J. Mol. Biol. (1994) 241, 110–124

Filamentous Hemagglutinin of *Bordetella pertussis*

A Bacterial Adhesin Formed as a 50-nm Monomeric Rigid Rod Based on a 19-residue Repeat Motif Rich in Beta Strands and Turns

A. M. Makhov[1], J. H. Hannah[2], M. J. Brennan[2], B. L. Trus[1,3]
E. Kocsis[1], J. F. Conway[1], P. T. Wingfield[4], M. N. Simon[5] and A. C. Steven[1]†

[1]*Laboratory of Structural Biology*
National Institute for Arthritis, Musculoskeletal and Skin Diseases
[2]*Division of Bacterial Products, Food and Drug Administration*
[3]*Computational Biology and Engineering Laboratory*
Division of Computer Research and Technology and
[4]*Protein Expression Laboratory, Office of the Director*
National Institutes of Health, Bethesda, MD 20892, U.S.A.
[5]*Department of Biology, Brookhaven National Laboratory*
Upton, NY 11973, U.S.A.

The filamentous hemagglutinin (FHA) of *Bordetella pertussis* is an adhesin that binds the bacteria to cells of the respiratory epithelium in whooping-cough infections. Mature FHA is a 220 kDa secretory protein that is highly immunogenic and has been included in acellular vaccines. We have investigated its structure by combining electron microscopy and circular dichroism spectroscopy (CD) with computational analysis of its amino acid sequence. The FHA molecule is 50 nm in length and has the shape of a horseshoe nail: it has a globular head that appears to consist of two domains; a 35 nm-long shaft that averages 4 nm in width, but tapers slightly from the head end; and a small, flexible, tail. Mass measurements by scanning transmission electron microscopy establish that FHA is a monomer. Its sequence contains two regions of tandem 19-residue pseudo-repeats: the first, of 38 cycles, starts at residue 344; the second, of 13 cycles, starts at residue 1440. The repeat motifs are predicted to consist of short β-strands separated by β-turns, and secondary structure measurements by CD support this prediction. We propose a hairpin model for FHA in which the head is composed of the terminal domains; the shaft consists mainly of the repeat regions conformed as amphipathic, hyper-elongated β-sheets, with their hydrophobic faces apposed; and the tail is composed of the intervening sequence. Further support for the model was obtained by immuno-labeling electron microscopy. The 19-residue repeats of FHA have features in common with the leucine-rich repeats (LRRs) that are present in many eukaryotic proteins, including some adhesion factors. The model is also compared with the two other classes of filamentous proteins that are rich in β-structure, i.e. viral adhesins and two β-helical secretory proteins. Our proposed structure implies how the functionally important adhesion sites and epitopes of FHA are distributed: its tripeptide (RGD) integrin-binding site is assigned to the tail; the putative hemagglutination site forms part of the head; and two classes of immunodominant epitopes are assigned to opposite ends of the molecule. Possible mechanisms are discussed for two modes of FHA-mediated adhesion.

Keywords: pertussis; filamentous hemagglutinin; adhesin; leucine-rich repeats; beta-helix proteins

1. Introduction

Filamentous hemagglutinin (FHA‡) is a surface-associated protein of *Bordetella pertussis*, the Gram-negative bacterium that is the etiological agent of whooping-cough (Weiss & Hewlett, 1986; Brennan *et al.*, 1991*a*). FHA is a prominent immunogen (Di Tommaso *et al.*, 1991; Shahin *et al.*, 1992) that has been found to serve as a protective antigen

† Author to whom all correspondence should be addressed. Bldg. 6, Rm. 425, N.I.H., Bethesda, MD 20892, U.S.A.

‡ Abbreviations used: FHA, filamentous hemagglutinin; CD, circular dichroism; LRR, leucine-rich repeats; STEM, scanning transmission electron microscopy; CTEM, conventional transmission electron microscopy; SSNR, spectral signal to noise ratio; S.D., standard deviation; S.E.M., standard error of mean.

0022-2836/94/310110–15 $08.00/0

in a mouse model system (Oda *et al.*, 1984; Kimura *et al.*, 1990), and has been included as a component of several experimental acellular vaccines (Rutter *et al.*, 1988; Rappuoli *et al.*, 1992).

B. pertussis has several adhesins that cause the bacterium to bind to various kinds of cells (Locht *et al.*, 1993; Brennan *et al.*, 1991*b*). These molecules include FHA (Tuomanen & Weiss, 1985; Urisu *et al.*, 1986), pertactin (Leininger *et al.*, 1991), fimbriae (Arai & Sato, 1976; Mooi *et al.*, 1987) and pertussis toxin (Tuomanen & Weiss, 1985). Of these, FHA appears to have multiple adhesin activities: it is lectin-like, as noted by its ability to agglutinate erythrocytes (Sato *et al.*, 1983) and to promote interactions that are inhibited by certain sugars (Tuomanen *et al.*, 1988); it binds to several kinds of cultured cells (Cowell *et al.*, 1986); and it participates in the adhesion reaction of greatest clinical significance, colonization of the respiratory tract by *B. pertussis*. In the latter reaction, FHA binds both to ciliated epithelial cells (Tuomanen & Weiss, 1985) and to the integrin CR3 presented by alveolar macrophages (Relman *et al.*, 1990). Its interaction with CR3 is mediated by the fibronectin-like tripeptide RGD which is present on FHA (Relman *et al.*, 1990; Leininger *et al.*, 1992).

Cloning of the FHA gene (Brown & Parker, 1987; Relman *et al.*, 1989; Delisse-Gathoye *et al.*, 1990) has led to the identification of an exceptionally long open reading-frame corresponding to a polypeptide chain of 357 kDa (Domenighini *et al.*, 1990). The mature form of FHA has a molecular weight of about 220 kDa, and corresponds to the amino-terminal portion of the predicted gene product, which presumably undergoes proteolytic processing at some stage of its biosynthetic pathway. Electron microscopy studies of purified FHA have described a filamentous molecule, 2 nm wide and 40 to 100 nm long (Arai & Sato, 1976; Askelof *et al.*, 1982; Sato *et al.*, 1983) and heterogeneous oligomeric populations with molecular weights of up to 10^6 Da have been reported (An der Lan *et al.*, 1986).

On account of the important roles played by FHA in adhesion and pathogenesis, and its immunological properties, there is much interest in its biosynthesis and mechanism(s) of adhesion. Hitherto, its mode of action, including the nature of its association with the cell surface of *B. pertussis* (Ashworth *et al.*, 1985), has remained obscure, in part because FHA is apparently secreted from the bacterium. In this context, we have investigated the molecular structure of FHA with the particular aim of establishing a frame of reference in which to localize its major binding sites. To this end, we combined various electron microscopic experiments with optical spectroscopy and computational analysis of the amino acid sequence of FHA.

Specifically, we applied computer image processing (for reviews, see Aebi *et al.*, 1984; Frank, 1989; Steven *et al.*, 1991) to conventional as well as to scanning transmission electron micrographs (STEM) of purified 220 kDa FHA. Dissection of FHA with chymotrypsin, which generates large fragments with molecular weights of 98 kDa and 140/150 kDa, respectively (Domenighini *et al.*, 1990), was also studied by electron microscopy. In parallel, we carried out a systematic computational analysis of its amino acid sequence. Because the conformations of fibrous proteins are often based on repeat motifs, particular attention was paid to the detection of repetitive elements. Based on these observations, we formulated a molecular model in which the overall folding of the polypeptide chain into a 50 nm-long monomeric hairpin is described. This model was tested by immuno-labeling electron microscopy, using a monoclonal antibody of known specificity. In addition, its functional implications are explored.

2. Materials and Methods

(a) *Experimental material*

Purified FHA was kindly provided by Pasteur Merieux Serums et Vaccins (Marcy l'Etoile, France), and by SmithKline Beecham Biologicals (Rixensart, Belgium). The former preparation was stored as an ammonium sulfate precipitate, and the latter in 50 mM NaH_2PO_4/K_2HPO_4 (pH 7·6) with 500 mM NaCl. The electron microscopy and circular dichroism experiments were performed mainly with the former sample which was diluted to the specified concentration just before use.

(b) *Chymotrypsin digestion*

FHA was diluted in an equal volume of buffer (0·01 M Tris·HCl (pH 8·0), 0·15 M NaCl, 0·01 M EDTA), and 1/3 vol. of 0·01 mM chymotrypsin (Sigma Chemical Co., St Louis, MO) was added. The mixture was incubated at room temperature for 30 min, and the reaction stopped by transferring to ice and adding the inhibitor AEBSF (ICN Biomedicals, Costa Mesa, CA) to a final concentration of 1 mM. The resulting digest was either subjected to chromatography or analyzed by SDS-PAGE on 4% to 20% gradient gels. The identification of the resulting fragments (see Results) was verified by immunoblotting with monoclonal antibodies specific for the amino-terminal region of FHA (Leininger *et al.*, 1993).

(c) *Affinity chromatography*

Chymotrypsin-treated FHA was chromatographed on heparin Sepharose CL-6B (Pharmacia-LKB Biotechnology, Piscataway, NJ) in the following TBS++ buffer (0·015 M Tris·HCl (pH 7·5), 0·05 M NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$). In separate experiments, either 8 M urea, 1 mM dithiothreitol, or 1 % Triton X-100 detergent was added to the TBS++ column buffer. After thorough washing of the column, the protein was eluted with a gradient of 0·1 M to 0·5 M NaCl in TBS++.

(d) *Ion exchange chromatography*

Chymotrypsin-treated FHA was chromatographed on CM Sepharose CL-6B (Pharmacia-LKB Biotechnology) in 0·02 M phosphate (pH 7·3) containing 8 M urea, thoroughly washed, and eluted with a linear 0 to 0·5 M NaCl gradient in the same buffer. The same material was also chromatographed on DEAE-Sepharose CL-6B (Pharmacia-LKB Biotechnology) in 0·05 M Tris·HCl

(pH 8·0), 0·01 M EDTA, 8 M urea, thoroughly washed, and eluted with a 0 to 0·5 M NaCl gradient in the same buffer.

(e) *Gel filtration chromatography*

FHA or digests thereof were chromatographed on a Sepharose 6HR 10-30 size-exclusion column (Pharmacia-LKB Biotechnology), equilibrated at 0·5 ml/min. Sterile Dulbecco's PBS buffer was used as column buffer.

(f) *Conventional transmission electron microscopy*

For negative staining, the protein was diluted to a concentration of ~30 μg/ml in 10 mM Tris·HCl (pH 7·2), 40 mM NaCl and adsorbed to a thin carbon film according to the method of Valentine & Green (1967), and negatively stained with 1% uranyl acetate. Microscopy was performed on a Philips 400RT electron microscope. The micrographs used for computer averaging were recorded at a magnification of 36,000×.

