Volume 215, number 1, 145–150 FEB 04648 May 1987

Predicted structure of tail-fiber proteins of T-even type phages

Isolde Riede, Heinz Schwarz and Fritz Jähnig

Max-Planck-Institut für Biologie, Corrensstrasse 38, D-7400 Tübingen, FRG

Received 22 February 1987

The sequences of the tail fiber protein 36 of the phages T_4, T_2, K_3, and Ox_2 were analyzed for homologies and for folding patterns using structure prediction methods. No repeating motif was found. A model for the fiber structure is proposed in which β-strands of about 6 amino acids are separated by turns. In the β-strand, hydrophobic amino acids are found alternating with hydrophilic ones. Such amphipathic β-strands can be stabilized by dimer formation. The dimerization occurs in a parallel fashion so that both N-termini are at one end of the dimer. This structure represents a rigid fiber. Our model is consistent with electron microscopic data and electron diffraction patterns for the T_4 tail fiber. The observation that all fiber components are found as dimers supports our model. Sequences of the receptor recognition proteins 38 of T-even type phages reveal an architecture different from the architecture of the fiber proteins 36 and 37 of these phages.

Tail fiber protein; T-even type phage; Structure prediction

1. INTRODUCTION

The long tail fibers of T-even type phages serve to attach the phage to the cell wall of *Escherichia coli*. These fibers are about 30 Å thick and 1500 Å long with a kink in the middle [1]. The protein composition of the tail fibers of phages T_2, Ox_2 and K_3 is shown in fig.1. Gene product (gp) 38 of T_4 phage apparently modifies a tail fiber protein without assembling to the tail fiber [2]. In contrast, gp38 of T_2, Ox_2 and K_3 is a constituent of the respective tail fibers.

Upon binding antibodies to a hybrid T_2-T_4 phage, Beckendorf [3] showed that gp37 is colinear with the distal half of the fiber, the C-terminus of the protein near the distal tip. Earnshaw et al. [4] examined distal half fibers of T_4, including proteins 36 and 37, in greater detail, using electron microscopy and electron and X-ray diffraction. They proposed a model for the structure of the fiber with a polypeptide chain in cross-β-conformation transverse to the axis of the fiber.

The principles involved in designing a long and relatively rigid structure are interesting, particularly in view of the absence of the α-helix structure. We therefore decided to work out a precise model for T-even fiber proteins, in order to understand the rigidity of the fiber on the basis of amino acids (aa). For that reason, the amino acid sequences of the fiber proteins 36 [5] of phages T_4, T_2, K_3 and Ox_2 were aligned and analyzed with the help of structure prediction methods.

Correspondence (present) address: I. Riede, European Laboratory of Molecular Biology, Meyerhofstr. 1, 6900 Heidelberg, FRG

2. RESULTS AND DISCUSSION

2.1. *Comparison of the aligned protein 36 sequences*

Protein 36 of T_4, T_2, K_3 and Ox_2 can be divided into three domains. The two highly conserved regions at the beginning and the end of the protein (more than 50 aa) are thought to mediate binding to the neighboring proteins (fig.1) [5]. The highly

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

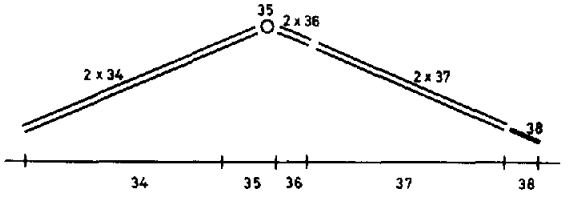

Fig.1. Composition of the long tail fibers of phages T_2, K_3 and Ox_2 from the baseplate (left) to the receptor recognizing tip (right). The five structural genes are clustered in the genome [2].

variable middle region contains 100 aa [5]. Comparing the proteins from the different phages, one observes that leucines, valines, isoleucines, methionines, tryptophans and phenylalanines replace each other. Strikingly, the content of the most hydrophobic amino acids is very constant, whereas the content of other groups of amino acids varies. This finding supports a model, in which the hydrophobic amino acids are involved in structure formation. Similarly, the extraordinarily high and constant content of glycine in these proteins, and indeed in all tail fiber proteins so far analyzed, indicates an important role also for this amino acid. Glycines are frequently found clustered with serine residues and in some cases either the glycine or the serine may be substituted by proline (see strand IV–V in fig.2). Thus a role for these amino acids in the turn formation of the polypeptide chains is suggested.

