Volume 322, number 2, 111–114 FEBS 12309 May 1993

The antigenic domain of flagellin from *S. paratyphi* shares a structural fold with subtilisin

N. Grewal and D.M. Salunke

National Institute of Immunology, JNU Complex, New Delhi 110 067, India

Received 9 February 1993

Bacterial flagellin has two domains: the polymerizing domain consisting of N- and C-terminal regions which are partly disordered in the monomeric state; and the central antigenic domain with compact globular structure. The polymerizing domain is highly conserved in flagellins from different species but the antigenic domain is diverse in sequence and size. Whereas the former has direct functional significance for bacterial motility, the latter has not been identified as having a specific function except for defining the distinct serotype of the bacterium. The sequence alignment of flagellin from *S. paratyphi* with proteins of known three-dimensional structure reveals significant homology of the central 265 residue stretch with the bacterial serine protease, subtilisin. This homology is evident also in the comparison of the predicted secondary structure of flagellin with the observed secondary structural features in subtilisin. The deletions/insertions arising due to optimal alignment of the two proteins occur on the surface loops in the structure. Thus, a domain of *S. paratyphi* flagellin and subtilisin appear to have similar structural folds.

Flagellin; Bacterial serine protease; Tertiary fold; Homology

1. INTRODUCTION

Flagellin, a globular protein the size of which varies for different species, is the monomer of the helical filament of the bacterial flagellum. The propellar motion of flagellar filaments, driven by a rotary motor, is responsible for bacterial motility. The significance of the structure, and interactions between flagellar monomers, to the mechanism of bacterial motility, has indeed been recognized [1]. Supramolecular details of the flagellum, including subunit interactions, and morphological features have been revealed by extensive biophysical studies [2], but the high resolution molecular description of the monomeric unit and its correlation with function has not been achieved. The protein polymerizes, forming infinitely long helical assemblages, and is therefore not amenable to single crystal X-ray crystallographic studies at high resolution. Therefore, structure prediction by the knowledge-based computer modeling [3] approach is particularly relevant. It is becoming more and more apparent, from the large number of protein crystal structures available, that the tertiary structures of proteins can be classified into a finite number of sub-structural modules which are used over and over again in different combinations and contexts. Incorporation of the redundancy, arising due to the repeated use of a finite number of sub-structural modules in diverse proteins, has enhanced reliability of the theoretical approaches to structure determination. Application of such an approach has led us to show that the structural fold of the non-conserved outer domain of flagellin from *S. paratyphi* [4] is similar to that of the subtilisin class [5] of serine proteases.

Correspondence address: D.M. Salunke, National Institute of Immunology, JNU Complex, New Delhi 110 067, India. Fax: (91)(11) 686-2125.

2. MATERIALS AND METHODS

A sequence database corresponding to the unique protein entries of known 3-D structures [6] was created and all the sequence homology searches were restricted to this database only. The initial homology searches were carried out using the mutation data matrix [7] in the NBRF sequence analysis software package. A modified unitary matrix, where the chemically similar amino acids are treated equivalent (viz. A = G, R = K, N = D = Q = E, V = I = L, F = W = Y and S = T), was used for subsequent optimization of the alignment. The secondary structural propensities were analyzed using the algorithms of Garnier et al. [8]. The molecular modeling software, QUANTA version 2.1 (Polygen Corp. Inc.), was used for model building applications. The model was subjected to energy minimization by the conjugate gradient method in CHARMm [9].

