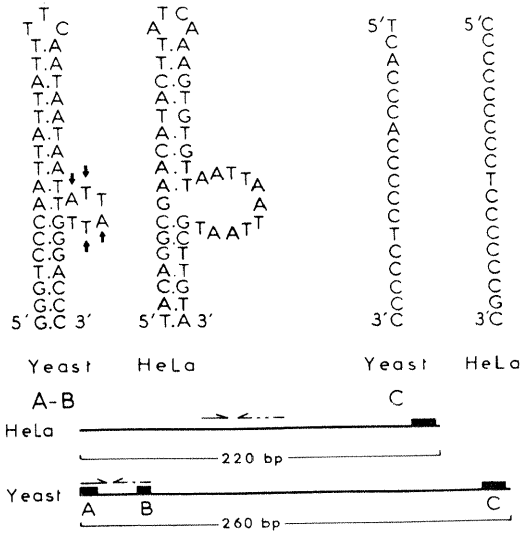

Fig. 4 Comparison of *ori* sequences of mitochondrial genomes from yeast (present work) and HeLa cells[28]. Homology of potential secondary structure is found for the inverted repeats in the A–B region (arrows indicate the base changes found in this region in different petite genomes). Homology of primary structure is found for cluster C. The bottom compares the two *ori* sequences; the arrows indicate the inverted repeats of the A–B region, the broken line corresponding to the looped-out sequence. bp, Base pairs.

Two explanations have been put forward to account for suppressivity. The first one proposes a replicative advantage of the mitochondrial genome of suppressive petites over that of wild-type cells[33–36]. It was directly inspired by the work of Mills *et al.*[37] on the replication of Qβ DNA but was not accompanied by any molecular model. The second one proposes a destructive recombination of the petite genome with the wild-type genome[38–42], and predicts that a number of different petite genomes are formed as the consequence of the increased parental genome instability due to the insertion of the petite genome. The present results contradict this latter explanation because most of the diploid petites studied here had genomes identical to those of the parental petites. Indeed, they provide for the first time a precise molecular basis for the former explanation of the replicative competition.

We thank Miss Claire Gaillard, Dr René Goursot and Mr Alain Huyard for help in mapping and sequencing studies, Miss Martine Brient for typing this manuscript and Mr Philippe Breton for the artwork.

Received 12 January; accepted 20 March 1981.

