Evidence for the Homology of the Main Determinant of the Human Encephalitogenic Protein and an Ancestral Histone IV Sequence

By KLAUSDIETER BAUER*

University of Heidelberg, Heidelberg, German Federal Republic

(Received 16 December 1971)

Sequence similarities between proteins, when exceeding random coincidences, are taken to be indicative of common ancestry (Florkin, 1966; Anfinsen, 1959), whether it is the 'same' protein, as it occurs in various species, that is compared, or whether the sequences belong to 'different' protein entities. Examples of the first case are abundant, and include cytochromes *c*, haemoglobin chains, immunoglobulin chains, ferredoxins and viral coat proteins. The second case has been observed less frequently; however, haemoglobin and myoglobin, lysozyme and α -lactalbumin and the serine proteases are well-known examples. A third kind of homology is the internal type, in which certain portions of a polypeptide chain are homologous with other parts of the same sequence, indicating that at least some present-day sequences have evolved through duplication processes from small peptides (Urbain, 1969).

Under favourable conditions all three types of sequence homology allow the reconstruction of ancestral peptides. I now report on a comparative investigation involving two of the homology types mentioned and two proteins differing in many respects from each other.

Ancestral histone IV A peptide

It has been demonstrated that histone IV is mainly composed of repetitions of two dodecapeptides, A and B, which are responsible for some 86% of the total polypeptide chain (Bauer, 1971*a*). The ancestral histone IV A peptide was, moreover, found to be homologous with partial sequences of a variety of non-histone proteins, including protamines, cytochromes *c*, ferredoxins, immunoglobulin L-chains and viral coat proteins (Bauer, 1971*a,b,c*, and unpublished work). This peculiar property of the histone IV A peptide may indicate that it is a very ancient piece of genetic information, as animal, plant, bacterial and viral proteins, many of them in the form of their nodal sequences (Dayhoff, 1969), are connected through their relationship with this sequence.

Experimental autoimmune encephalomyelitis

This disease can be induced in animals by injections of nervous tissue, i.e. autologous, homologous and

* Postal address: Heidelberg, Rohrbacher Strasse 8/III/3, German Federal Republic.

heterologous brain and spinal cord. Paterson (1968) states: 'The fact that certain immune responses made by animals with EAE [experimental autoimmune encephalomyelitis] are also demonstrable in patients with multiple sclerosis together with the histopathologic similarities which exist between the animal and the human disease, suggest that EAE and demyelinating disease of man have common pathogenic mechanisms ... The encephalitogenic activity of nervous tissue is due to organ-specific, rather than species-specific, antigenic constituents', even if cross-reactions with other tissues have been observed.

Encephalitogenic protein and its connexion with histone IV A peptide

Experimental autoimmune encephalomyelitis is of considerable importance and interest as a model for the investigation of human demyelinating disease. It was a major advance when Carnegie (1971) was able, not only to report the total sequence of the (or at least one of the) antigen(s) responsible, but also to distinguish the main determinant connected with the encephalitogenic potency of the protein.

The antigen is a basic protein of 170 amino acid residues; the main encephalitogenic determinant consists of the 11 residues 112–122 of the chain (Carnegie, 1971). In Fig. 1 I have aligned this determinant with histone IV A peptide. With the introduction of one gap there is a very close similarity between the two sequences. One major dissimilarity is the presence in the determinant of a tryptophan residue, which might be expected to be responsible for its antigenicity at least in part, since histones are known to be very weak antigens (Bauer & Beheim, 1968; Bauer, 1968, and unpublished work), as discussed below.

It should be mentioned that the homology with histone IV A peptide is not restricted to the main encephalitogenic determinant itself, but that many more partial sequences of the encephalitogenic protein are also related more or less closely to the sequence of histone IV A peptide (Fig. 1). Altogether a total of 78.8% of the encephalitogenic protein was observed to be homologous with histone IV A peptide, when the criteria described below are applied.

Statistical evaluation

The procedure described previously (Bauer, 1971a,b,c) was followed throughout, with all calculations made at the base level rather than the amino acid level.

(1) In histone IV A peptide an average of 1.39 bases occurs at every base position; therefore random sequences, when aligned under exactly equal conditions (i.e. one sequence being without ambiguity, the other to the same degree as is the A peptide), will coincide with an average probability of $Q_r = 0.32$ ($= 0.25 + 0.39 \times 0.25 - 0.39 \times 0.25^2$) (Bauer, 1971b). The random coincidences follow a binomial distribution. By use of the correction employed in the calculation of Q , it is thus possible to compare actual sequence similarities with a random model that takes into account all ambiguities of the A peptide. I have calculated $q_{\max.}$ values, defined as $q_{\max.} = 1 - (Q_r + 3\sigma)$ and compared them with the relative minimal base deviations, q , for each peptide compared with the A-peptide sequence. Only q values not exceeding $q_{\max.}$ were accepted as evidence for homology, since they are outside the 3σ range of the random-coincidence distribution.

It might be argued that the four bases are not equally distributed now and that therefore the binomial distribution must not be employed in this form. However, with the extremely long periods of evolution time involved the variety of base distributions found at present in different groups of species should with high probability average out.

I should mention that in previous analyses I used the somewhat more direct approach of random models made by use of random numbers and the codon bases involved and directly estimated the σ values (Bauer, 1971a,c), which in all cases examined were observed to be in agreement with the results obtained by the use of the criteria described in the present communication.