3. RESULTS

The flagellin monomer from *S. paratyphi* is a single polypeptide chain of 488 residues. The sequence alignment with proteins available in the Brookhaven protein database led to the identification of significant homology of a segment of flagellin with subtilisin, leaving about 25 residues on either termini. The 242 residue flagellin segment, corresponding to residues fla202–444, aligns with subtilisin segment sbc23–257, with about 30% chemically equivalent residues of which 19% are

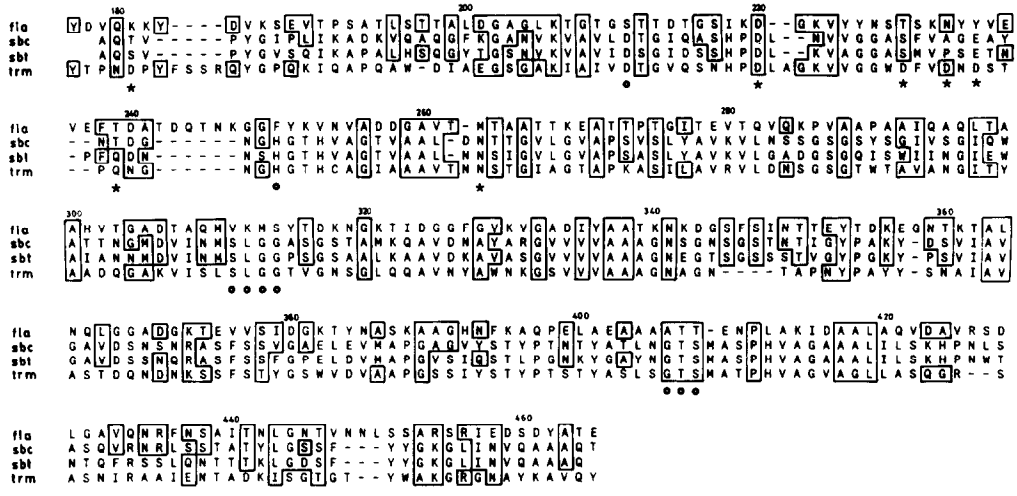

Fig. 1. Sequence alignment of flagellin from *S. paratyphi* (fla) with subtilisin Carlsberg (sbc), subtilisin BPN′ (sbt) and thermitase (trm). The numbering at the top corresponds to flagellin. Chemically equivalent residues are highlighted by boxes. The residues involved in defining substrate binding site are identified by (○) and those involved in calcium binding site are identified by (*).

identical. This alignment could then be extended on both the N- and C-terminal sides covering the entire subtilisin sequence (Fig. 1). Out of the four insertions, two (fla183 and fla448–450) are very close to the terminal ends of the domain. The homology can also be extended to thermitase (trm) and subtilisin BPN (sbt), two other proteins belonging to the subtilisin class of serine proteases. The homologous sequence in flagellin, fla180–465, corresponds very closely to the non-conserved region, the antigenic domain, that defines the serological properties of the molecule.

Homology of the antigenic domain of flagellin with subtilisin was also observed in the secondary structural features of the two molecules. The secondary structural propensities of the antigenic domain of flagellin were compared with the regular secondary structural elements identified from the X-ray crystal structure of subtilisin [10] (Fig. 2). The correspondence between the two is significant. In particular, the flagellin sequences corresponding to 5 strands of the core β-sheet of subtilisin have strong β-sheet propensities. All the helices in subtilisin also find corresponding sequences in flagellin with high helical propensity except for the long solvent-exposed helix, scb132–146.

The identification of homology between the two proteins led to the construction of the 3-D model for the antigenic domain using the backbone conformation of subtilisin as the template (Fig. 3). As the insertions were occuring on the exposed loop, these were modeled with the single constraint that the polypeptide can be closed through a peptide bond without introducing any steric hindrance. Extensive refinement of the model was not considered since the observed homology is in the range indicative of a common structural fold [11]; a correspondence in the atomic details is not anticipated.

4. DISCUSSION

It has been established by fibre diffraction [12], electron microscopic image reconstruction [13] and thermodynamic [14,15] studies of flagellar preparations from different bacteria that the subunit protein, flagellin, can be divided into two structurally distinct domains. The domain buried in the intact flagellum, which is involved in stabilizing the intersubunit interactions, can be described as the polymerizing domain, whereas the one that is exposed outwards and is responsible for defining serotype can be called the antigenic domain. Comparison of the 3-D images reconstructed from electron micrographs of the *C. cresentus* flagella, lacking an antigenic domain, and *S. typhimurium* flagella reveals that this domain is not involved in polymerization [13]. This has also been shown by studies involving systematic deletion mutations in the central non-conserved regions of the sequence [16,17].