1. Ephrussi, B. in *Unités Biologiques Douées de Continuité Génétique*, 165–180 (Editions du CNRS, Paris, 1949).
2. Ephrussi, B. in *Nucleocytoplasmic Relations in Microorganisms*, 13–47 (Clarendon, Oxford, 1952).
3. Ephrussi, B., de Margerie-Hottinguer, H. & Roman, H. *Proc. natn. Acad. Sci. U.S.A.* **41**, 1065–1071 (1955).
4. Bernardi, G. *Trends biochem. Sci.* **4**, 197–201 (1979).
5. Faugeron-Fonty, G. *et al. J. molec. Biol.* **134**, 493–537 (1979).
6. Gaillard, C. & Bernardi, G. *Molec. gen. Genet.* **174**, 335–337 (1979).
7. Gaillard, C., Strauss, F. & Bernardi, G. *Nature* **283**, 218–220 (1980).
8. Baldacci, G., de Zamaroczy, M. & Bernardi, G. *FEBS Lett.* **114**, 234–236 (1980).
9. Bernardi G. & Bernardi, G. *FEBS Lett.* **115**, 159–162 (1980).
10. Bernardi, G. *et al.* in *DNA Recombination, Interactions and Repair* (eds Zadrazil, S. & Sponar, J.) 77–84 (Pergamon, Oxford, 1980).
11. Bernardi, G. *et al.* in *Mobilization and Reassembly of Genetic Information* (eds Scott, W. A., Werner, R., Joseph, D. R. & Schultz, J.) 119–132 (Academic, New York, 1980).
12. Bernardi, G. *et al.* in *The Organization and Expression of the Mitochondrial Genome* (eds Kroon, A. M. & Saccone, C.) 21–31 (Elsevier, Amsterdam, 1980).
13. Prunell, A. & Bernardi, G. *J. molec. Biol.* **110**, 53–74 (1977).
14. de Zamaroczy, M., Baldacci, G. & Bernardi, G. *FEBS Lett.* **108**, 429–432 (1979).
15. Goursot, R., de Zamaroczy, M., Baldacci, G. & Bernvardi, G. *Curr. Genet.* **1**, 173–176 (1980).
16. Nobrega, F. G. & Tzagoloff, A. *J. biol. Chem.* **255**, 9828–9837 (1980).
17. Macino, G. & Tzagoloff, A. *Cell* **20**, 507–517 (1980).
18. Sanders, J. P. M., Heyting, C., Verbeet, M. Ph., Meijlink, F. C. P. W. & Borst, P. *Molec. gen. Genet.* **157**, 239–261 (1977).
19. Wesolowski, M., Monnerot, M. & Fukuhara, H. *Curr. Genet.* **2**, 121–129 (1980).
20. Tabak, H. F., Hecht, N. B., Menke, M. M. & Hollenberg, C. P. *Curr. Genet.* **1**, 33–43 (1979).
21. Coruzzi, G. & Tzagoloff, A. *J. biol. Chem.* **254**, 9324–9330 (1979).
22. Heyting, C. *et al. Molec. gen. Genet.* **168**, 231–250 (1979).
23. Thalenfeld, B. E. & Tzagoloff, A. *J. biol. Chem.* **255**, 6173–6180 (1980).
24. Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E. & Tzagoloff, A. *J. biol. Chem.* (in the press).
25. Tzagoloff, A., Nobrega, M., Akai, A. & Macino, G. *Curr. Genet.* **2**, 149–157 (1980).
26. Berlani, R. E., Bonitz, S. G., Coruzzi, G., Nobrega, M. & Tzagoloff, A. *Nucleic Acids Res.* **8**, 5017–5030 (1980).
27. Mathews, S., Schweyen, R. J. & Kaudewitz, F. in *Mitochondria 1977* (eds Bandlow, W. *et al.*) 133–138 (de Gruyter, Berlin, 1977).
28. Crews, S., Ojala, D., Posakony, J., Nishiguchi, J. & Attardi, G. *Nature* **277**, 192–198 (1979).
29. Gillum, A. M. & Clayton, D. A. *J. molec. Biol.* **135**, 353–368 (1979).
30. Kobayashi, M., Yagimuma, K., Seki, T. & Koike, K. in *The Organization and Expression of the Mitochondrial Genome* (eds Kroon, A. M. & Saccone, C.) 221–229 (Elsevier, Amsterdam, 1980).
31. Blanc, H. & Dujon, B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3942–3946 (1980).
32. Gingold, E. B. *10th Int. Conf. of Yeast Genetics and Molecular Biology*, Abstr. no 333, 143 (1980).
33. Carnevali, F., Morpurgo, G. & Tecce, G. *Science* **163**, 1331–1333 (1969).
34. Rank, G. H. *Can. J. Genet. Cytol.* **12**, 129–136 (1970).
35. Rank, G. H. *Can. J. Genet. Cytol.* **12**, 340–346 (1970).
36. Rank, G. H. & Bech-Hansen, N. T. *Can. J. Microbiol.* **18**, 1–7 (1972).
37. Mills, D. R., Peterson, R. L. & Spiegelman, S. *Proc. natn. Acad. Sci. U.S.A.* **58**, 217–224 (1967).
38. Coen, D., Deutsch, J., Netter, P., Petrochilo, E. & Slonimski, P. P. in *Control of Organelle Development*, 24th Symp. Soc. exp. Biol., 449–496 (Cambridge University Press, London, 1970).
39. Deutsch, J. *et al. Genetics* **76**, 195–219 (1973).
40. Michaelis, G., Petrochilo, E. & Slonimski, P. P. *Molec. gen. Genet.* **123**, 51–65 (1973).
41. Perlman, P. S. & Birky, C. W. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4612–4616 (1974).
42. Slonimski, P. P. & Lazowska, J. in *Mitochondria 1977* (eds Bandlow, W., Schweyen, R. J., Wolf, K. & Kaudewitz, F.) 39–52 (de Gruyter, Berlin, 1977).
43. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).