(g) *Immuno-electron microscopy*

These experiments were conducted using a monoclonal antibody, called 2/1 : 5E, from a murine hybridoma raised against purified FHA. Its reactivity was established by Western blotting of a set of chimeric proteins consisting of FHA fragments fused with maltose binding protein (E. Leininger, M. J. Brennan, unpublished results). By this criterion, the epitope of 2/1 : 5E was determined to lie in the carboxy-terminal part of FHA, between residues 1655 and 2012. FHA and IgG antibodies were mixed at a molar ratio of 1 : 1, incubated in 0·1 M PBS (pH 7·6) at 4°C, and sampled by negative staining at various times up to 24 h. For visualization, small samples were adsorbed to glow-discharged grids and stained briefly with 2% uranyl acetate. Control incubations of FHA with purified mouse IgG were carried out under the same conditions. Several variations on this procedure were tried, but the highest level of labeling and clearest visualization of the complexes formed were obtained as described above.

(h) *Scanning transmission electron microscopy*

STEM observations were performed at the Brookhaven National Laboratory facility (Wall, 1979; Wall & Hainfield, 1986). To prepare freeze-dried specimens, FHA stored at a concentration of ~0·3 mg/ml protein was diluted 10-fold in buffer (10 mM Tris·HCl, 40 mM NaCl, pH 7·2), then adsorbed according to the "wet film" technique (Wall *et al.*, 1985) to a thin (~2·5 nm) carbon film. Immediately beforehand, tobacco mosaic virus particles were adsorbed to this film to serve as an internal mass standard. The grid was then washed 10 times with 20 mM ammonium acetate (a volatile salt). Finally, excess fluid was drawn off with filter paper, and the specimen was rapidly frozen by quenching in liquid-N_2 slush, and dried for 6 to 8 h at a constant sublimation rate (Mosesson *et al.*, 1981). After locating areas with suitable densities of adsorbed molecules, digital micrographs of 512×512 pixels were recorded at a sampling step of 1·0 nm, and an average radiation dose of 1×10^2 to 5×10^2 electrons/nm^2.

For negative staining STEM microscopy, FHA preparations were adsorbed to thin carbon substrates as described above, then stained with 1% uranyl acetate. Micrographs used for image averaging were recorded at sampling steps of 0·25 and 0·5 nm, i.e. at 2 to 4-fold higher magnifications than with the unstained data, and at correspondingly higher electron doses.

(i) *Image processing*

Conventional transmission electron microscopy (CTEM) negatives were digitized with a Perkin Elmer 1010MG flatbed microdensitometer at a scanning rate corresponding to 0·39 nm per pixel at the specimen. General image processing and display operations were performed using the PIC system of software (Trus & Steven, 1981), implemented on a VAX 3500 computer (Digital Equipment Corp., Maynard, MA) with a Gould Model 9000 image processor (Gould Imaging and Graphics, Fremont, CA). Photographs were recorded with an Imagecorder Plus camera station (Focus Graphics, Foster City, CA).

Image averaging was performed as previously described (Kocsis *et al.*, 1991). The data were sorted into two sets by a supervised classification (Fraser *et al.*, 1990) and the distinction between them confirmed by inspection of a principal components factorial map (Unser *et al.*, 1989). After proofing the data sets with the OMO algorithm (Unser *et al.*, 1986), the numbers of particles ultimately averaged were 178 and 177, respectively (CTEM, negative stain); 75 (STEM, negative stain); 143 (STEM, unstained). The resolution of the averaged images was 2·6 nm (CTEM, negative stain); 3·0 nm (STEM, negative stain); 4·0 nm (STEM, unstained), according to the spectral signal-to-noise-ratio (SSNR) criterion (Unser *et al.*, 1987). Dimensions given for various structural features were measured from the averaged images using the edge detection criterion described by Steven *et al.* (1988).

(j) *Computational analysis of amino acid sequence*

Secondary structure predictions were calculated according to Garnier *et al.* (1978) and Ptitsyn & Finkelstein (1983); hydropathy profiles according to Kyte & Doolittle (1982); and flexibility profiles according to Karplus & Schulz (1985). Matrix homology analysis and homology searches using the GENEBEE program (Brodsky *et al.*, 1991; Combee, Moscow, Russia) were calculated on either a VAX 3500 or an IBM PS2(386). The reference data base was Swiss-Prot (release 21). Fourier transforms were calculated using the PIC program (Trus & Steven, 1981). Use was also made of the AASAP program kindly made available by Dr D.A.D. Parry.

(k) *Circular dichroism*

Spectra were recorded on a Jasco J-600 spectropolarimeter as described by Wingfield *et al.* (1991). The protein concentration was estimated by measuring the absorbance at 280 nm. A correction for light scattering was made by extrapolation of a plot of λ^{-4} *versus* absorbance between 320 and 350 nm. A calculated molar absorption coefficient (ε) of 106·7 mM/cm was used, based on a mass of 220,251 which corresponds to residues 1 to 2165 (Delisse-Gathoye *et al.*, 1990). A mean residue molar mass of 101·7 Da was used. The fractional contents of the various kinds of secondary structure were calculated according to Perczel *et al.* (1991).

3. Results

(a) *Ultrastructural analysis of isolated FHA*

A field of negatively stained molecules is shown in Figure 1a. The great majority are narrow filaments that are rather uniform in length (45 to 50 nm) and

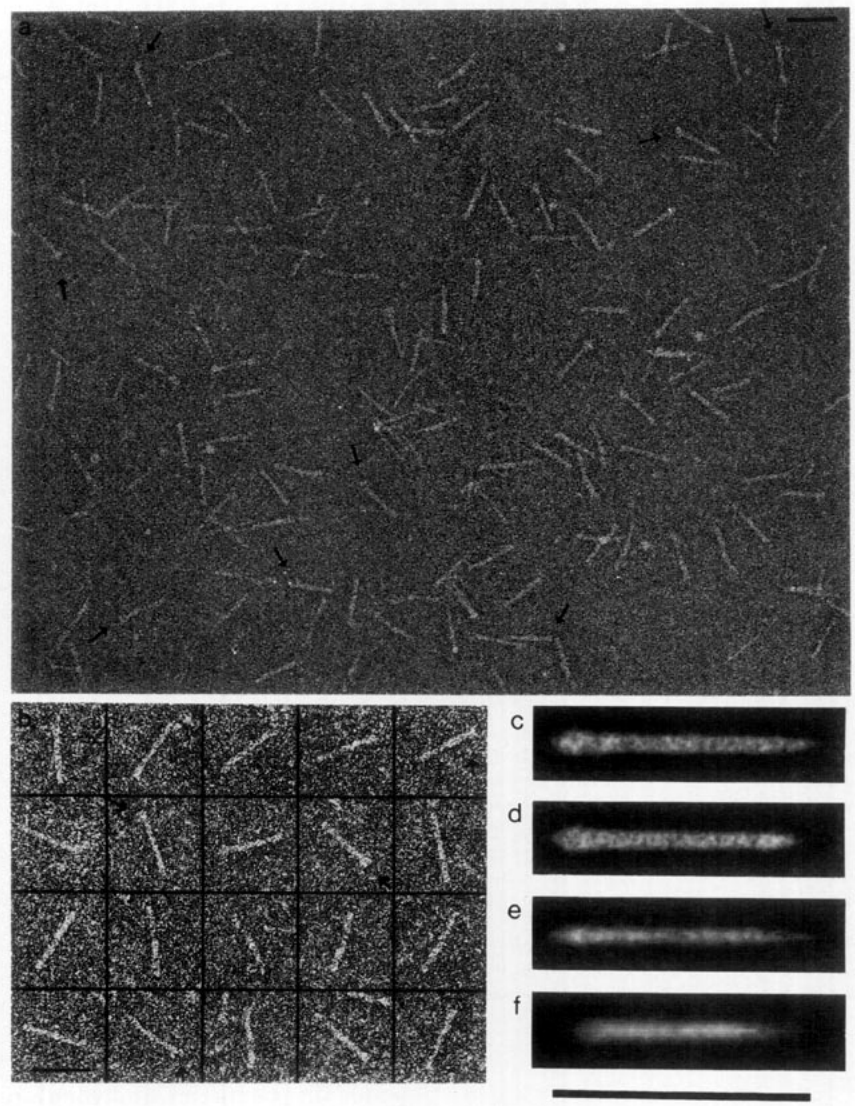

Figure 1. a, Electron micrograph of dispersed FHA molecules, negatively stained with uranyl acetate. b, Representative gallery. Most of the molecules show a globular head at one end which in some cases (arrows) appears heart-shaped. c to f, Averaged images of the FHA molecule: c and d are subsets of images from the same negative staining experiment by conventional transmission electron microscopy. They match perfectly over the entire length of the molecule (thus attesting the reproducibility of the analysis) except at the tail (right-hand end), indicating some conformational variability or sensitivity to specimen preparation in this part of the molecule. A virtually identical structure is depicted in a similar analysis of negatively stained specimens by STEM (e). Unstained, frozen-dried, molecules visualized by dark-field STEM (f) show a slightly shorter molecule with a less conspicuous head (see Results), but otherwise exhibit essentially the same features. Bars represent 50 nm.

show little curvature, suggesting a rigid conformation. A representative gallery is compiled in Figure 1b. Most of the molecules exhibit a knob or "head" at one end, which in some cases is heart-shaped, suggesting two separate domains. The filament shaft is about 4 nm wide, and tapers slightly from the head end to the tail end. Thus the FHA molecule is polar in structure and resembles a horseshoe nail.

To put these observations on a more quantitative basis, a large set of molecular images ($N = 387$) was analyzed by computational averaging. After straightening and alignment, the images were sorted into two length classes (longer and shorter filaments, respectively) which were averaged separately (Figure 1c, d). The resulting images reveal that FHA has three domains: the head, the shaft, and the tail. The two images match perfectly except at the tail end, where the shorter molecules terminate in a small knob instead of a thin tip. Thus they differ primarily in terms of the conformation(s) assumed by the tail domain. A set of STEM images of negatively stained FHA molecules was also averaged, yielding a representation (Figure 1e) that is slightly thinner, but otherwise consistent with the foregoing results. The dimensions of the three domains, as determined in both analyses, are given in Table 1.

Although FHA molecules are relatively rigid, some straightening was usually required in the

Table 1
Dimensions and masses of FHA and its domains

	CTEM neg. stain long	CTEM neg. stain short	STEM neg. stain	STEM unstained
Length of FHA (nm)	51·9	49·6	50·7	45·0
Length of head† (nm)	9·0	9·0	8·4	8·8
Length of shaft† (nm)	35·4	34·8	33·4	29·6
Length of tail† (nm)	7·5	5·9	8·9	8·8
Width of head (nm)	5·6	5·6	3·9	5·5
Average width of shaft‡ (nm)	4·6	4·4	3·4	4·3
Mass of FHA (kDA)	–	–	–	200
Mass of head§ (kDa)	42	39	41	42·5
Mass of shaft§ (kDa)	159	162	155	144
Mass of tail§ (kDa)	20	20	25	13·5

These measurements were made from averaged images (Figure 1c to f) which represent cylindrically averaged side-projections of the FHA molecule.