(2) In addition, the probability, P_0 , was calculated that an uninterrupted series of coincidences as

observed in each pair of compared sequences occurs at random. With four different elements (bases) this parameter is $P_0 = 4^{-n}$ in a run of length n . When, moreover, this run is considered in comparison with the total sequence, iteration tests can be applied, leading to the probability, α , for such an event to happen by chance (Geigy Tables, 1968).

Although these three parameters are used exclusively when the probability of true homology versus accidental similarity is estimated, both the underlined presentation of amino acids found equally in encephalitogenic protein and in histone IV A peptide (Fig. 1) and the relative amino acid deviations, p , are given, not only to add a less abstract feature to the comparison, but rather to show clearly the necessity of investigating relationships of this type at the base level itself.

Taken together, the three parameters mentioned strongly indicate the non-randomness of the coincidences between the main encephalitogenic determinant and the histone IV A peptide.

Conclusions

The finding reported here that an ancestral histone sequence and the larger part of the encephalitogenic protein are apparently homologous has several implications.

(1) It is another example of the ubiquity of the peculiar genetic information expressed in histone IV A peptide. This supports the concept that at an earlier era of molecular evolution only a very restricted amount of genetic information existed (Zuckerkindl & Pauling, 1965), and, moreover, that gene duplication played an important role in the development that followed this phase (Ohno, 1970).

(2) Carnegie (1971) states in his description of the encephalitogenic protein: 'The protein has a high isoelectric point and is similar in size, shape and lack of helical content to histones ... but differs from histones in amino acid composition ...'. The relative ease with which evidence for a homology between

Fig. 1. Alignment of the sequences of histone IV A peptide and of the various portions of the human encephalitogenic protein that are similar in their primary structures to this peptide

In the first position the main encephalitogenic determinant is aligned with histone IV A peptide. For each comparison the statistical parameters (calculated as described in the text) are given. Numbers in parentheses with q are the $q_{\max.}$ values, indicating the 3σ limit of a random distribution of coincidences that is corrected for the ambiguities of the histone IV A peptide. All q values but one are smaller than $q_{\max.}$; they are thus outside the 3σ range. For further details see the text. K is the proportion of the amino acids of histone IV A peptide in the aligned peptides. Underlining indicates amino acids common to the A peptide and the partial sequences from encephalitogenic protein. The statistical treatment was at the base level. Key to one-letter amino acid code: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine; - and X, gap (Dayhoff, 1969). Other abbreviations: α , A or U; β , C or G; γ , A, C or U; n , A, U, C or G.

the two proteins can be established when partial sequences are compared, as in the present investigation, should encourage more comparative analyses of this type: first, a reconstruction of ancestral peptides is made by an analysis of internal homologies; in a second step such peptides from different proteins are tested for homology.

(3) It has already been mentioned that histones are very weak immunogens and antigens (Bauer & Beheim, 1968; Bauer, 1968, and unpublished work). Histone IV A peptide is known to participate in the structure of at least three different histone molecules to various degrees, i.e. in its histone modifications this sequence is apparently either only a very weak or no antigenic determinant at all. In the encephalitogenic protein, on the other hand, it becomes a very important and prominent antigenic structure. This is probably due to the introduction of one tryptophan residue into the peptide, since heterocyclic (and aromatic) amino acids are known to be of great importance for protein antigenicity. At the base level a single exchange is responsible for the introduction of this key residue, which makes experimental autoimmune encephalomyelitis a molecular disease of a new immunological type.

(4) The relationship between the encephalitogenic protein and histone IV A peptide and thus with a variety of other proteins may be connected with its pathogenic abilities. For example, a considerable portion of the encephalitogenic protein is homologous with not much less than half of the histone IV molecule. Histones are known to influence mem-

brane permeability considerably (Bauer, 1968). It may be that the encephalitogenic protein fulfils a similar function in the normal state, but that this is disturbed in the experimental disease. A study of its relationship with histone IV A peptide and its associated group of proteins might be one way to attack this problem, which is of paramount importance for neurology.

- Anfinsen, C. B. (1969) *The Molecular Basis of Evolution*, John Wiley and Sons, New York
 Bauer, K. (1968) *Deut. Med. J.* **19**, 425
 Bauer, K. (1971a) *Int. J. Protein Res.* **3**, 165
 Bauer, K. (1971b) *Naturwissenschaften* **58**, 364
 Bauer, K. (1971c) *Int. J. Protein Res.* in the press
 Bauer, K. & Beheim, P. (1968) *Z. Krebsforsch.* **70**, 241
 Carnegie, P. R. (1971) *Biochem. J.* **123**, 57
 Dayhoff, M. O. (1969) *Atlas of Protein Sequence and Structure*, vol. 4, National Biomedical Research Foundation, Silver Spring
 Florkin, M. (1966) *A Molecular Approach to Phylogeny*, Elsevier, Amsterdam, London and New York
 Geigy Tables (1968) *Wissenschaftliche Tabellen*, 7th. edn., J. R. Geigy A.G., Basle
 Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag, Berlin, Heidelberg and New York
 Paterson, P. K. (1968) in *Textbook of Immunopathology*, vol. 1: *Experimental Autoimmune (Allergic) Encephalomyelitis* (Miescher, P. A. & Müller-Eberhard, H. J., eds.), p. 132, Grune and Stratton, New York and London
 Urbain, J. (1969) *Biochem. Genet.* **3**, 249
 Zuckerkandl, E. & Pauling, L. (1965) in *Evolving Genes and Proteins* (Bryson, V. & Vogel, H. J., eds.), p. 97, Academic Press, New York