Structure of C-terminal half of two H–2 antigens from cloned mRNA

F. Brégégère*, J. P. Abastado*, S. Kvist†, L. Rask*§, J. L. Lalanne*, H. Garoff†, B. Cami*, K. Wiman‡, D. Larhammar‡, P. A. Peterson‡, G. Gachelin*, P. Kourilsky* & B. Dobberstein†

*Unité de Biologie Moléculaire du Gène, ER CNRS no. 201 and SCN INSERM no. 20, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cédex 15, France
†European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 102 209, D-6900 Heidelberg, FRG
‡Department of Cell Research, The Wallenberg Laboratory, University of Uppsala, Uppsala, Sweden

The classical cell-surface histocompatibility antigens (H–2 antigens in the mouse), known to have key roles in cell-to-cell recognition[1], are encoded by at least three highly polymorphic genes (*H–2D*, *K* and *L*)[2]. Like their human (HLA) counterparts[3], H–2 heavy chains span the cell membrane with a short C-terminal cytoplasmic region and an N-terminal extracellular stretch of about 280 amino acids. HLA antigens seem to be organized in three domains containing β-pleated sheets, with disulphide loops within the second and third domains, but the relative scarcity of material has hampered biochemical studies of the H–2 antigens[4–6]. We now report the sequencing of plasmids carrying H–2 cDNA as a means of inferring the amino acid sequence of the antigens, and especially of their previously poorly described C-terminal half.

The isolation of recombinant plasmids pH–2^d-1 and pH–2^d-3 is described in Fig. 1 legend and elsewhere[7]. Restriction maps of the cDNA inserts, 1,150 and 980 base pairs (bp) long, respectively, are different, but can be tentatively aligned on *Pvu*II, *Sst*I and *Pst*I sites (Fig. 1). Both inserts contain a noncoding stretch of about 480 bp next to the poly(A) sequence. The 627- and 479-bp long coding sequences and their corresponding amino acid sequences are given in Fig. 2. They show extensive homologies with available sequences of H–2 and HLA molecules (82% with H–$2K^b$, 73% with HLA B7) (Fig. 3)[3–6,8–11], allowing unequivocal alignment in the third domain. With reference to HLA[3], we assigned nucleotide 133 to the first tryptophan residue in pH–2^d-1 and nucleotide 181 to the first arginine in pH–2^d-3. Both clones should, accordingly, code for the entire third domain, the membrane spanning region and the cytoplasmic segment.

§Present address: Department of Cell Research, The Wallenberg Laboratory, University of Uppsala, Uppsala, Sweden.