† The domainal divisions were made on the basis of visual evaluation of these images.

‡ The widths represent the average width along the shaft (despite slight tapering, see below), and the widths at the mid-points of the head and tail, respectively. Some lateral smearing is probably introduced in the straightening and averaging procedures, although this may be compensated in width measurements by the edge-detection criterion applied. Nevertheless, taking into consideration the limited resolution (see Materials and Methods) and variability in staining, there may well be a discrepancy of the order of 1 to 2 nm between the measured width and the actual width of the molecule, which represents a sizable fractional error. (In contrast, the fractional error for the lengths of the shaft and the molecule should be small because the uncertainty should be of the same magnitude as in the width measurement and these dimensions are much greater.) Nevertheless, we consider that the slight difference in width measured at opposite ends of the shaft represents a real effect, both because relative dimensions in a given averaged image may be determined much more precisely than the above margin of (absolute) error would suggest; also, because it is a reproducible feature of the different averaged images. Averaging the transverse density scan over 3 nm lengths of shaft at the head and tail ends, respectively, and measuring the ratio between resulting widths gave values of 0·82, 0·89, 0·97 and 0·87 for the images in Figure 1c, d, e, f. Some individual images (see Figure 1a, b) appear to exhibit more pronounced tapering and others less than the average, and may represent the molecule as viewed from differing orientations.

§ The domainal masses given for negatively stained data are calculated as the corresponding fractions of the total projected stain-excluding density, and calibrated to a sequence-derived mass of 220 kDa. The images from which these measurements were made are shown in Figure 1c to f.

analysis described above. From the straightening operations, an average curvature profile was calculated (data not shown). Its peaks mark the sites most susceptible to bending. The two major peaks are located in the head and close to the thin end. Flexibility at the latter site correlates with its observed structural variability (compare Figure 2a and b).

(b) *FHA is a monomer: mass determination by STEM*

The masses of individual molecules may be measured from dark-field STEM micrographs of unstained specimens by integrating the corresponding image intensity (Wall, 1979). The stoichiometry of an oligomeric protein may thus be determined if the molecular weight of its monomer is known. Dark-field STEM was performed on frozen-dried preparations of FHA (e.g. Figure 2a). The distribution of mass measurements (Figure 2b) is slightly skewed to the low mass side (see Figure legend), but most of the data lie in a peak that averages 200 kDa (S.D. = 33 kDa; S.E.M. = 3·5 kDa; $N = 173$). Taking into account this margin of error and the fact that no particles with masses of ~400 kDa (dimer), or ~600 kDa (trimer) etc, were observed, we conclude that FHA is a monomer under these experimental conditions.

(c) *Domainal mass analysis*

Correlation averaging was applied to the unstained STEM data (Figure 1f), yielding a representation of FHA that is of somewhat lower resolution, ~4·0 nm *versus* 2·6 nm, than the negatively stained images (Figure 1c to e). It shows a slightly shorter molecule, which probably represents shrinkage caused by freeze-drying (Furcinitti *et al.*, 1989; Makhov *et al.*, 1993). The structural polarity is also less pronounced. Nevertheless, there is a high degree of consistency between the stained and unstained representations. These images were used to estimate the masses and dimensions of the distinct domains (Table 1).

(d) *Proteolytic dissection of FHA*

In SDS-PAGE, FHA migrates with a mobility corresponding to a monomer mass of 220 kDa. However, breakdown products of ~150 kDa and ~98 kDa have been observed (Irons *et al.*, 1983; An der Lan *et al.*, 1986; Domenighini *et al.*, 1990). An arginine-rich site RRARR starting at residue 1097

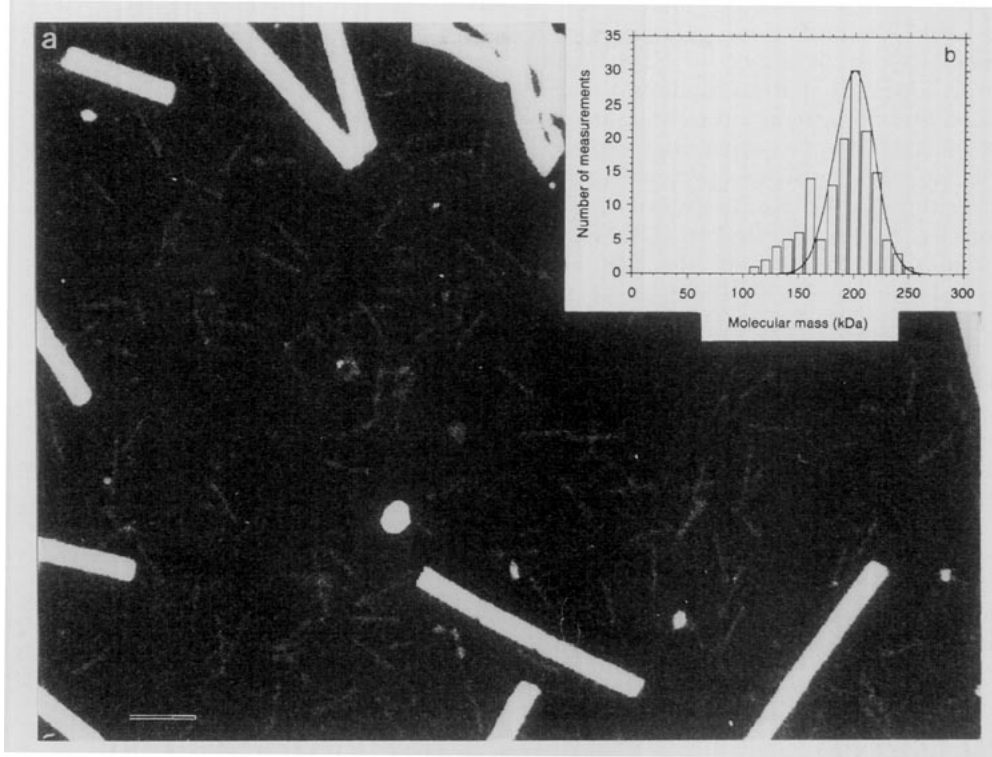

Figure 2. a, Dark-field STEM micrograph of a field of unstained FHA molecules, prepared by freeze-drying. The thick particles are TMV virions, used as internal mass standard. Bar represents 50 nm. b, Histogram of individual molecular mass measurements made from such micrographs. The distribution is slightly skewed to the low-mass side. There are 2 possible factors that may underlie the observed asymmetry. First, negative staining showed a few molecules (about 5%) that are significantly shorter than 50 nm. They may be contaminants, and may account for a few low-mass measurements. Second, some indistinct portion of the molecule may have been inadvertantly cropped out in a few cases when the mass integral was performed (the projected density of FHA is quite low). Thus the maximum at ~200 kDa which marks the mode of the distribution, gives our best estimate for the molecular mass of FHA.

is particularly vulnerable to proteases. Digestion with chymotrypsin yields fragments of 98 kDa (the amino-terminal portion) and a 140/150 kDa doublet which represents the carboxy-terminal portion (Domenighini *et al.*, 1990). With the aim of identifying these fragments as the head-containing or tail-containing domains of FHA, we attempted to visualize them separately by negative staining.

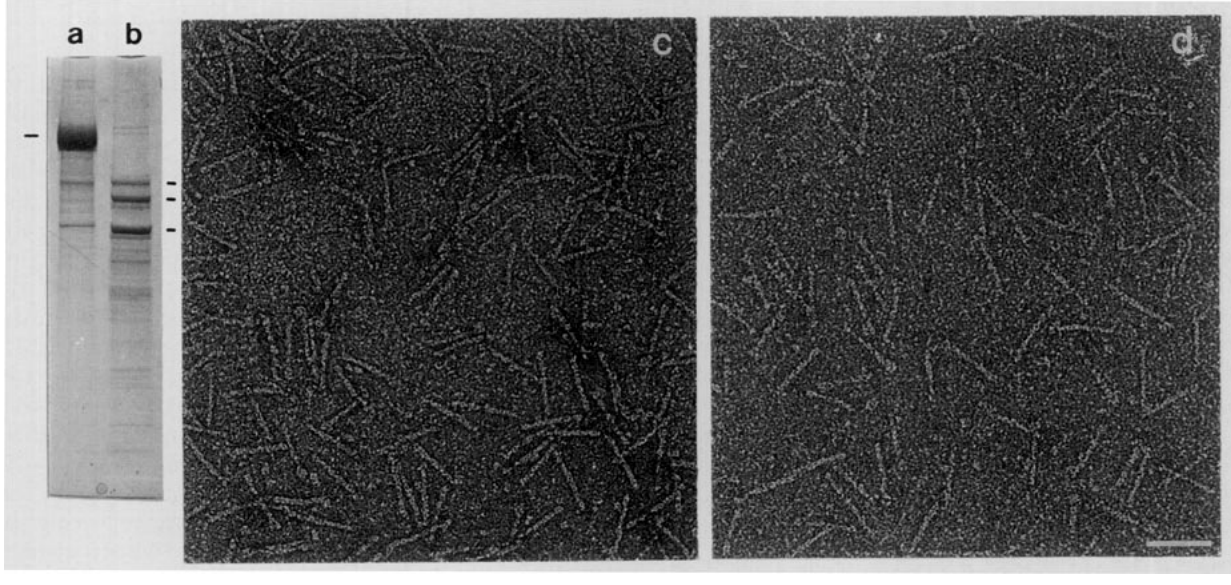

Figure 3. Effect of chymotrypsin digestion on FHA. As detected by SDS-PAGE, this treatment quantitatively digests intact 220 kDa FHA (a), yielding a doublet at 140 and 150 kDa, and a fragment of 98 kDa (b). Negatively stained electron micrographs of samples taken before (c) and after (d) this treatment show no significant morphological change. Bar represents 50 nm.

Digestion with chymotrypsin generated the expected fragments, as detected by SDS-PAGE (Figure 3a, b). Knowing that FHA binds to heparin (Menozzi *et al.*, 1991), we attempted to separate the fragments by heparin affinity chromatography but were unable to do so, even in the presence of 8 M urea or 1% Triton X-100. Nor were we able to separate them by carboxy-methyl or DEAE ion exchange chromatography, despite the fact that their expected pI values are very different, at 10·3 (98 kDa fragment) and 6·4 (140/150 kDa fragment), respectively (Relman *et al.*, 1989). Moreover, chymotryptic digestion did not affect the migration of FHA by gel filtration chromatography (data not shown). Upon examination by electron microscopy, the length and appearance of the molecules were found not to have been significantly altered by treatment with the protease (see Figure 3c and d). We conclude that the observed cleavage represents a local nicking, and the resulting fragments do not separate until denatured by boiling in SDS prior to gel electrophoresis.