2.2. *Absence of repetitions in proteins 36*

Proteins folding in cross-β-sheet structure, like the adenovirus fiber [6], frequently contain a repeating motif which can be correlated with repetition in the secondary structure. Analysis of T_4 protein 36 (and 37) sequences has already excluded the possibility that such repetitions exist within these proteins [7]. Sequence comparisons of T_2, K_3 and Ox_2 gp36 with the program DOTPLOT [8] revealed no repetitive motif in any of the protein 36 sequences (not shown). Therefore, another approach to understand the molecular structure had to be taken.

2.3. *Structural predictions*

Structure prediction plots for Ox_2 protein 36 are shown in fig.3. The other gp36 (and gp37) exhibit similar patterns. Secondary structure predictions according to Chou and Fasman [9] (fig.3A) reveal about 27 segments regularly separated by turns (fig.3B) with β-potential and turn potential alternating. Most of the protein is predicted to be in β-structure and only about 10 aa are predicted to be in α-structure. Thus, the prediction is consistent with the high content of β-structure observed by Earnshaw et al. [4].

The hydrophobicity plot according to Kyte and Doolittle [10] (fig.3C) revealed that the most hydrophobic amino acids are evenly distributed within the sequence. Secondly, the amphipathic β-strand profile (fig.3E) shows a regular zigzag pattern, which indicates amphipathic β-strands [11]. The regular pattern of sided β-strands is interrupted by regions of non-amphipathic structure. These regions often coincide with predicted turns (fig.3B) [9]. Fig.3D shows an amphipathic α-helix profile. Only aa 20–30 and 110–120 show a regular zigzag pattern typical of amphipathic α-helices. However, since according to the Chou and Fasman predictions, these regions exhibit a high β-turn potential, it is unlikely that they form α-helices.

In summary, hydrophobic amino acids alternate with non-hydrophobic amino acids to form an amphipathic β-sheet.

2.4. *Proposed folding of proteins 36*

From the structural predictions for all four phage proteins the folded structure of the polypeptide chains is proposed as outlined in fig.2. Hydrophobic amino acids are circled and turn predictions are indicated by arrows. β-Strands of about 6 aa alternate with turns of about 4 aa. From this model, the role of the most hydrophobic amino acids and of the glycines can be deduced: the most hydrophobic aa are almost exclusively located in the β-strands whereas the glycines are involved in turns.

2.5. *Expectations from electron microscopy*

gp36 contributes 130 Å to the tail fiber of 30 Å thickness [1]. Taking into account that 1 aa in a β-strand is about 3.5 Å long, a strand of 30 Å length (thickness of the fiber) should contain about 8 aa. Since the distance between two β-strands is about 4.7 Å, 28 β-strands would be required for a fiber segment of 130 Å. Thus, the folding of the proteins proposed in fig.3 fits exactly with the dimension of the fiber.

In a hydrophilic environment, an amphipathic β-strand may fold back on itself or dimerize in a parallel or antiparallel fashion to shield the hydrophobic amino acid from water. Since the phage tail fiber proteins are always found as dimers, the second possibility is more likely. For gp37 it has already been shown that the proteins are arranged in parallel with the C-terminus at the

