exists between this middle span (match positions 173 – 236 in Figure 4) and the C-terminal region of the *hsdS* protein (match positions 400 – 494 of Figure 4). The C-terminal span is also strongly conserved when considering the homology amongst the entire *hsdS* amino acid sequences. It is suggested here that the two regions, best conserved amongst the complete proteins and best conserved over the repeating domains, would provide the structural configuration necessary for interaction at the two specificity sites in the DNA recognition span. A similar situation is found in the repressor proteins (Pabo and Lewis, 1982; Ohlendorf *et al.*, 1983) where the sequence regions displaying the strongest conservation also provide the basic tertiary configuration for nucleic acid interaction despite differing DNA specificities in the various repressor molecules.

Gough and Murray (1983) suggest that the conserved C-terminal regions of the *hsdS* proteins may be important for recognition of the other subunits as methylases and ATPases are interchangeable in various bacterial systems. It is suggested here that this putative subunit recognition region would confine itself to alignment positions 463 – 494 as the remaining strong C-terminal homologies may exist for DNA recognition. Gough and Murray (1983) emphasized that the two regions of strongest homology in the *hsdS* proteins were for subunit interaction while in this report it is proposed that the homologies primarily exist for recognition of the DNA methylation and cleavage sites.

Materials and methods

The protein sequences were correlated by comparing every possible span of length L residues in one protein with all such spans in the second protein. Two scoring procedures were used. The first was based on the Dayhoff relatedness odds matrix (Dayhoff, 1969; McLachlan, 1971; Staden, 1982) whose elements express relative weights with which amino acids substitute in aligned sequences of 71 protein families. The second scoring method involved calculation of the mean correlation coefficient for each oligopeptide comparison over six residue physical characteristics thought to be the primary forces directing protein folding (cf. Creighton, 1978): helix, sheet and turn secondary structural conformational preferences; residue polarity; and two amino acid hydrophobicity measures (hydration potential and surrounding hydrophobicity). The parameters are listed and discussed by Argos *et al.* (1983). The use of physical characteristics in comparing sequences has been previously discussed (Argos *et al.*, 1983; Argos and Siezen, 1983; Kubota *et al.*, 1981,1982). The final search matrix was constructed by averaging the scores from the two techniques after subtracting the mean value from all elements of each search matrix and then scaling the respective elements such that the sum of the differences between the respective mean matrix values and those elements greater than the mean were made equal. The resulting matrix for proteins compared here had a lower noise level and indicated longer stretches for alignment than matrices calculated from either technique.

Search matrix plots were made by attaching symbols to the matrix element values that fell within particular fractional standard deviation ranges. No matrix value $<4.0\sigma$ was considered for determining the overall alignments. In the 4.0σ to 5.5σ range, the theoretical probability of such a matrix value occurring is between 10^{-5} and 10^{-6} (McLachlan, 1971). A series of diagonally co-linear, broken lines were easily detected by visual inspection through the use of the same symbol in all L positions. The symbol corresponding to the higher σ range was allowed to dominate if symbol overlap occurred. A window search length of 30 residues resulted in minimal noise. Once the sequences had been matched, an assessment of the overall structural relatedness of the two proteins was calculated using mean correlation coefficients for all the aligned residues over the six physical characteristics.

Plots of the sequence number *versus* the conformational preference parameter (helix, β-strand, and turn) (Argos *et al.*, 1983) for a given amino acid were determined for each protein sequence and then smoothed over three cycles such that every successive group of three points (i to i+2) was fitted by a least-squares line and the value at (i+1) replaced by that calculated from the line. The smoothed curves for each potential were averaged over the aligned sequences, a procedure which should yield a better prediction than that from any one sequence (Argos *et al.*, 1976). The structural type assigned at each aligned residue position corresponded to the largest of the three mean potentials that were > 1.00, the neutral preference value (Chou and Fasman, 1974). Five such successive values were required for helix initiation and three for strands or turns. If Pro or Gly occurred at the fifth or greater position of a helically predicted region, the span was assigned as coil due to the rare appearance of such residues at these helical sites (Argos and Palau, 1982). For all other conditions (e.g., all mean potentials < 1.00), the coil structure was predicted.