(e) *FHA contains two extended tracts of tandem 19-residue repeats*

In an earlier analysis of the primary sequence of FHA, Delisse-Gathoye *et al.* (1990) detected six cycles of a quasi-repeat of about 40 residues in length between residues 382 and 664, followed by three more cycles from residue 701 to 992. Relman *et al.* (1989) also commented on the presence of two to three cycles of the same repeat with a subrepeat of about half its length. We have re-examined the sequence for the presence of repetitive elements, using several computational tools. Our findings confirm and extend the earlier observations. We conclude that the first repeat region (which we call R_1) is much longer than was originally thought, and consists of 38 copies of a 19-residue cycle that extend without interruption from residue 344 to residue 1065. It is followed by a second progression of 19-residue repeats (R_2) that have not been detected before.

The presence of both R_1 and R_2 is revealed by a self-homology matrix (Figure 4). Their extents were assessed by appraisal of the sequences flanking the ones detected by the homology matrix. The quasi-repeats and their consensus motifs are tabulated in Figure 5. In R_1, the super-repeat of 38 residues is evident between residues 383 and 571 both by inspection (Table 1) and from the homology matrix (Figure 4), but is not evident in the rest of R_1, where the basic 19-residue repeat is dominant. Extending from residue 1440 to 1688, R_2 contains 13 cycles of a somewhat different repeat. There is about the same level of homology among these repeats as is observed in R_1, but there is little homology between the two sets.

To determine the periodicities of R_1 and R_2, these sequences were coded for hydropathy (Kyte & Doolittle, 1982) and flexibility (Karplus & Schulz, 1985) and the corresponding Fourier transforms

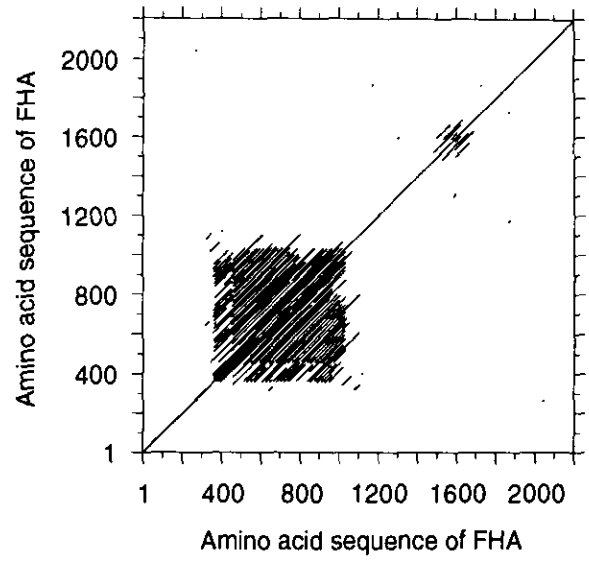

Figure 4. Self-homology matrix of the amino acid sequence of FHA. This plot was calculated using the DotHelix algorithm in the GeneBee package (Brodsky *et al.*, 1991), which is designed to overcome the problem of a fixed-width comparison window (Leontovich *et al.*, 1990). The "power of similarity" parameter is used as a measure of statistical significance for putative homologies. It is defined as the number of standard deviations (S.D.) over the expectation of the random comparison score, assuming that this score is normally distributed. This value is related to the probability of accumulating the given (or higher) score upon comparing segments of random sequences of the same lengths and with the same amino acid compositions as the sequences actually compared. Usually, values above 7 represent strong evidence for a genuine relationship between the compared sequences, whereas with similarity values of 3 to 7, additional data are generally required for this conclusion. This matrix plot was thresholded with a power of similarity value of 7. The 2 sets of parallel diagonal lines denote the presence of the 19-residue repeats in R_1 and R_2, respectively.

were calculated. The hydropathy spectrum of R_1 (Figure 6(a)) contains a series of sharp, equally spaced, peaks that index as harmonics of a 19-residue periodicity. The peak corresponding to the basic repeat does not rise above background, but the second, third, seventh and ninth orders are strongly expressed. The flexibility spectrum (Figure 6(b)) shows strong second and third orders of the same repeat. The absence of a visible first order reflection, followed by strong second and third orders suggest that the repeat consists of two similar components. The hydropathy and flexibility spectra of R_2 reveal several orders of a slightly longer repeat of 19·5 residues (Figure 6(c), (d)). These peaks are broader and proportionately less high than the peaks in the R_1 spectra, reflecting the fact that R_2 is shorter, and its repetitive structure is less stringently conserved.

Such peaks were not observed in hydropathy or flexibility spectra calculated for other segments of the FHA sequence, nor with other proteins (e.g.

a

```
344  L S L K G A G V V S A G K L A S G G G A  363
364  V N V A G G G A V K I A S A S S V G N    382
383  L A V Q G G G K V Q A T L L N A G G T    401
402  L L V S G R Q A V Q L G A A S S R Q A    420
412  L S V N A G G A L K A D K L S A T R R    439
440  V D V D G K Q A V A L G S A S S N A      457
458  L S V R A G G A L K A G K L S A T G R    476
477  L D V D G K Q A V T L G S V A S D G A    495
496  L S V S A G G N L R A N E L V S S A Q    514
515  L E V R G Q R E V A L D D A S S A R G    533
534  M T V V A A G A L A A R N L Q S K G A    552
553  I G V Q G G E A V S V A N A N S D A E    571
572  L R V R G R G Q V D L H D L S A A R G    590
591  A D I S G E G R V N I G R A R S D S D    609
610  V K V S A H G A L S I D S M T A L G A    628
629  I G V Q A G G S V S A K D M R S R G A    647
648  V T V S G G G A V N L G D V Q S D G Q    666
667  V R A T S A G A M T V R D V A A A A D    685
686  L A L Q A G D A L Q A G F L K S A G A    704
705  M T V N G R D A V R L D G A H A G G Q    723
724  L R V S S D G Q A A L G S L A A K G E    742
743  L T V S A A R A A T V A E L K S L D N    761
762  I S V T G G E R V S V Q S V N S A S R    780
781  V A I S A H G A L D V G K V S A K S G    799
800  I G L E G W G A V G A D S L G S D G A    818
819  I S V S G R D A V R V D Q A R S L A D    837
838  I S L G A E G G A T L G A V E A A G S    856
857  I D V R G G S T V A A N S L H A N R D    875
876  V R V S G K D A V R V T A A T S G G G    894
895  L H V S S G R Q L D L G A V Q A R G A    913
914  L A L D G G A G V A L Q S A K A S G T    932
933  L H V Q G G E H L D L G T L A A V G A    951
952  V D V N G T G D V R V A K L V S D A G    970
971  A D L Q A G R S M T L G I V D T T G D    989
990  L Q A R A Q Q K L E L G S V K S D G G   1008
1009 L Q A A A G G A L S L A A A E V A G A   1027
1028 L E L S G Q G V T V D R A S A S R A R   1046
1047 I D S T G S V G I G A L K A G A V E A   1065
```

consensus

```
l
i - v s g - g a l - v - - l - a - g -
v   l   a       v   a     a   a
```

b

```
1440 S N K I R L M G P L Q V N A G G P V      1455
1456 S N T G N L K V R E G V T V T A A S F    1474
1475 D N E T G A E V M A K S A T L T T S G A A 1495
1496 R N A G K M Q V K E A A T I V A A S V    1514
1515 S N P G T F T A G K D I T V T S R G G F  1534
1535 D N E G K M E S N K D I V I K T E Q F    1553
1554 S N G R V L D A K H D L T V T A S G Q A  1573
1574 D N R G S L K A G H D F T V Q A Q R I    1592
1593 D N S G T M A A G H D A T L K A P H L    1611
1612 R N T G Q V V A G H D I H I I N S A K L  1631
1632 E N T G R V D A R N D I A L D V A D F    1650
1651 T N T G S L Y A E H D A T L T L A Q G    1669
1670 T Q R D L V V D Q D H I L P V A E G T    1688
```

consensus

```
t
s N t G - l - a - h d i t i - a - - -
d       v               v
```

Figure 5. Tabulation of the 19-residue repeats from regions R_1 (a) and R_2 (b) of the amino acid sequence of FHA, and their consensus motifs.

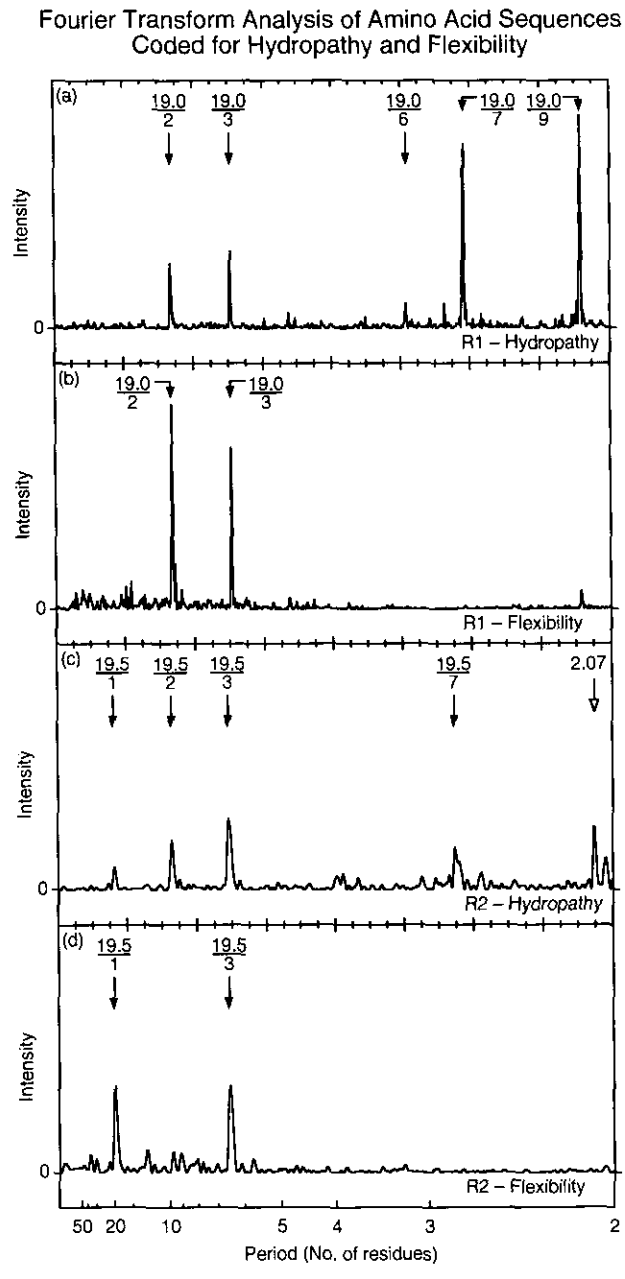

Figure 6. Fourier transform analysis of the R_1 (a,b) and R_2 (c,d) regions of the FHA amino acid sequence after coding the sequences according to hydrophathy (Kyte & Doolittle, 1982), (a) and (c), and flexibility (Karplus & Schulz, 1985), (b) and (d). No smoothing (local averaging) was performed on the sequences prior to calculating the transforms. The spectra show pronounced peaks corresponding to harmonics of the 19-residue repeats present in both sequences. Also evident in (a) and (c) are peaks corresponding to a 2-residue hydropathy repeat, reflecting the presence of peptides with alternating polar and apolar residues.