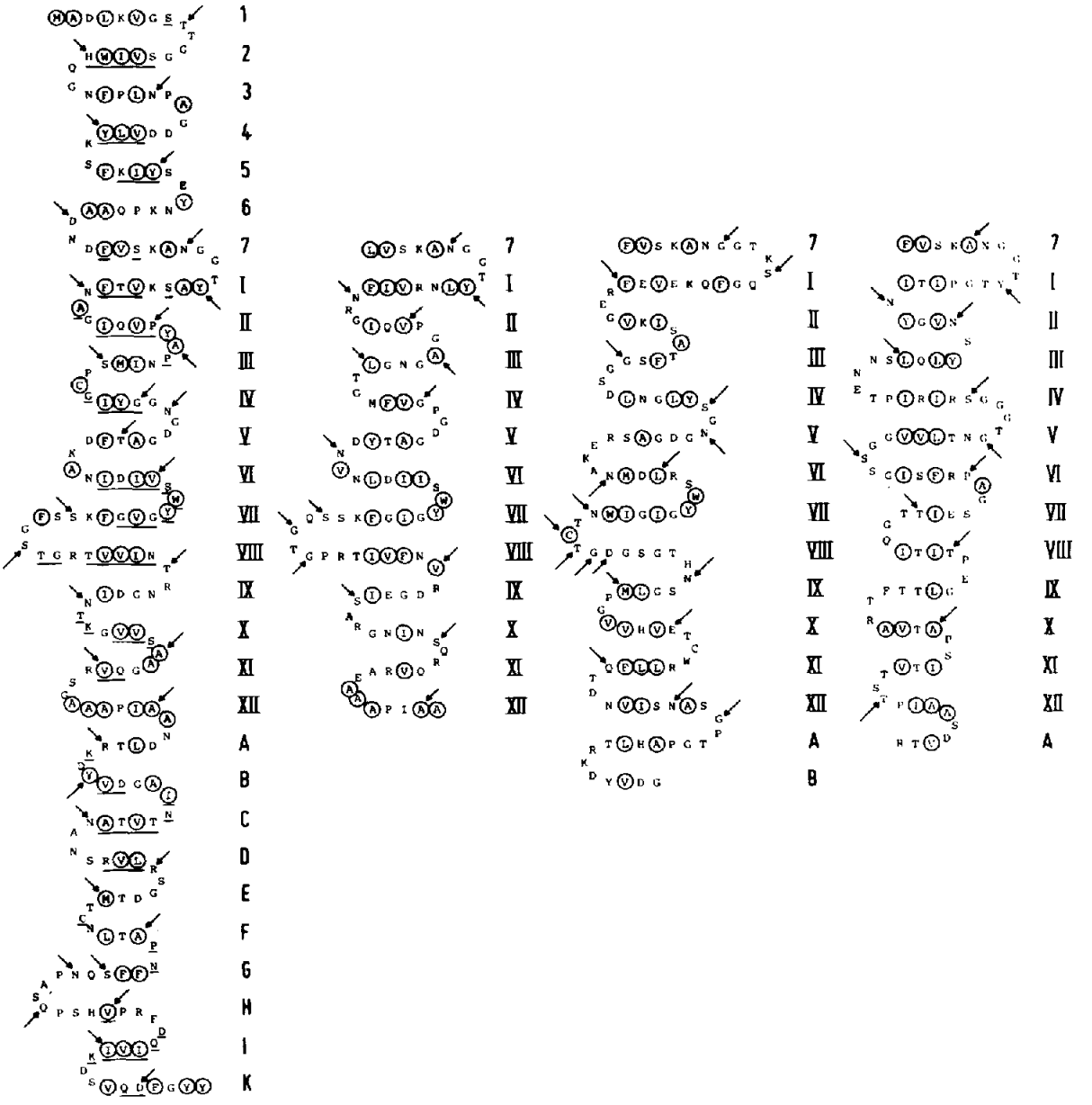

Fig.2. Proposed folding of gp36 of the phages T_4, T_2, K_3 and Ox_2 (from left to right). Because of the similarities in the conserved regions, β-strands 1–7 (N-terminal strands, conserved) and A–K (C-terminal strands, conserved) are only outlined for the T_4 protein. Hydrophobic amino acids are circled, amino acids predicted in turns are indicated by arrows, amino acids predicted to be in β-structure are underlined only for the T_4 sequence.