0028–0836/81/270078—04$01.00

Nature Vol. 292 2 July 1981

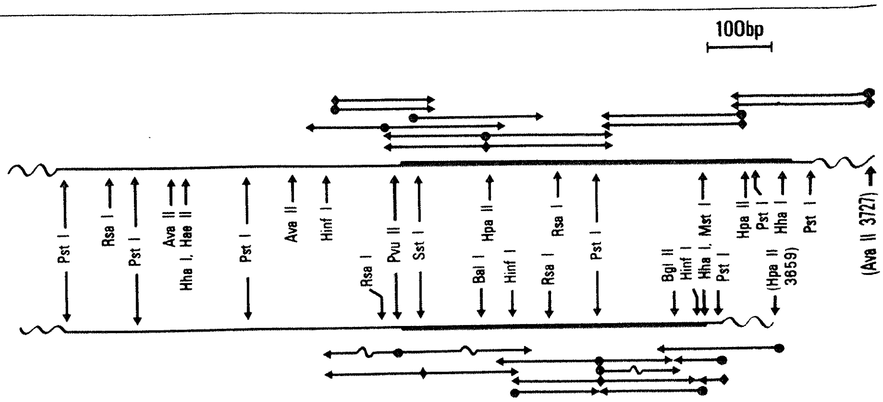

Fig. 1 Restriction maps of pH-2^d-1 and pH-2^d-3 inserts and strategies to sequence them. The cDNA library from which pH-2^d-1 (ref. 7) was selected had been constructed using mRNA from SL2 lymphoma cells grown as ascites in DBA/2 mice (H-2^d haplotype). The 400 independent bacterial clones of this library were further screened by *in situ* hybridization[29], using a fragment of the insert of the first H-2 clone as a probe. The DNAs of the positive responders were then tested for the specific binding of H-2 mRNA as already described[7]. At least two of them, pH-2^d-2 and pH-2^d-3, were found positive in this test. The cDNA insert of pH-2^d-2 was found identical to part of that of pH-2^d-1, and was not analysed further. Plasmid DNA was prepared from cleared lysates[30], partially purified by centrifugation in a CsCl/ethidium bromide gradient, and further purified by fractionation through a 5–40% sucrose gradient[31]. Digestion with restriction endonucleases (Biolabs, Boehringer or BRL) were carried out in standard conditions. Restriction maps were constructed from the size of the DNA fragments, estimated from electrophoretic patterns on agarose or acrylamide gels[32,33]. As indicated, each cDNA insert is bordered by two reconstituted *Pst*I sites[34]. Both inserts have the same orientation with respect to pBR322 map. pBR322 sequences are presented here in their usual orientation, so that the parts of the cDNA sequences corresponding to the 3′ ends of the messengers are on the left-hand side of the inserts. The coding regions are represented by thick lines. The restriction fragments sequenced are represented by arrows. They were labelled at their 5′ (●) or 3′ (◆) ends, and cleaved secondarily to generate subfragments with only one labelled end[35]. Sequencing techniques used were those of Maxam and Gilbert[35] (——) or Maat and Smith[36] (—~—).

```
                    135                     140                     145                     150                     155                     160
         Trp Thr Ala Ala Asp Met Ala Ala Gln Ile Thr Arg Arg Lys Trp Glu Gln Ala Gly Ala Ala Glu Arg Asp Arg Ala Tyr Leu
pH-2d-1  TGG ACG GCG GCG GAC ATG GCG GCG CAG ATC ACC CGA CGC AAG TGG GAG CAG GCT GGT GCT GCA GAG AGA GAC CGG GCC TAC CTA

                        165                     170                     175                     180                     185
         Glu Gly Glu Cys Val Glu Trp Leu Arg Arg Tyr Leu Lys Asn Gly Asn Ala Thr Leu Leu Arg Thr Asp Pro Pro Lys Ala His
pH-2d-1  GAG GGC GAG TGC GTG GAG TGG CTC CGC AGA TAC CTG AAG AAC GGG AAT GCT ACG CTG CTG CGC ACA GAT CCC CCA AAG GCC CAT
                                                                                 TG CGC ACA GAT TCC CCA AAG GCA CAT
                                                                                    Arg Thr Asp Ser Pro Lys Ala His

             190                     195                     200                     205                     210                     215
         Val Thr His His Arg Arg Pro Glu Gly Asp Val Thr Leu Arg Cys Trp Ala Leu Gly Phe Tyr Pro Ala Asp Ile Thr Leu Thr