Makhov *et al.*, 1993), indicating that they are specific to R_1 and R_2. Because there are no significant peaks at harmonics of a 38-residue period that are not also harmonics of a 19-residue repeat, we conclude that the latter period is the dominant repetitive element throughout both R_1 and R_2, even though, in the first section of R_1, this is a sub-repeat of a 38-residue cycle (Delisse-Gathoye *et al.*, 1990). Moreover, the 19-residue repeats in R_1 subdivide into two components of 9 or 10 residues.

(f) *Predicted secondary structure of repeats*

According to the algorithms of Garnier *et al.* (1978) and Ptitsyn & Finkelstein (1983), the repeat regions are expected to be rich in β-structure, with alternating strands and turns. The consensus

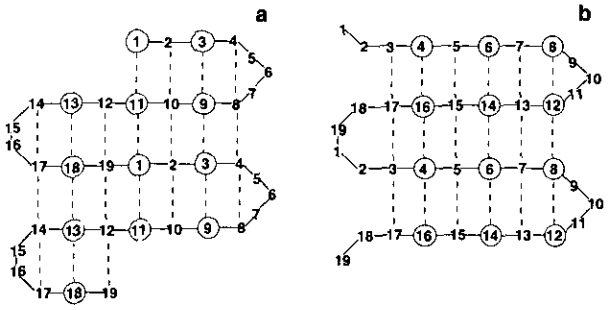

Figure 7. Possible secondary structures for the 19-residue repeats of FHA. a, R_1 repeat motif; b, R_2 repeat motif. Both consist of short β-strands connected by 2 β-turns. The numbering of the amino acid residues follows the ordering given in Figure 5. Circled positions are mostly occupied by hydrophobic residues.

prediction is for two β-turns per 19-residue motif: possible foldings for the repeats of R_1 and R_2 are shown in Figure 7a and b, respectively. Interestingly, the predicted β-strands are markedly amphipathic, with alternating polar and non-polar residues, a pattern that makes one side of the β-sheet hydrophilic and the other side hydrophobic. The non-random alternation of polar and non-polar residues is also reflected in the hydropathy spectra of both R_1 and R_2 (Figure 6(a), (c)), in terms of strong peaks at a frequency of $\sim(2\,\text{res.})^{-1}$. This peak, which is particularly strong in the case of R_1, is sharpened by being sampled as the ninth order of the 19-residue repeat. That this peak relates to the hydrophobicity/hydrophilicity characteristics of the sequences is confirmed by its virtual absence from the flexibility spectra (Figure 6(b), (d)).

(g) *Secondary structure determination by circular dichroism*

We have recorded the CD spectrum of purified FHA (Figure 8). Its shape implies a high content of β-structure. Estimation of secondary structure by deconvolution of this spectrum according to Perczel *et al.* (1991) yielded 56% β-strand and 43% β-turn. These values are in good agreement with the predicted conformations of the R_1 and R_2 motifs (above), even though the repeat regions account for only about half of the molecule.

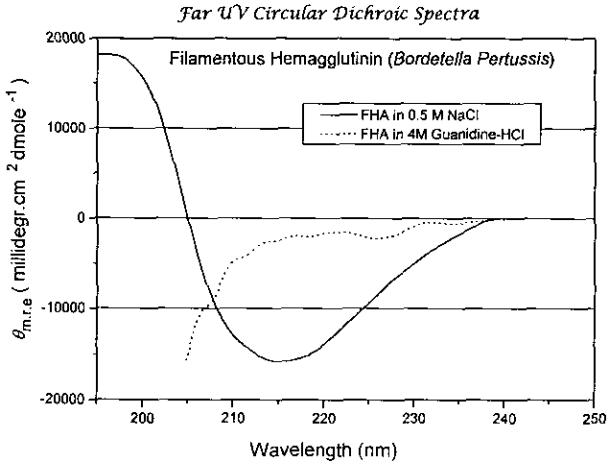

Figure 8. Circular dichroism spectra of purified FHA, both native and after denaturation in guanidine hydrochloride.

(h) *A hairpin molecular model of FHA*

Based on the morphological, conformational, and sequence-related observations presented above, we derived a molecular model for FHA. Its amino acid sequence falls into five distinct regions: the two repeat regions (R_1 and R_2), the intervening sequence (S_i), and the end-sequences (S_N and S_C), as demarcated in Figure 9(b). First, we noted that the length of R_1, folded into a ladder of 38 modules of cross-β structure (Figure 7) with an inter-strand spacing of 0·48 nm, should be $\sim$35 nm, which closely matches the observed length of the shaft. Next, since the hydrophobic face of this hyper-extended β-sheet should be sequestered from the solvent, we postulated that R_2 forms a similar β-ladder folded back to pair with part of R_1; and the remainder of R_1 is interfaced to parts of S_i and S_C, which also assume some other extended, amphipathic conformations. From this point, only two formulations are possible: (1) the head contains both termini, and the tail contains all or part of S_i; (2) the head represents S_i, with the terminal domains located in the tail. However, it is not possible for any class 2 model to assign enough mass to the head domain and have the thicker end of the shaft at the same end, as required by the electron microscopy evidence (Figure 1). The best agreement with our estimates of domainal masses is obtained with a class 1 model, with the start of R_2 offset by about 10 nm relative to the end of R_1 (Figure 9(a)).

(i) *Test of model by immuno-electron microscopy*

To test this model, we used electron microscopy to localize the binding site on the molecule of an antibody of defined specificity. These experiments were conducted with the monoclonal antibody 2/1 : 5E, which was raised against purified FHA and whose epitope has been mapped between residues 1655 and 2012 (see Materials and Methods). The number of (FHA+IgG) complexes observed increased with time of co-incubation up to approximately 6 hours, and did not change significantly thereafter. Under these conditions, 10 to 15% of the FHA molecules were distinctly labeled with an antibody (e.g. Figure 10a); in contrast, essentially no labeling (<1%) was detected with a control murine IgG. That the binding was specific was further attested by the clustering of the attachment points of the large majority of bound antibodies around a single site (Figure 10b). This site is either in the head or in the shaft portion immediately adjacent to the head, which is where the model predicts residues 1655 to 2012 to be (Figure 9(a)). In fact, these data

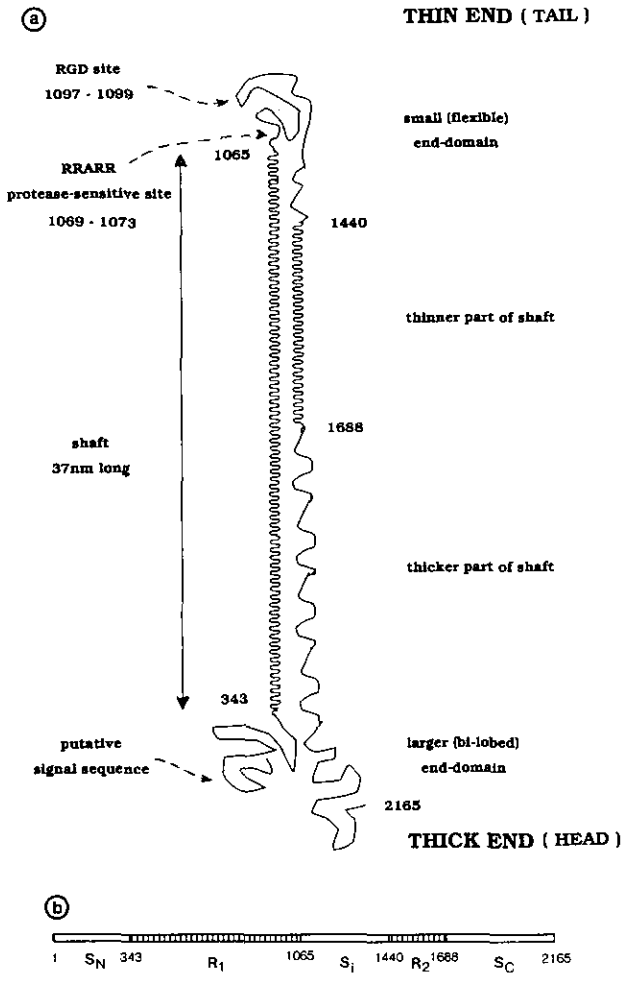

Figure 9. (a) Model for FHA, showing proposed locations of adhesion sites and immunodominant epitopes. (b) Diagram depicting the 5 distinct segments of the amino acid sequence of FHA. R_1 and R_2 are the repeat regions. S_N, S_i and S_C are the amino-terminal, intervening, and carboxy-terminal sequences, respectively. In common with other investigators (Domenighini *et al.*, 1990; Delisse-Gathoye *et al.*, 1990), we have found that the amino terminus of FHA is blocked. Thus, it is not clear where the polypeptide chain actually starts. There are some resemblances between the predicted amino-terminal sequence, and prokaryotic signal sequences (Delisse-Gathoye *et al.*, 1990), suggesting the possibility of a truncation from the amino terminus. However, the entire sequence of the polypeptide chain contains only 3 cysteine residues, all in the extreme amino terminus, at positions 24, 31 and 62. Chemical analysis indicates that 220 kDa FHA contains at least 1 cysteine (R. Boykins, personal communication), implying that any proteolytic modification of its amino terminus does not remove more than 61 residues. The molecular model described below is based on the assumption of a 2165-residue polypeptide chain with a predicted molecular mass of 220 kDa, which is quite consistent with the independent estimates obtained from SDS-PAGE and STEM mass analysis. As drawn, the diagram depicts planar β-sheets, although they would be expected to rotate around the molecular axis, although sufficiently slowly that the observed correlation between the length of R_1 and the shaft would not be affected.

suggest a more precise mapping of the epitope than was given by the fusion protein analysis. The model assigns it to the carboxy-terminal part of the mapped region. Noting that there is some uncertainty as to where R_2 should be placed along R_1, our best estimate is that the epitope lies between residues 1900 and 2012.