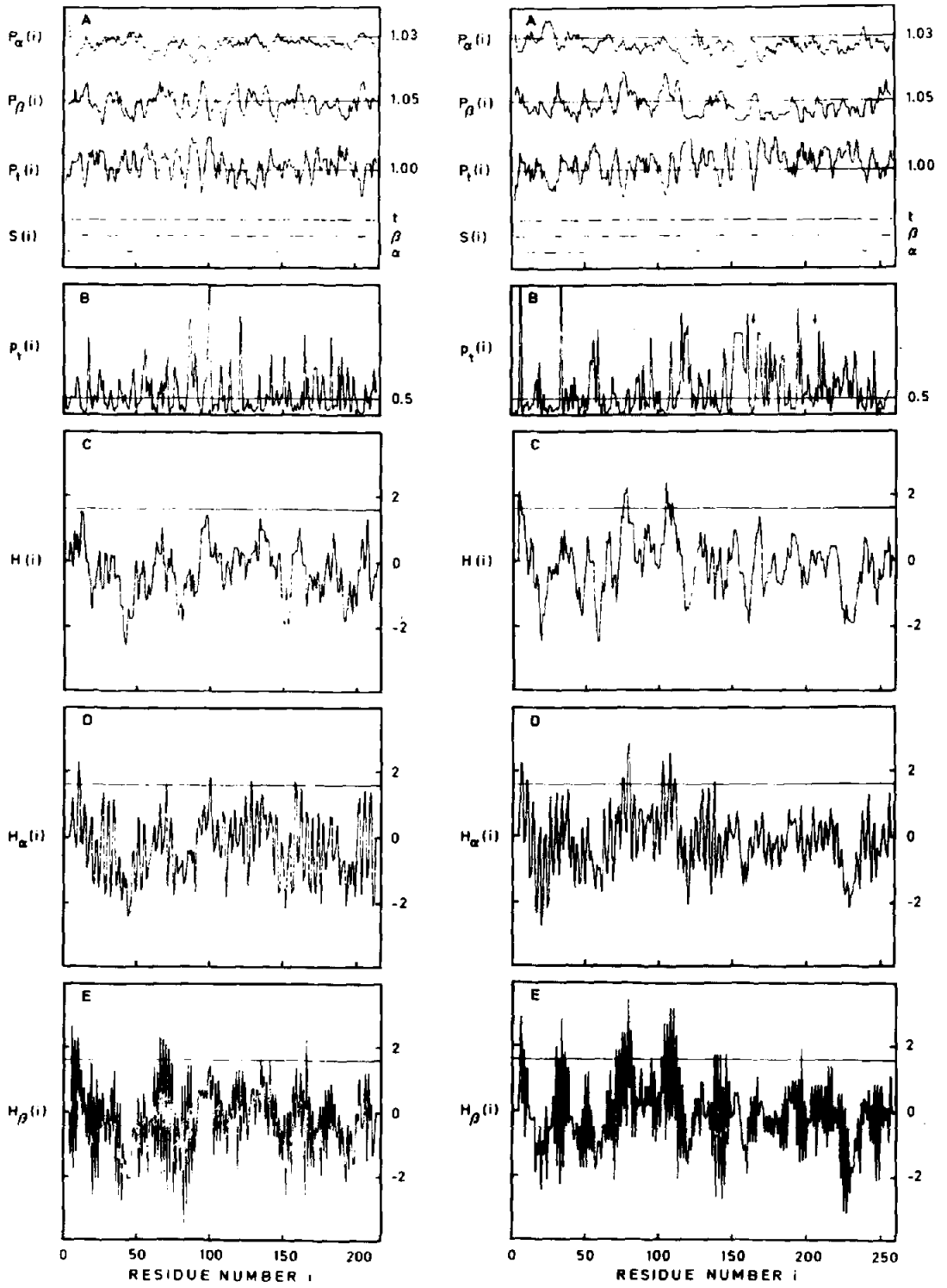
A
$P_\alpha(i)$
$P_\beta(i)$
$P_t(i)$
S(i)
1.03
1.05
1.00
t
β
α
B
$P_t(i)$
0.5
C
H(i)
D
$H_\alpha(i)$
E
$H_\beta(i)$
2
0
-2
0
50
100
150
200
250
RESIDUE NUMBER i