pH-2d-1  GTG ACC CAT CAC CGC AGA CCT GAA GGT GAT GTC ACC CTG AGG TGC TGG GCC CTG GGC TTC TAC CCT GCT GAC ATC ACC CTG ACC
pH-2d-3  GTG ACC CAT CAC CCC AGA TCT GAA GGT GAA GTC ACC CTG AGG TGT TGG GCC CTG GGC TTC TAC CCT GCT GAC ATC ACC CTG ACC
         Val Thr His His Pro Arg Ser Glu Gly Glu Val Thr Leu Arg Cys Trp Ala Leu Gly Phe Tyr Pro Ala Asp Ile Thr Leu Thr

                     220                     225                     230                     235                     240
         Trp Gln Leu Asn Gly Glu Glu Leu Thr Gln Glu Met Glu Leu Val Glu Thr Arg Pro Ala Gly Asp Gly Thr Phe Gln Lys Trp
pH-2d-1  TGG CAG TTG AAT GGG GAG GAG CTG ACC CAG GAA ATG GAG CTT GTG GAG ACC AGG CCT GCA GGG GAT GGA ACC TTC CAG AAG TGG
pH-2d-3  TGG CAG TTG AAT GGG GAG GAG CTG ACC CAG GAC ATG GAG CTT GTG GAG ACC AGG CCT GCA GGG GAT GGA ACC TTC CAG AAG TGG
         Trp Gln Leu Asn Gly Glu Glu Leu Thr Gln Asp Met Glu Leu Val Glu Thr Arg Pro Ala Gly Asp Gly Thr Phe Gln Lys Trp

         245                 250                     255                     260                     265                     270
         Ala Ser Val Val Val Pro Leu Gly Lys Glu Leu Lys Tyr Thr Cys His Val Glu His Glu Gly     Pro Glu Pro Leu Thr Leu
pH-2d-1  GCA TCT GTG GTG GTG CCT CTT GGG AAG GAG CTG AAG TAC ACA TGC CAT GTG GAA CAT GAG GGG     CCT GAG CCC CTC ACC CTG
pH-2d-3  GCA TCT GTG GTG GTG CCT CTT GGG AAG GAG CAG AAT TAC ACA TGC CGT GTG TAC CAT GAG GGG CTG CCT GAG CCC CTC ACC CTG
         Ala Ser Val Val Val Pro Leu Gly Lys Glu Gln Asn Tyr Thr Cys Arg Val Tyr His Glu Gly Leu Pro Glu Pro Leu Thr Leu

                 275                                 280                     285                     290                     295
         Arg Trp Gly Lys Glu Glu Pro Pro Ser Ser Thr Lys Thr Asn Thr Val Ile Ile Ala Val Pro Val Val Leu Gly Ala Val Val
pH-2d-1  AGA TGG GGC AAG GAG GAG CCT CCT TCA TCC ACC AAG ACT AAC ACA GTA ATC ATT GCT GTT CCG GTT GTC CTC GGA GCT GTG GTC
pH-2d-3  AGA TGG             GAG CCT CCT CCG TTC ACT GAC TCT TAC ATG GTG ATC GTT GCT GTT CTG GGT GTC CTT GGA GCT ATG GCC
         Arg Trp             Glu Pro Pro Pro Phe Thr Asp Ser Tyr Met Val Ile Val Ala Val Leu Gly Val Leu Gly Ala Met Ala
                                                     TM

                 300                     305                     310                     315                     320                     325
         Ile Leu Gly Ala Val Met Ala Phe Val Met Lys Arg Arg Arg Asn Thr Gly Gly Lys Gly Gly Asp Tyr Ala Leu Ala Pro Gly
pH-2d-1  ATC CTT GGA GCT GTG ATG GCT TTT GTG ATG AAG AGG AGG AGA AAC ACA GGT GGA AAA GGA GGG GAC TAT GCT CTG GCT CCA GGC
pH-2d-3  ATC ATT GGA GCT GTG GTG GCT TTT GTG ATG AAG AGA AGG AGA AAC ACA GGT GGA AAA GGA GGG GAC TAT GCT CTG GCT CCA GGC
         Ile Ile Gly Ala Val Val Ala Phe Val Met Lys Arg Arg Arg Asn Thr Gly Gly Lys Gly Gly Asp Tyr Ala Leu Ala Pro Gly
                                     TM

                     330                 335
         Ser Gln Ser Ser Asp Met Ser Leu Pro Asp Cys Lys Val  -COOH
pH-2d-1  TCC CAG AGC TCT GAT ATG TCT CTC CCA GAT TGT AAA GTG TGA
pH-2d-3  TCC CAG AGC TCT GAA ATG TCT CTC CGA GAT TGT AAA GCG TGA
         Ser Gln Ser Ser Glu Met Ser Leu Arg Asp Cys Lys Ala  -COOH
```