(j) *Resemblance to leucine-rich repeat (LRR) proteins*

We conducted homology searches for other proteins that may be similar to the R_1 or R_2 sequences, but found no significant matches. However, their overall characteristics are reminiscent of the LRRs that have been detected in over 30 different proteins (e.g. Takahashi *et al.*, 1985; Rothberg *et al.*, 1990; Krantz *et al.*, 1991), almost all from eukaryotic cells. The FHA repeats are of comparable length to LRRs, albeit slightly shorter: most LRRs are 20 to 30, and usually 24 to 28 residues long. FHA repeats also have a higher than normal content of leucine residues, usually in conserved positions (Figure 5) although they are less Leu-rich than most LRRs. Malvar *et al.* (1992) distinguished a subset of LRR proteins whose motifs contain a conserved Asn residue, a feature that is shared by the R_2 repeats, but not the R_1 repeats, of FHA (Figure 5). The FHA repeats have no overall homology with LRRs reported to date (which themselves exhibit quite wide sequence diversity; see, e.g. Rothberg *et al.*, 1990), although the R_1 consensus motif does contain a peptide, LxLxxL (see Figure 5a), that occurs in many LRRs. The proposition of structural similarity between FHA repeats and at least some LRRs is further supported by the virtual superimposability of the CD spectrum of FHA (Figure 8) with those of two synthetic peptides corresponding to single-copy LRRs. These peptides, both from *Drosophila* proteins, are a 24-residue LRR from chaoptin when associated with lipid vesicles (Figure 5 of Krantz *et al.*, 1991); and a 23-residue LRR from the Toll gene product, polymerized *in vitro* into extended filaments (Figure 2E of Gay *et al.*, 1991).

4. Discussion

(a) *Comparison between FHA and other filamentous proteins rich in β-structure*

Most filamentous proteins characterized to date form coiled-coil α-helices or collagen-like (3_{10}) helices; conversely, there have been relatively few cases of filamentous proteins whose conformations are based on β-sheets and β-turns. Recently, however, several such molecules have come to light and they are discussed briefly here in the context of structural comparison with FHA and also because many of them, like FHA, have been implicated in adhesion reactions.

(i) *Virus fibers, viral adhesins*

Many large, non-enveloped, viruses display fibrous proteins that extend from the virion and

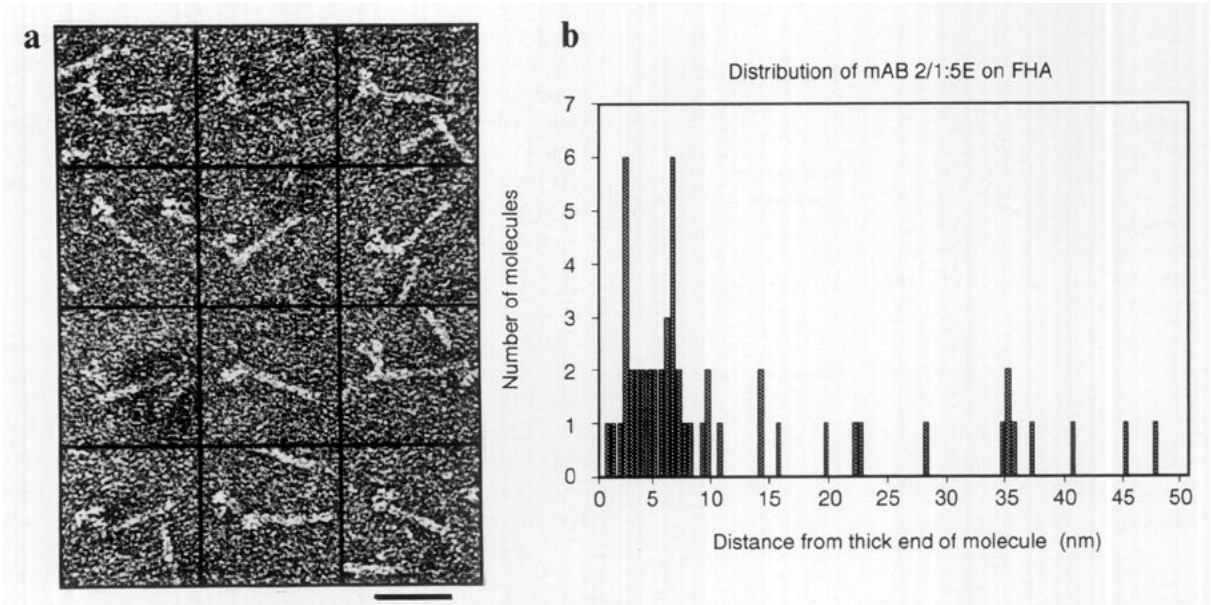

Figure 10. Immuno-labeling electron microscopy of FHA molecules labeled with the monoclonal antibody 2/1:5E whose epitope lies between residues 1655 and 2012 in the carboxy-terminal part of the molecule. a, Gallery of (FHA + IgG) complexes visualized by negative staining. Bar represents 50 nm. b, Distribution of measurements of point of attachment of antibody relative to the head end of molecule.

bind to receptors on host cells. One of the most studied is the adenovirus fiber, whose sequence contains a run of 15-residue repeats, and whose structure contains β-sheets whose strands run perpendicular to the fiber axis (Green *et al.*, 1983). These authors proposed a model for a dimeric fiber based on two long amphipathic antiparallel β-sheets paired back to back. Subsequent research concluded that this fiber is a trimer (van Oostrum & Burnett, 1985), and Stouten *et al.* (1992) have proposed a three-stranded β-helical model.

Bacteriophage fibers have similar structural properties. β-Sheets were detected by electron diffraction on the distal part of the long tail-fiber of phage T4, leading to a model based on cross-β structure (Earnshaw *et al.*, 1979) which the Green model (above) resembles. The long tail-fiber has long been thought to be dimeric (King & Laemmli, 1971; Dickson, 1973), but recent evidence suggests that it is trimeric (Cerritelli *et al.*, 1994). The short tail-fiber of T4 is a trimer (Kells *et al.*, 1975), and has a repeat motif and a rod domain whose properties are similar to those of the adenovirus fiber (Makhov *et al.*, 1993). The tail-fibers of phages P22 (Sargent *et al.*, 1988) and T7 (distal half-fiber, Steven *et al.*, 1988) are also trimers with high contents of β-structure. Thus, there is an emerging trend: these viral adhesins are elongated molecules, tend to be trimers and to be rich in β-strands and β-turns. In the two cases for which evidence is available, these strands have been found to run transverse to the fiber axis.

Some of these properties are reminiscent of FHA: in particular, our model resembles the Green proposal for the adenovirus fiber with the important distinction that FHA is a monomer and its two amphipathic β-sheets run in opposite directions to form an intramolecular, not an intermolecular, pairing.

(ii) *Pectate lyase and alkaline protease*

Recently, high resolution structures have been determined by X-ray crystallography for two proteins that adopt extended conformations rich in β-structure. Pectate lyase, a plant pathogen secreted by *E. chrysanthemi*, was found to form a right-handed β-helix consisting of short β-strands coiled around the molecular axis (Yoder *et al.*, 1993*a*,*b*) to form three long parallel β-sheets. There are, in minimum, 22 residues per turn, corresponding to an axial rise of 0·022 nm per residue. The alkaline protease of *P. aeruginosa* has a 219-residue C-terminal domain of which a section is coiled into a similar conformation, (Baumann *et al.*, 1993).

An alternative model for FHA would envisage the shaft as a right-handed β-helix like that of pectate lyase, thus requiring about 1600 residues to give the observed length of 35 nm. To obtain a mass ratio of head : tail of 2 : 1 (Table 1) and have the carboxy terminus in the head, the β-helix should start at about residue 200. The aspects of this model that we find unappealing are that its structural domains do not correlate with sequence-based division of the molecule into R_1, R_2, etc; also, this model places the protease-sensitive site in the middle of the shaft, whereby the failure of the two fragments to separate after chymotrypsin digestion is less easily explained. However, one positive aspect of this model is that it would allow the conserved Asn residues in R_2 to stack into an "asparagine ladder", as in pectate lyase (Yoder *et al.*, 1993*a*). An additional point of correspondence is that, like both β-helix-containing

proteins, FHA is a bacterial secretory protein (Willems *et al.*, 1994).

(iii) *Leucine-rich repeat (LRR) proteins*

Recently, the first high resolution structure for an LRR protein, a porcine RNase inhibitor, has been determined by Kobe & Deisenhofer (1993). The fold of this 28/29-residue LRR motif contains one β-strand and one α-helix, with the α-helix folded back antiparallel to the β-strand. Successive copies of the motif are stacked together so that the β-strands form a long parallel β-sheet, backed by a layer of parallel α-helices. Because the width of an α-helix is greater than the spacing between β-strands, the molecule has a horseshoe shape, with the β-sheet wrapped around the inside and the layer of α-helices on the outside (Kobe & Deisenhofer, 1993). There is an evident topological resemblance between this fold and the β-helix (Yoder *et al.*, 1993*a*; see above): both are one-start right-handed solenoids distinguished by the substitution of an α-helix in the former for a strand-turn-strand motif in the latter.

Our model for FHA has certain features in common with the RNase inhibitor structure. Both proteins are monomers organized as linear stacks of their respective repeat motifs, linked in part by hydrogen bonding between β-strands from adjacent motifs. However, in the RNase inhibitor, adjacent β-strands are parallel, not antiparallel. Whereas the RNase inhibitor is curved, FHA is quite straight, indicating that its repeat motif should not contain an α-helix, as already implied by the sequence analysis and the spectroscopic evidence. We anticipate the possibility of further variations on the same general theme, allowing for different numbers (B) of β-strand steps per repeat, parallel or antiparallel β-sheets, and multimeric as well as monomeric molecules. In general, we can predict the lengths of the corresponding domains to be $0{\cdot}48 \times B \times N$ nm, where N is the number of consecutive repeats (in addition to any contributions from flanking sequences). Thus the RNase inhibitor structure corresponds to ($B=1$; $N=15$), and our model for the FHA rod to ($B=2$; $N=38$).

Functionally, we note that roles in forming contacts between eukaryotic cells have been attributed to some LRR-containing proteins (e.g. Hashimoto *et al.*, 1988; Reinke *et al.*, 1988; Lopez *et al.*, 1988). It remains to be seen whether their LRR domains serve as spacers that link receptor-binding domains or whether they play more direct roles in adhesion.

(b) *Monomeric hairpin model of FHA*

FHA has a marked tendency for self-aggregation, and high salt concentrations (>0·5 M) are required to maintain its solubility. Our STEM data show that FHA is a monomer under these conditions. In this respect, it is somewhat unusual since most filamentous proteins characterized to date consist of side-by-side associations of two or three elongated subunits. In the following, we discuss our molecular model further, bearing in mind that the capacity of FHA to self-associate and/or to associate with other components of the *B. pertussis* cell envelope may be relevant to how it functions as an adhesin *in vivo*.