In comparison to the sequence homology of the evolutionarily related proteins belonging to a functionally related family, the observed homology between flagellin and subtilisin is weak, however, it is interesting to note that the homology extends over a length of more than 250 residues with very few breaks. Also, as can be seen from Fig. 3, the breaks arising in flagellin correspond to the surface exposed loop regions in the subtilisin fold. The feature which makes this homology more significant is that the homologous sequence, fla180–465, corresponds very closely to an entire domain which has been defined independently [14]. The two termini (fla1–179 and fla466–488) fold together constituting the polymerizing domain, and the intervening region (fla180–465) is organized to form the antigenic domain. If the two domains are structurally distinct, as seen in the fibre

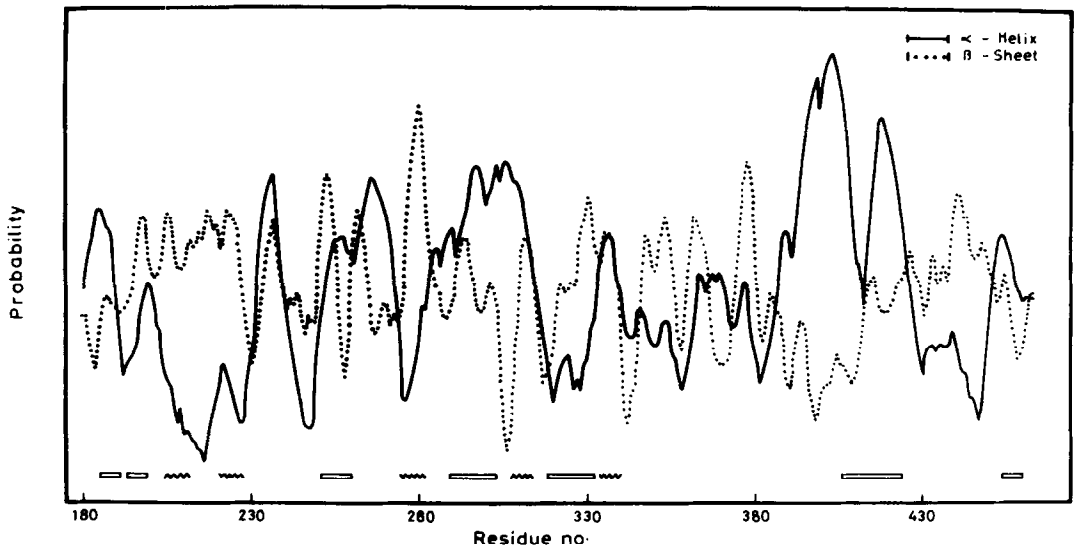

Fig. 2. The regular secondary structural propensities of the antigenic domain (fla180–466) of flagellin. The probabilities are plotted to an arbitrary scale. The observed secondary structural elements corresponding to the homologous segments in subtilisin are indicated as rod (helix) and zigzag line (β-strand) for corresponding residues.

diffraction electron density map of Namba et al. [12], the two ends of the antigenic domain ought to come into close proximity exposed to solvent on the same side of the protein. This is consistent with the observed juxtaposition of the N- and C-terminal ends in the model based on the subtilisin structure. The domain structure developed on the basis of the homology with subtilisin, although corresponding closely, does not exactly overlap with that defined earlier. Often the domain boundaries defined on the basis of controlled digestion by proteolytic enzymes have inherent uncertainty in them since a loop protruding out within a single domain can be as accessible to protease digestion as the linking sequence between the two domains. Indeed, the previously defined cleavage site [14] corresponds to an extensively exposed loop (fla422–fla428) in our model.