Acknowledgements

The author is grateful for financial and computer assistance from the EMBL and for support and stimulating discussions from Drs. Fritz Winkler and Demetrius Tsernoglou. Preparation of the manuscript by Christine Barber is also acknowledged.

References

Abdel-Meguid,S.S., Grindley,N.D.F., Templeton,N.S. and Steitz,T.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2001-2005.

Argos,P., Schwarz,J. and Schwarz,J. (1976) *Biochim. Biophys. Acta*, **439**, 261-273.

Argos,P. and Palau,J. (1982) *Int. J. Peptide Protein Res.*, **19**, 380-393.

Argos,P., Hanei,M., Wilson,J.M. and Kelley,W.N. (1983) *J. Biol. Chem.*, **258**, 6450-6457.

Argos,P. and Siezen,R.J. (1983) *Eur. J. Biochem.*, **131**, 143-148.

Bickle,T.A., Brack,C. and Yuan,R. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 3099-3103.

Bickle,T.A. (1982) *Nucleases*, published by Cold Spring Harbor Laboratory Press, NY, USA, pp. 85-108.

Chou,P.Y. and Fasman,G.D. (1974) *Biochemistry (Wash.)*, **13**, 211-245.

Creighton,T.E. (1978) *Biophys. Mol. Biol.*, **33**, 231-297.

Dayhoff,M.O. (1969) *Atlas of Protein Sequences and Structure*, published by National Biomedical Research Foundation, Silver Springs, MD, USA.

Endlich,B. and Linn,S. (1981) in Boyer,P.D. (ed.), *The Enzymes*, Vol.**14**, Academic Press, NY, pp. 137-156.

Fuller-Pace,F.V., Bullas,L.R., Delius,H. and Murray,N.E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6095-6099.

Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.*, **166**, 1-19.

Kan,N.C., Lautenberger,J.A., Edgell,M.H. and Hutchison,C.A.,III (1979) *J. Mol. Biol.*, **130**, 191-209.

Kubota,Y., Takahashi,S., Nishikawa,K. and Ooi,T. (1981) *J. Theor. Biol.*, **91**, 347-361.

Kubota,Y., Nishikawa,K., Takahashi,S. and Ooi,T. (1982) *Biochim. Biophys. Acta*, **701**, 242-252.

McKay,D.B. and Steitz,T.A. (1981) *Nature*, **290**, 744-749.

McLachlan,A.D. (1971) *J. Mol. Biol.*, **61**, 409-424.

Ohlendorf,D.H., Anderson,W.F., Lewis,M., Palo,C.O. and Matthews,B.W. (1983) *J. Mol. Biol.*, **169**, 757-769.

Otto,J., Argos,P. and Rossmann,M.G. (1980) *Eur. J. Biochem.*, **109**, 325-330.

Pabo,C.O. and Lewis,M. (1982) *Nature*, **298**, 443-447.

Rosamond,J., Endlich,B. and Linn,S. (1979) *J. Mol. Biol.*, **129**, 619-635.

Sauer,R.T., Yocum,R.R., Doolittle,R.F., Lewis,M. and Pabo,C.O. (1982) *Nature*, **298**, 447-451.

Sommer,R. and Schaller,H. (1979) *Mol. Gen. Genet.*, **168**, 331-335.

Staden,R. (1982) *Nucleic Acids Res.*, **10**, 2951-2961.

Steitz,T.A., Ohlendorf,D.H., McKay,D.B., Anderson,W.F. and Matthews,B.W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3097-3100.

Steitz,T.A. and Weber,I.T. (1984) in McPherson,A. and Jurnak,F. (eds.), *Structural Biology*, John Wiley, NY, in press.

Yuan,R., Hamilton,D. and Burckhardt,J. (1980) *Cell*, **20**, 237-244.

Yuan,R. (1981) *Annu. Rev. Biochem.*, **50**, 285-315.

Received on 1 February 1985