(i) *Differing lengths of R_1 and R_2*

One unsettled aspect of our model is that we have not given a specific prescription for the conformation of the sequences that are hypothesized to cover the part of the hydrophobic face of R_1 that is not mated with R_2. Here, pectate lyase may offer a clue: its structure is repetitive, of short strand/turn motifs, although no hint of the presence of tandem repeats in its amino acid sequence was detected by Yoder *et al.* (1993*a*), which has also been our experience using Fourier analysis (data not shown). It may be, therefore, that the 19-residue structural motif of FHA is sustained on either side of R_2 even though no repeats are discernible from sequence homologies. This conjecture is consistent with the spectroscopic estimate of secondary structure which describes a molecule consisting almost entirely of β-strands and β-turns, even though R_1 and R_2 together account for less than 50%. To explain the observed tapering of the shaft from head to tail, we suppose that the postulated β-strands may be somewhat longer towards the head end and shorter towards the tail end.

(ii) *Distribution of functionally significant sites in FHA*

Our model readily explains why the fragments produced by proteolysis at the RRARR site do not separate: the cleavage site is located in S_i at the thin end of the molecule (Figure 10a), and the interaction between R_1 and R_2 which is hypothesized to be primarily responsible for stabilizing the fragments' association is not affected by this cleavage event. The RGD integrin-binding site is also assigned to the thin, flexible, end of the molecule (Figure 9(a)).

Recently, monoclonal antibodies and patient-derived antisera have been used to map the immunodominant epitopes of FHA with reference to fusion proteins that contain various parts of the FHA sequence (E. Leininger & M. J. Brennan, unpublished results). These analyses identify two classes of such epitopes. One class (type 1), which includes the monoclonal 2/1 : 5E used in this study, is distal to residue 1655, and we expect them to lie in the head of the molecule. This region may be important for eliciting an immune response that confers protection against pertussis infection. Similarly, our model assigns the other class (type 2 epitopes) to the tail half of FHA.

The segment of FHA between residues 91 and 205 is almost 50% homologous with amino-terminal sequences of the hemolysins of *Serratia marcescens* and of *Proteus mirabilis* (Delisse-Gathoye *et al.*, 1990). The same segment also bears a significant homology to amino-terminal sequences of the HMW-1 and HMW-2 proteins of *Haemophilus influenzae* (Barenkamp & Leininger, 1992). Bearing

in mind that the hemolysins and FHA (and possibly, also HMW1 and HMW2) have in common an affinity for the surface of erythrocytes, it seems plausible that this portion of FHA is implicated in its hemagglutination activity. Our model assigns it to the head domain (Figure 9(a)).

In summary, our model predicts that at least two sites involved in separate adhesion reactions, the RGD motif and the hemagglutination site, are located at opposite ends of the FHA molecule. It remains to be seen how a secreted, monomeric, protein contrives to fulfil the adhesion functions that have been attributed to FHA. Present data are compatible with at least two possibilities: (1) the 35 nm-long, rigid shaft may serve as a linker or spacer that attaches the bacterium *via* its tail domain to CR3 integrins presented on the surface of alveolar macrophages (Relman *et al.*, 1990). (2) Alternatively, the amphipathic nature of the repeats in the FHA shaft suggest that they may interact directly with membranes. This alternative is supported, albeit indirectly, by the observed membrane insertion of the chaoptin LRR (Krantz *et al.*, 1991). The mechanism of action of FHA as a multivalent adhesin may also involve other, yet-to-be-identified, sites that cause self-association or binding of FHA to other components of the *B. pertussis* cell envelope.

We thank Drs J. Wall and J. Hainfeld for support at the Brookhaven STEM facility, and Drs M. Kessel, J. King and B. Kobe for helpful discussions.

References

Aebi, U., Fowler, W. E., Buhle, L. E. & Smith, P. R. (1984). Electron microscopy and image processing applied to the study of protein structure and protein-protein interactions. *J. Ultrastruct. Res.* **88**, 143–176.

An der Lan, B., Cowell, J. L., Burstyn, D. G. & Manclark, C. R. (1986). Characterization of the filamentous hemagglutinin from *Bordetella pertussis* by gel electrophoresis. *Mol. Cell. Biochem.* **70**, 31–55.

Arai, H. & Sato, Y. (1976). Separation and characterization of two distinct hemagglutinins contained in purified leukocytosis-promoting factor from *Bordetella pertussis*. *Biochim. Biophys. Acta*, **444**, 765–782.

Ashworth, L. A. E., Dowsett, A., Irons, L. & Robinson, A. (1985). The location of surface antigens of *Bordetella pertussis* by immunoelectron microscopy. *Develop. Biol. Stand.* **61**, 143–152.

Askelof, P., Granstrom, M., Gillenius, P. & Lindberg, A. A. (1982). Purification and characterization of a fimbrial haemagglutinin from *Bordetella pertussis* for use in an enzyme-linked immunosorbent assay. *J. Med. Microbiol.* **15**, 73–83.

Barenkamp, S. J. & Leininger, E. (1992). Cloning, expression, and DNA sequence analysis of genes encoding nontypeable *Haemophilus influenzae* high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of *Bordetella pertussis*. *Infect. Immunol.* **60**, 1302–1313.

Baumann, U., Wu, S., Flaherty, K. M. & McKay, D. B. (1993). Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J.* **12**, 3357–3364.

Brennan, M. J., Burns, D. L., Meade, B. D., Shahin, R. D. & Manclark, C. R. (1991*a*). Recent advances in the development of pertussis vaccines. In *Vaccines: New Approaches to Immunological Problems* (Ellis, R., ed.), pp. 23–52, Butterworth, Shoreham, MA.

Brennan, M. J., Hannah, J. H. & Leininger, E. (1991*b*). Adhesion of *Bordetella pertussis* to sulfatides and to the GalNAcbeta4Gal sequence found in glycosphingolipids. *J. Biol. Chem.* **266**, 18827–18831.

Brodsky, L. I., Drachev, A. L., Tatuzov, P. L. & Chumakov, K. M. (1991). Package of programs for analysis of the sequences of biopolymers-GENEBEE. *Biopolymers Cells*, **7**, 10–14.

Brown, D. R. & Parker, C. D. (1987). Cloning of the filamentous hemagglutinin of *Bordetella pertussis* and its expression in *Escherichia coli*. *Infect. Immunol.* **55**, 154–161.

Cerritelli, M. E., Simon, M. N., Vaca, M., Conway, J. F. & Steven, A. C. (1994). The long tail-fibers of bacteriophage T4: determination of stoichiometry and domainal structure by dark-field STEM. In Proc. XIIIth Int. Cong. Elec. Micr. (Paris), in the press.

Cowell, J. L., Urisu, A., Zhang, J. M. Steven, A. C. & Manclark, C. R. (1986). Filamentous hemagglutinin and fimbriae of *Bordetella pertussis*: properties and roles in attachment. *Microbiol. 1986*, 55–58.

Delisse-Gathoye, A.-M., Locht, C., Jacob, F., Raaschou-Nielsen, M., Heron, I., Ruelle, J.-L., DeWilde, M. & Cabezon, T. (1990). Cloning, partial sequence, expression, and antigenic analysis of the filamentous hemagglutinin gene of *Bordetella pertussis*. *Infect. Immunol.* **58**, 2895–2905.

Dickson, R. C. (1973). Assembly of the bacteriophage T4 tail-fibers. IV. Subunit composition of tail fibers and fiber precursors. *J. Mol. Biol.* **79**, 633–647.

Di Tommaso, A., Domenighini, M., Bugnoli, M., Taglianue, A., Rappuoli, R. & De Magistris, M. T. (1991). Identification of subregions of *Bordetella pertussis* filamentous hemagglutinin that stimulate human T-cell responses. *Infect. Immunol.* **59**, 3313–3315.

Domenighini, M., Relman, D., Capiau, C., Falkow, S., Prugnola, A., Scarlato, V. & Rappuoli, R. (1990). Genetic characterization of *Bordetella pertussis* filamentous haemagglutinin: a protein processed from an unusually large precursor. *Mol. Microbiol.* **4**, 787–800.

Earnshaw, W. C., Goldberg, E. B. & Crowther, R. A. (1979). The distal half of the tail fiber of bacteriophage T4. *J. Mol. Biol.* **132**, 101–131.

Frank, J. (1989). Image analysis of single molecules. *Electron Microscopy Rev.* **2**, 53–74.

Fraser, R. D. B., Furlong, D. B., Trus, B. L., Nibert, M. L., Fields, B. N. & Steven, A. C. (1990). Molecular structure of the cell-attachment protein of reovirus: correlation of computer-processed electron micrographs with sequence-based predictions. *J. Virol.* **64**, 2990–3000.

Furcinitti, P. S., van Oostrum, J. & Burnett, R. M. (1989). Adenovirus polypeptide IX revealed as capsid cement by difference images from electron microscopy and crystallography. *EMBO J.* **8**, 3563–3570.

Garnier, J., Osguthorp, J. D. & Robson, B. (1978). Analysis of the accuracy and implications of simple methods for providing the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97–120.

Gay, N. J., Packman, L. C., Weldon, M. A. & Barna, J. C. J. (1991). A leucine-rich repeat peptide derived from the *Drosophila* Toll receptor forms extended filaments with a beta-sheet structure. *FEBS Letters*. **291**, 87–91.

Green, N. M., Wrigley, N. G. Russel, W. C., Martin, S. R. & McLachlan, A. D. (1983). Evidence for a repeating cross-beta sheet structure in the adenovirus fibre. *EMBO J.* **2**, 1357–1365.

Hashimoto, C., Hudson, K. L. & Anderson, K. V. (1988). The Toll gene of *Drosophila*, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. *Cell*, **52**, 269–279.

Irons, L. I., Ashworth, L. A. E. & Wilton-Smith, P. (1983). Heterogeneity of the filamentous hemagglutinin of *Bordetella pertussis*. *J. Gen. Microbiol.* **129**, 2769–2778.

Karplus, P. A. & Schulz, G. E. (1985). Prediction of chain flexibility in proteins. *Naturwissen*, **72**, 212–213.

Kells, S. S., Ohtsuki, M. & Haselkorn, R. (1975). The structure of bacteriophage T4 gene 12 protein. *J. Mol. Biol.* **99**, 349–351.

Kimura, A., Mountzouros, K. T., Relman, D. A., Falkow, S. & Cowell, J. L. (1990). *Bordetella pertussis* filamentous hemagglutinin: evaluation as a protective antigen and colonization factor in a mouse respiratory infection model. *Infect. Immunol.* **58**, 7–16.

King, J. & Laemmli, U. K. (1971). Polypeptides of the tail fibers of bacteriophage T4. *J. Mol. Biol.* **62**, 465–477.

Kobe, B. & Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature (London)*, **366**, 751–756.

Kocsis, E., Trus, B. L., Steer, C. J., Bisher, M. E. & Steven, A. C. (1991). Image averaging of flexible fibrous macromolecules: the clathrin triskelion has an elastic proximal segment. *J. Struct. Biol.* **107**, 6–14.