Fig. 2 Coding sequences of pH-2^d-1 and pH-2^d-3. Both sequences have been aligned as described in the text, the 5′ terminus is to the left, the 3′ terminus to the right. The putative glycosylation (◆) and phosphorylation (◇) sites are labelled. The positions at which a difference is found between the two nucleotide sequences are labelled (●).

A continuous stretch of 26 uncharged amino acids, mostly hydrophobic, extends from amino acids 282 to 307, displaying at the expected position the characteristic features of a membrane-spanning segment. Noticeably, it contains a repetition (Val-Val-Leu-Gly-Ala-Val, followed by Val-Ile-Leu-Gly-Ala-Val) in pH-2^d-1, also seen at the nucleotide level in pH-2^d-3. The amino acid residues differ in 10 out of the 26 positions, suggesting that the major constraint is the sole maintenance of hydrophobicity. No homology with membrane-spanning segments of other membrane proteins was found.

Amino acids 308–338 correspond to intra-cytoplasmic sequences which have been reported to be phosphorylated[12] and associated with components of the cytoskeleton[13]. A possible phosphorylation site (Arg-Asn-Thr)[14] is found at position 313 in

both H-2 clones. At the border with the membrane, a cluster of four basic residues (Lys-Arg-Arg-Arg) is found in both clones. As clusters of basic amino acids in similar positions have been found in HLA-A2 and HLA-B7 (ref. 15), membrane-bound IgM[16], human glycophorin[17] and several viral glycoproteins[18], we propose that they might be involved in the positioning of transmembrane proteins.

The amino acid sequence located at the external membrane border shows many variations. The conserved proline residues at positions 276–278 indicate breakage of the α-helical structure, suggesting that this segment can form a flexible link, in agreement with the accessibility of this region to papain[6].

pH-2^d-1 codes for the third domain and half of the second, whereas pH-2^d-3 codes for the third domain only. Cysteinyl residues at positions 164, 203 and 269 are likely to be those involved in intrachain disulphide bridges, as they are in H-2K^b (refs 4, 5, 19). Possible glycosylation sites (Asn-Tyr-Thr)[20] are found at positions 176 and 256 in pH-2^d-1 and pH-2^d-3, respectively. The two amino acid sequences are extremely similar. Divergences are found mainly as clusters (positions 193–198, 225–227, 255–268, 275–303) also seen in comparisons with HLA with additional variations.

The third domain of HLA shares significant homologies with the constant domains of immunoglobulin heavy chains[21,22]. Using the alignment frame designed for HLA[21], we found that the third domains of the H-2 molecules encoded by the two plasmids display the same type of homology (in preparation). At 20 out of the 23 aligned positions corresponding to hydrophobic amino acids involved in the β-pleated sheet structure in immunoglobulins, hydrophobic residues are also found in H-2 sequences. These results suggest strongly that the third domain of H-2 antigens, like that of HLA, is folded in an immunoglobulin-like three-dimensional structure. When the aforementioned clusters of amino acid differences between pH-2^d-1 and pH-2^d-3 are placed in the three-dimensional immunoglobulin model, they fall in loop areas (in 8 differences out of 10), while β-pleated sheets correspond to highly conserved regions. This suggests that the three-dimensional structure of the third domain imposes constraints on divergences. This could be true for other parts of the molecule as well and be important in the understanding of the basis of alloantigenicity.

Comparisons with available data on H-2D^d, K^d and L^d (Fig. 3) show that pH-2^d-1 differs from H-2L^d at positions 155, 156, 169 and 262. It has a methionine at position 138, whereas the cyanogen bromide cleavage map of H-2K^d indicates that there is no such residue in the molecule[23]. At 57 out of 58 assigned positions the pH-2^d-1 amino acid sequence is identical to that of H-2D^d, the only difference being at amino acid 255, denoted as 'tentatively assigned'[10]. Therefore, pH-2^d-1 cannot code for H-2L^d or H-2K^d, but could well code for H-2D^d. The pH-2^d-3 sequence differs from H-2D^d at position 262 (ref. 11) and is compatible with the 15 assigned positions reported for H-2L^d in the corresponding area[11]. It has a possible glycosylation site[20] at positions 252–258 as would be expected for H-2K^d (P. Robinson, personal communication). Protein sequence data are thus too limited to allow conclusive assignments, especially as the cloned sequences could also specify Tla, Qa1, Qa2 or other H-2-like antigens[24], in line with the finding that the mouse genome contains multiple H-2-related sequences[25].

Table 1 Analysis of the nucleotide changes between pH-2^d-1 and pH-2^d-3 sequences

	Replacements	Silent substitutions	Total substitutions
Codons 183–284 (third domains)	17/704 = 0.024	5/214 = 0.023	22/918 = 0.024
Codons 285–308 (membrane spanning regions)	7/148.5 = 0.047	2/58.5 = 0.034	9/207 = 0.043
Codons 309–339 (cytoplasmic fragments)	3/221.5 = 0.014	1/66.5 = 0.015	4/288 = 0.014
Codons 183–339	27/1,074 = 0.025	8/339 = 0.024	35/1,413 = 0.025

pH-2^d-1 and pH-2^d-3 were compared over their aligned sequences (Fig. 2). The rate of 'silent substitutions' (see text) was determined by a computation similar to that described by Lomedico *et al.*[28]: all possible single-step mutations (that is, three possible changes for each base) were totalled over the 157 aligned codons, and classified as replacements if they involved an amino acid change, and as silent substitutions if they did not. The numbers were then averaged for the two genes. The fractions displayed in the table indicate the number of replacements (or silent or total substitutions) actually recorded over the total number of possible replacements (or silent or total substitutions).