Fig.3. Structure prediction plots for gp36 of bacteriophage Ox_2 (left) and gp38 of K_3 (right). (A) The α-helix potential P_α, β-strand potential P_β, turn potential P_t, and the prediction of α-helix structure ($s = \alpha$), β-strand structure ($s = \beta$), or turn structure (s = t), according to Chou and Fasman [9]. (B) The position-dependent turn potential p_t according to Chou and Fasman [9]. (C) The hydrophobicity H according to Kyte and Doolittle [10]. (D) The hydrophobicity H_α of one side of an α-helix around residue i according to Vogel and Jähnig [11]. (E) The hydrophobicity H_β of one side of a β-strand around residue i according to Vogel and Jähnig [11].

fiber tip [3]. However, nothing is known, so far, about gp34 and gp36. Reinspection of electron micrographs, taken previously to study the serological relationships of different phages [12], reveals that monovalent antibody fragments (Fab) always bind pairwise to the regions of the tail fiber where gp34, 36 and 37 are located (fig.4). Hence, dimerization of gp34 and 36 also occurs in a parallel fashion. This suggests that all tail fiber proteins follow the same principle of parallel dimerization.

2.6. *Model for the fiber structure*

The structural predictions for the tail fiber proteins support a model of two amphipathic cross-β-sheets forming a dimer, which is stabilized by the

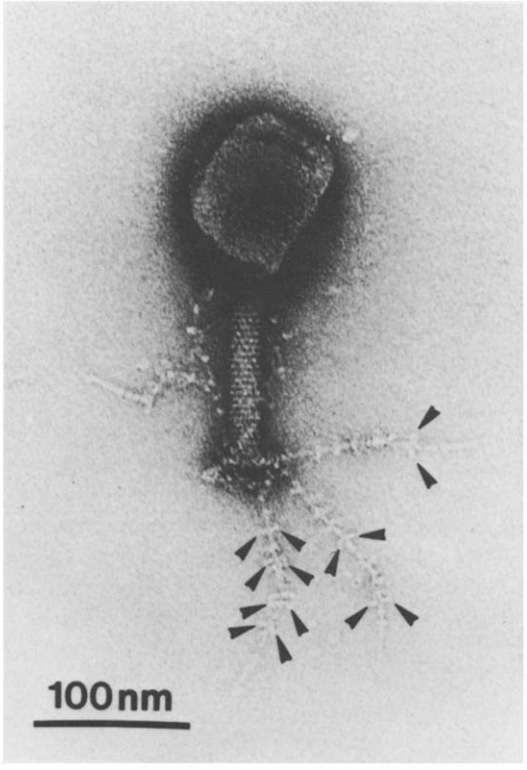

Fig.4. Electron microscopy. Phage Tull*-46 is shown with antibodies against phage Tull*-6 [12]. Monovalent antibody fragments (Fab) (arrowheads) bind always pairwise to regions including gp34, 36 and 37.

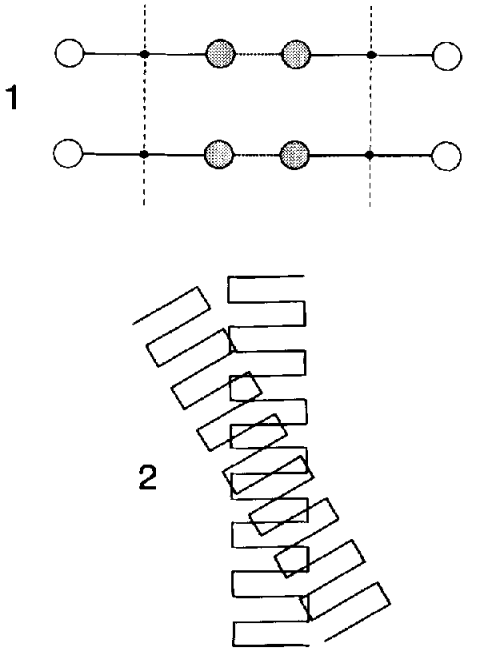

Fig.5. Formation of the dimer gp36. (1) β-Strands shown in fig.3 interact via H-bonds with each other (broken lines). These H-bonds indicate the axis of the fiber. The β-strands are amphipathic, therefore hydrophilic amino acids (open circles) are at the outside and hydrophobic amino acids (dotted circles) on the inside attracting each other via hydrophobic forces (dotted lines). (2) Two β-sheets pack with an angle of 30° on top of each other [13]. If they are amphipathic, they start to twist around each other to shield the hydrophobic sides from contact with water.