Krantz, D. D., Zidovetzki, R., Kagan, B. L. & Zipursky, S. L. (1991). Amphipathic beta structure of leucine-rich repeat peptide. *J. Biol. Chem.* **266**, 16801–16807.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**, 105–132.

Leininger, E., Roberts, M., Kenimer, J. G., Charles, I. G., Fairweather, N., Novotny, P. & Brennan, M. J. (1991). Pertactin, an Arg-Gly-Asp-containing *Bordetella pertussis* surface protein that promotes adherence of mammalian cells. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 345–349.

Leininger, E., Ewanovich, C. A., Bhargava, A., Peppler, M. S., Kenimer, J. G. & Brennan, M. J. (1992). Comparative role of the Arg-Gly-Asp sequence present in the *Bordetella pertussis* adhesins pertactin and filamentous hemagglutinin. *Infect. Immunol.* **60**, 2380–2385.

Leininger, E., Probst, P. G., Brennan, M. J. & Kenimer, J. G. (1993). Inhibition of *Bordetella pertussis* filamentous hemagglutinin-mediated cell adherence with monoclonal antibodies. *FEMS Microbiol. Letters*, **106**, 31–38.

Leontovich, A. M., Brodsky, L. I. & Gorbalenya, A. E. (1990). Construction of full local similarity map of two biopolymers. *Biopolymers Cells*, **6**, 14–22.

Locht, C., Bertin, P., Menozzi, F. D. & Renauld, G. (1993). The filamentous haemagglutinin, a multifaceted adhesin produced by virulent *Bordetella* spp. *Mol. Microbiol.* **9**, 653–660.

Lopez, J. A., Chung, D. W., Fujikawa, K., Hagen, F. S., Davie, E. & Roth, G. (1988). The alpha and beta chains of human platelet glycoprotein Ib are both transmembrane proteins containing a leucine-rich amino-acid sequence. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 2135–2139.

Makhov, A. M., Trus, B. L., Conway, J. F., Simon, M. N., Zurabishvili, T. G., Mesyanzhinov, V. V. & Steven, A. C. (1993). The short tail-fiber of bacteriophage T4: molecular structure and a mechanism for its conformational transition. *Virology*, **194**, 117–127.

Malvar, T., Biron, R. W., Kaback, D. B. & Denis, C. L. (1992). The CCR4 protein from *Saccharomyces cerevisiae* contains a leucine-rich repeat region which is required for its control of ADH2 gene expression. *Genetics*, **132**, 951–962.

Menozzi, F. D., Gantiez, C. & Locht, C. (1991). Interaction of the *Bordetella pertussis* filamentous hemagglutinin with heparin. *FEMS Microbiol. Letters*, **78**, 59–64.

Mooi, F. R., van der Heide, H. G. J., Ter Avest, A. R., Welinder, K. G., Livey, I., van der Zeijst, B. A. M. & Gaastra, W. (1987). Characterization of fimbrial subunits from *Bordetella* species. *Microbiol. Pathol.* **2**, 473–484.

Mosesson, M. W., Hainfeld, J., Wall, J. & Haschmeyer, R. H. (1981). Identification and mass analysis of human fibrinogen molecules and their domains by scanning transmission electron microscopy. *J. Mol. Biol.* **153**, 695–718.

Oda, M., Cowell, J. L., Burstyn, D. G. & Manclark, C. R. (1984). Protective activities of the filamentous hemagglutinin and the lymphocytosis-promoting factor of *Bordetella pertussis* in mice. *J. Infect. Diseases*, **150**, 823–833.

Perczel, A., Hollosi, M., Tusnady, G. & Fasman, G. D. (1991). Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins. *Protein Eng.* **4**, 669–679.

Ptitsyn, O. B. & Finkelstein, A. V. (1983). Theory of protein secondary structure and algorithm of its prediction. *Biopolymers*, **22**, 15–25.

Rappuoli, R., Pizza, M., Covacci, A., Bartolini, A., de Magistris, T. M. & Nencioni, L. (1992). Progress towards the development of new vaccines against whooping cough. *Vaccine*, **10**, 1027–1032.

Reinke, R., Krantz, D. E., Yen, D. & Zipursky, S. L. (1988). Chaoptin, a cell surface glycoprotein required for *Drosophila* photoreceptor cell morphogenesis, contains a repeat motif found in yeast and human. *Cell*, **52**, 291–301.

Relman, D. A., Domenighini, M., Tuomanen, E., Rappuoli, R. & Falkow, S. (1989). Filamentous hemagglutinin of *Bordetella pertussis*, nucleotide sequence and crucial role in adherence. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 2637–2641.

Relman, D., Tuomanen, E., Falkow, S., Golenbock, D. T., Saukkonen, K. & Wright, S. D. (1990). Recognition of a bacterial adhesin by an integrin: macrophage CR3 (alfa M-beta 2, CD11b/CD18) binds filamentous hemagglutinin of *Bordetella pertussis*. *Cell*, **61**, 1375–1382.

Rothberg, J. M., Jacobs, J. M., Goodman, C. S. & Artavanis-Tsakonas, S. (1990). Slit: an extracellular protein necessary for development of midline glia and commisural axon pathways contains both EGF and LRR domains. *Genes Develop.* **4**, 2169–2187.

Rutter, D. A., Ashworth, L. A. E., Day, A., Funnell, S., Lovell, F. & Robinson, A. (1988). Trial of a new acellular pertussis vaccine in healthy adult volunteers. *Vaccine*, **6**, 29–32.

Sargent, D., Benevides, J. M., Yu, M. H., Kiung, J. & Thomas, G. J. (1988). Secondary structure and thermostability of the phage P22 tailspike. XX.

Analysis by Raman spectroscopy of the wild-type protein and a temperature-sensitive folding mutant. *J. Mol. Biol.* **199**, 491–502.

Sato, Y., Cowell, J. L., Sato, H., Burstyn, D. G. & Manclark, C. R. (1983). Separation and purification of the hemagglutinins from *Bordetella pertussis*. *Infect. Immunol.* **41**, 313–320.

Shahin, R. D., Amsbaugh, D. F. & Leef, M. F. (1992). Mucosal immunization with filamentous hemagglutinin protects against *Bordetella pertussis* respiratory infection. *Infect. Immunol.* **60**, 1482–1488.

Steven, A. C., Trus, B. L., Maizel, J. V., Unser, M., Parry, D. A. D., Wall, J. S., Hainfeld, J. F. & Studier, F. W. (1988). Molecular substructure of a viral receptor recognition protein. The gp17 tail-fiber of bacteriophage T7. *J. Mol. Biol.* **200**, 351–365.

Steven, A. C., Kocsis, E., Unser, M. & Trus, B. L. (1991). Spatial disorders and computational cures. *Int. J. Biol. Macromol.* **13**, 174–180.

Stouten, P. F. W., Sander, C., Ruigrok, R. W. H. & Cusack, S. (1992). New triple-helical model for the shaft of the adenovirus fiber. *J. Mol. Biol.* **226**, 1073–1084.

Takahashi, N., Takahashi, Y. & Putnam, F. W. (1985). Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich alpha2-glycoprotein of human serum. *Proc. Nat. Acad. Sci., U.S.A.* **82**, 1906–1910.

Trus, B. L. & Steven, A. C. (1981). Digital image processing of electron micrographs—the PIC system. *Ultramicroscopy*, **6**, 383–386.

Tuomanen, E. & Weiss, A. (1985). Characterization of two adhesins of *Bordetella pertussis* for human ciliated respiratory-epithelial cells. *J. Infect. Diseases*, **152**, 118–125.

Tuomanen, E., Towbin, H., Rosenfelder, G., Braun, D., Larson, G., Hansson, G. & Hill, R. (1988). Receptor analogs and monoclonal antibodies that inhibit adherence of *Bordetella pertussis* to human ciliated respiratory epithelial cells. *J. Exp. Med.* **168**, 267–277.

Unser, M., Steven, A. C. & Trus, B. L. (1986). Odd men out: a quantitative objective procedure for identifying anomalous members of a set of noisy images of ostensibly identical specimens. *Ultramicroscopy*, **19**, 337–348.

Unser, M., Trus, B. L. & Steven, A. C. (1987). A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy*, **23**, 39–52.

Unser, M., Trus, B. L. & Steven, A. C. (1989). Normalization procedures and factorial representations for classification of correlation-aligned images: a comparative study. *Ultramicroscopy*, **30**, 299–310.

Urisu, A., Cowell, J. L. & Manclark, C. R. (1986). Filamentous hemagglutinin has a major role in mediating adherence of *Bordetella pertussis* to human WiDr cells. *Infect. Immunol.* **52**, 695–701.

Valentine, R. C. & Green, N. M. (1967). Electron microscopy of an antibody-hapten complex. *J. Mol. Biol.* **27**, 615–617.

van Oostrum, J. & Burnett, R. M. (1985). Molecular composition of the adenovirus type 2 virion. *J. Virol.* **56**, 439–448.

Wall, J. S. (1979). Mass measurements with the electron microscope. In *Introduction to Analytical Electron Microscopy* (Hren, J. J., Goldstein, J. I. & Joy, D. C., eds), pp. 333–342, Plenum Publishing Corp., New York.

Wall, J. S. & Hainfeld, J. F. (1986). Mass mapping with the scanning transmission electron microscope. *Annu. Rev. Biophys. Chem.* **15**, 355–376.

Wall, J. S., Hainfeld, J. F. & Chung, K. D. (1985). Films that wet without glow-discharge. Proc. 43rd Ann. Mtg. EMSA (Bailey, G. W., ed.), pp. 716–717, San Francisco Press, San Francisco.

Weiss, A. A. & Hewlett, E. L. (1986). Virulence factors of *Bordetella pertussis*. *Annu. Rev. Microbiol.* **40**, 661–686.

Willems, R. J. L., Geuijen, C., van der Heide, H. G. J., Renauld, G., Bertin, P., van den Akker, W. M. R., Locht, C. & Mooi, F. R. (1994). Mutational analysis of the *Bordetella pertussis* fim/fha gene cluster: identification of a gene with sequence similarities to haemolysin accessory genes involved in export of FHA. *Mol. Microbiol.* **11**, 337–347.

Wingfield, P. T., Stahl, S. J., Payton, M. A., Venkatesan, S., Misra, M. & Steven, A. C. (1991). HIV-1 Rev expressed in recombinant *Escherichia coli*: purification, polymerization and conformational properties. *Biochemistry*, **30**, 7527–7534.

Yoder, M. D., Keen, N. T. & Jurnak, F. (1993*a*). New domain motif: The structure of pectate lyase C, a secreted plant virulence factor. *Science*, **260**, 1503–1507.

Yoder, M. D., Lietzke, S. E. & Jurnak, F. (1993*b*). Unusual structural features in the parallel β-helix in pectate lyases. *Structure*, **1**, 241–251.

Edited by F. Cohen

(Received 30 September 1993; accepted 18 April 1994)