The nucleotide sequences of pH-2^d-1 and pH-2^d-3 diverge in only 47 (9.7%) of the 485 bases aligned for comparison (including 12 aligned with empty positions, Fig. 2). The third domain shows remarkable conservation with only one base change in a stretch of 156 nucleotides (positions 204–255). Surprisingly, silent changes (with no corresponding amino acid change) are unusually rare, compared with replacement changes (Table 1), whereas in other genes the former arise more often than the latter[26,27]. This raises several hypotheses: a special conservative constraint might exist on the nucleotide sequences themselves, making the silent changes not neutral. Alternatively, the two proteins may have diverged too rapidly to allow the accumulation of neutral mutations in their genes[26]. Whether this feature is related to a mechanism involved in the generation of the natural polymorphism of these genes or not remains to be investigated.

We thank Miss O. Lebail for technical assistance, Mr B. Caudron for assistance in computer usage and Dr D. Meyer for a critical reading of the manuscript. The work described here was done in accordance with the French and German guidelines for

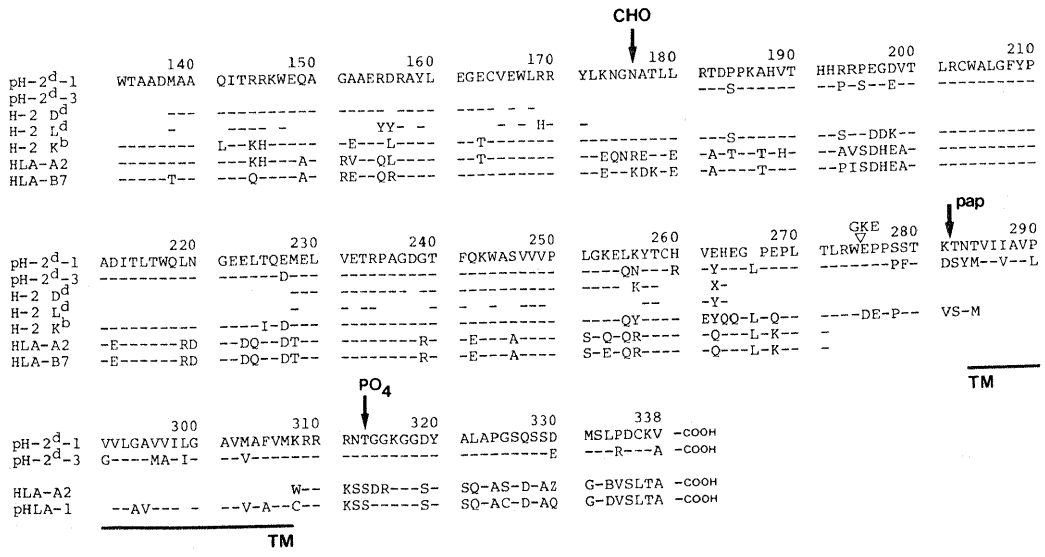

Fig. 3 Comparison of the amino acid sequences encoded by pH-2^d-1 and pH-2^d-3 with sequences of murine and human histocompatibility antigens[3–6,8–11,15,23]. pHLA-1 is a recombinant plasmid carrying an HLA cDNA sequence[37]. Its deduced amino acid sequence shows no difference from the published data on HLA-B7 COOH-terminal fragment[15]. The amino acids are indicated according to the one-letter code[38]. pH-2^d-1 insert has been taken as a reference to align the other sequences. Three residues have been taken out of the alignment between positions 274 and 275, as indicated, to keep to numbering conventions already used for H-2 and HLA sequences[3]. The dashes indicate identity with pH-2^d-1 sequence. X indicates an undetermined amino acid, different from tyrosine.

recombinant research. This work was supported by grants to P.K. from CNRS (ER 201 and ATP 4246), INSERM (SCN 20 and CRAT 72-79-104) and DGRST (79-7-0500 and 78-7-2931), to G.G. from DGRST (79-7-0790) and to P.P. from the Swedish Cancer Society.