hydrophobic interaction between the two monomers (fig.5). According to the finding that antibodies always bind pairwise, the two β-sheets are arranged in parallel.

When β-sheets pack face to face, the angle between the strand directions of the two β-sheets is about 30° [13]. Hence, if two long β-sheets are packed in this way, the contact between the two polypeptides is minimal and thermodynamically unstable (fig.5). To stabilize this structure the two β-sheets may then twist around each other. Such a twisted structure would confer a high rigidity to the fiber. This structure would agree with the proposal of Earnshaw et al. [4], that tail fibers are turned along their axis.

2.7. *Structure of the receptor recognizing proteins 38*

Protein 38, which is located at the fiber tip of T_2, K_3 and Ox_2 and which is responsible for recognition of the bacterial receptor, can be divided into two functional and structural domains: the 100 N-terminal aa which are known to bind to the neighboring protein 37 (fig.1) and the C-terminal 160 aa which are responsible for receptor recognition [14]. Whereas the N-terminus of proteins 38 is rather conserved, the C-terminus is highly divergent within the phages. Also, as with other tail fiber proteins, the N-terminus is unique, but the C-terminus is repetitive. Fig.3 shows the structure prediction plots for protein 38. The regularities identified for tail fiber proteins, β-strands alternating with β-turns to form an amphipathic β-sheet, can be identified in the 100 N-terminal aa.

The second half of the protein, which is responsible for receptor recognition, shows much less regularity. The frequency of turns is especially high. This phenomenon points towards a receptor binding site given by several turns forming a planar surface region as observed for antibody binding sites [15]. Two host range mutations, both of which alter the receptor specificity of the phages, have been localized at the sites indicated by arrows [14] which indeed lie near two predicted β-turns. Thus, the presence of many turns within gp38 make this protein a suitable candidate for a mediator of host binding.

ACKNOWLEDGEMENTS

We are very grateful to U. Henning and S. MacIntyre for critical reading of the manuscript.

REFERENCES

[1] Bishop, R.J., Conley, M.P. and Wood, W.B. (1974) J. Supramol. Struct. 2, 196–201.
[2] Wood, W.B. and Revel, H.R. (1976) Bacteriol. Rev. 40, 847–868.
[3] Beckendorf, S.K. (1973) J. Mol. Biol. 73, 37–53.
[4] Earnshaw, W.C., Goldberg, E.B. and Crowther, R.A. (1979) J. Mol. Biol. 132, 101–131.
[5] Riede, I., Drexler, K. and Eschbach, M.-L. (1985) Nucleic Acids Res. 13, 605–616.
[6] Green, N.M., Wrigley, N.G., Russel, W.C., Martin, S.R. and McLachlan, A.D. (1983) EMBO J. 2, 1357–1365.
[7] Oliver, D.B. and Crowther, R.A. (1981) J. Mol. Biol. 153, 545–568.
[8] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[9] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 5–148.
[10] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[11] Vogel, H. and Jähnig, F. (1986) J. Mol. Biol. 190, 191–199.
[12] Schwarz, H., Riede, I., Sonntag, I. and Henning, U. (1983) EMBO J. 2, 375–380.
[13] Chothia, C. and Janin, J. (1981) Proc. Natl. Acad. Sci. USA 78, 4146–4150.
[14] Riede, I., Drexler, K., Eschbach, M.-L. and Henning, U. (1986) J. Mol. Biol., in press.
[15] Amit, A.G., Mariuzza, R.A., Phillips, S.E.V. and Poljak, R.J. (1985) Nature 313, 156–158.