The antigenic domain, since it is responsible for defining distinct serotypes, ought to be very different in bacteria belonging to different species. Indeed, the size of the domain varies from a few residues in the case of *C. cresentus* [13] and *B. subtilis* [18] to over 250 residues in the case of *S. paratyphi*. Therefore, this domain may not necessarily have a common structural motif among flagellins from various species. The doubly wound α/β

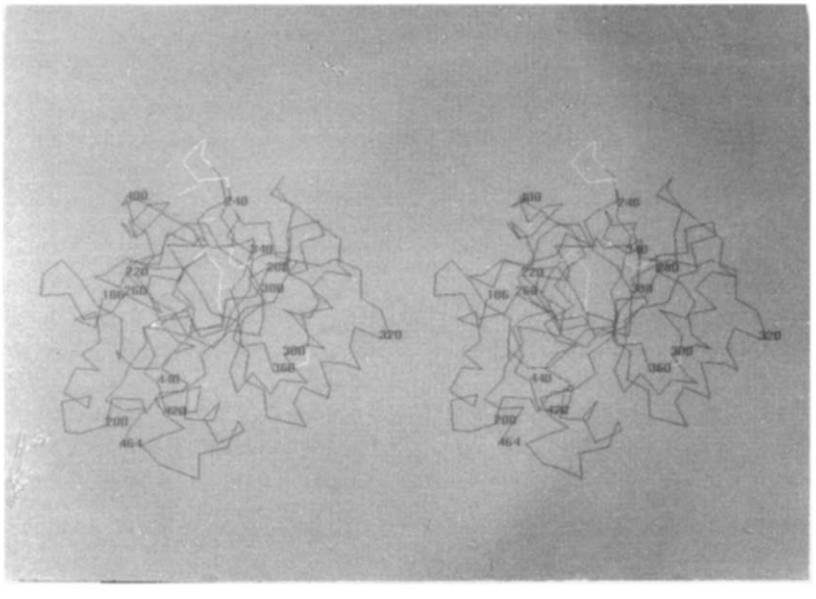

Fig. 3. The α carbon stereoscopic drawing of the model of flagellin antigenic domain based on the subtilisin structure. The inserted regions are highlighted with white.

fold, with a parallel β-sheet wrapped over by α-helices, predicted for the antigenic domain of *S. paratyphi*, occurs in diverse proteins with unrelated functions. The broad structural features of this fold were originally associated with the NAD binding proteins [19]. Unlike in the case of the TIM-barrel [20] or the 4-helix bundle motif [21,22], it has not been possible to recognize any consensus sequence or the signature residues that can be attributed to this structural motif. One feature, however, is obvious in that the core β-sheet has the charged side chains at the ends of the strands leaving primarily hydrophobic residues within the sheet, as anticipated by structural considerations [19]. The fact that the 5-stranded twisted β-sheet is at the core of the structure is significant. It is this β-sheet that is most conserved between the flagellin antigenic domain and subtilisin in terms of secondary structural propensities.

Thermodynamic and circular dichroism studies show that the terminal regions have high flexibility and are probably α-helical on folding [14,23]. Analysis of this region of the sequence led Federov and Afimov [24] to identify a sequence motif corresponding to the coiled-coil packing of α-helices. They have proposed 4-helix bundle folding for this domain in the N-terminal region and an $\alpha\alpha$-hairpin in the C-terminal region. Thus the assembly of α-helices in the polymerizing domain, as proposed by Federov and Afimov, and the α/β subtilisin type folding of the antigenic domain emerging from our observations, together provide plausible structural organization of the entire flagellin molecule.

While extending the structural relationship between the subtilisin family of proteins and flagellin to its function, two important aspects are (i) the active site geometry and (ii) the metal coordination. We find that not all the residues that are directly involved in defining the substrate binding site of subtilisin are conserved in flagellin. This is also true about the calcium binding site (Fig. 1). Two proteins having a common structural fold but distinct functional properties is not so very surprising. For example, the 4-helix bundle motif has been used in the oxygen transport protein, hemerithrin [25], and also in the structural protein of a plant virus, the subunit protein of TMV [26]. Another such example of weak seqeunce homology indicating a common structural fold has been reported recently [27], where hsp70 has been modeled on the basis of the X-ray crystal structure of HLA.

Why does the bacterial protease and a domain of the bacterial flagellin have common structural folds? The observed similarity of fold between subtilisin and one domain of flagellin may just be another example suggesting that nature uses a finite number of structural modules again and again in different combinations and contexts and endows them with diverse functions. On the other hand, this structural homology may indeed provide a clue to a possible functional role for this domain besides just defining serological properties.

Acknowledgements: We thank Drs. A. Qadri and M.R.N. Murthy for useful discussions. Financial support from Department of Biotechnology (Government of India) is gratefully acknowledged.

REFERENCES

[1] Asakura, S. (1970) Adv. Biophys. 1, 99–155.
[2] DeRosier, D.J. (1992) Curr. Opin. Str. Biol. 2, 280–285.
[3] Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E. and Thornton, J.M. (1987) Nature 326, 347–352.
[4] Wei, L. and Joys, T.M. (1985) J. Mol. Biol. 186, 791–803.
[5] Wells, J.A., Cunningham, B.C., Graycar, T.P. and Carter, P. (1987) Cold Spring Harber Symp. Quant. Biol. 52, 647–652.
[6] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[7] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 524–545.
[8] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[9] Brooks, B.R., Broccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) J. Comput. Chem. 4, 187–217.
[10] McPhalen, C.A., Schnebli, H.P. and James, M.N.G. (1985) FEBS Lett. 188, 55–58.
[11] Chothia, C. and Lesk, A.M. (1986) EMBO J. 5, 823–826.
[12] Namba, K., Yamashita, I. and Vonderviszt, F. (1989) Nature 342, 648–654.
[13] Trachtenberg, S. and DeRosier, D.J. (1988) J. Mol. Biol. 202, 787–808.
[14] Vonderviszt, F., Uedaira, H., Kidokoro, S. and Namba, K. (1990) J. Mol. Biol. 214, 97–104.
[15] Kostyukova, A.S., Pyatibratov, M.G., Filimonov, V.V. and Fedorov, O.V. (1988) FEBS Lett. 241, 141–144.
[16] Kuwajima, G. (1988) J. Bacteriol. 170, 3305–3309.
[17] Homma, M., Fujita, H., Yamaguchi, S. and Iino, T. (1987) J. Bacteriol. 169, 291–296.
[18] Delange, R.J., Chang, J.Y., Shaper, J.H., Martinez, R.J., Komatsu, S.K. and Glazer, A.N. (1973) Proc. Nat. Acad. Sci. USA 70, 3428–3431.
[19] Richardson, J. (1981) Adv. Protein Chem. 34, 167–339.
[20] Rice, P.A., Goldman, A. and Steitz, T.A. (1990) Proteins 8, 334–340.
[21] Banner, D.W., Kokkinidis, M. and Tsernoglou, D. (1987) J. Mol. Biol. 196, 657–675.
[22] Gupta, H.M. and Salunke, D.M. (1992) Curr. Sci. 62, 374–376.
[23] Fedorov, O.V., Kostyukova, A.S. and Pyatibratov, M.G. (1988) FEBS Lett. 241, 145–148.
[24] Fedorov, O.V. and Efimov, A.V. (1990) Protein Eng. 3, 411–413.
[25] Smith, J.L., Hendrickson, W.A. and Addison, A.W. (1983) Nature 303, 86–88.
[26] Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R. and Klug, A. (1978) Nature 276, 362–368.
[27] Rippman, F., Taylor, W.R., Rothbard, J.B. and Green, N.M. (1991) EMBO J. 10, 1053–1059.