Received 12 March; accepted 8 May 1981.

1. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27,** 51–177 (1979).
2. Klein, J. *Science* **203,** 516–521 (1979).
3. Strominger, J. L. *Prog. Immun.* **4,** 541–554 (1980).
4. Uehara, H. *et al. Biochemistry* **19,** 306–315 (1980).
5. Uehara, H. *et al. Biochemistry* **19,** 6182–6188 (1980).
6. Martinko, J. A. *et al. Biochemistry* **19,** 6188–6193 (1980).
7. Kvist, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 2272–2276 (1981).
8. Orr, H. T., Lopez De Castro, J. A., Lancet, D. & Strominger, J. L. *Biochemistry* **18,** 5711–5720 (1979).
9. Orr, H. T., Lopez De Castro, J. A., Parham, P., Ploegh, H. L. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **76,** 4395–4399 (1979).
10. Nairn, R., Nathenson, S. G. & Coligan, J. E. *Eur. J. Immun.* **10,** 495–503 (1980).
11. Coligan, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **77,** 1134–1138 (1980).
12. Pober, J. S., Guild, B. C. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **75,** 6002–6006 (1978).
13. Geiger, B. & Singer, S. J. *Cell* **16,** 213–222 (1979).
14. Kemp, D. E., Graves, D. J., Benjamini, E. & Krebs, E. G. *J. biol. Chem.* **252,** 4888–4894 (1977).
15. Robb, R. J., Terhorst, C. & Strominger, J. L. *J. biol. Chem.* **253,** 5319–5324 (1978).
16. Rogers, J. *et al. Cell* **20,** 303–312 (1980).
17. Tomita, M. & Marchesi, V. T. *Proc. natn. Acad. Sci. U.S.A.* **72,** 2964–2968 (1975).
18. Wickner, W. *Science* **210,** 861–868 (1980).
19. Ewenstein, B. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75,** 2909–2913 (1978).
20. Marshall, R. D. *A. Rev. Biochem.* **41,** 673–709 (1972).
21. Orr, H. T., Lancet, D., Robb, R. J., Lopez De Castro, J. A. & Strominger, J. L. *Nature* **282,** 266–270 (1979).
22. Tragardh, L., Rask, L., Wiman, K., Fohlman, J. & Peterson, P. A. *Proc. natn. Acad. Sci. U.S.A.* **77,** 1129–1133 (1980).
23. Kimball, E. S., Maloy, E. S., Martinko, J. M., Nathenson, S. G. & Coligan, J. E. *Molec. Immun.* **17,** 1283–1291 (1980).
24. Vitetta, E. S. & Capra, J. D. *Adv. Immun.* **26,** 147–193 (1978).
25. Cami, B., Bregegere, F. & Kourilsky, P. *Nature* (in the press).
26. Perler, F. *et al. Cell* **18,** 545–558 (1979).
27. Miyata, T., Yasunaga, T. & Nishida, T. *Proc. natn. Acad. Sci. U.S.A.* **77,** 7328–7332 (1980).
28. Lomedico, P. *et al. Cell* **18,** 545–558 (1979).
29. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72,** 3961–3966 (1975).
30. Bastia, D. *J. molec. Biol.* **124,** 601–639 (1978).
31. Rougeon, F., Kourilsky, P. & Mach, B. *Nucleic Acids Res.* **2,** 2365–2378.
32. Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3,** 2387–2398 (1975).
33. Garoff, H., Frischauf, A. M., Simons, K., Lehrach, H. & Delius, H. *Nature* **288,** 236–241 (1980).
34. Villa-Komaroff, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **75,** 3721–3727 (1978).
35. Maxam, J. & Gilbert, M. *Meth. Enzym.* **65,** 499–560 (1980).
36. Maat, J. & Smith, A. J. H. *Nucleic Acids Res.* **5,** 4537–4545 (1978).
37. Ploegh, H. L., Orr, H. T. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **77,** 6081–6085 (1980).
38. IUPAC-IUB Commission on Biochemical Nomenclature *J. biol. Chem.* **243,** 3557–3559 (1968).

Neutron diffraction reveals oxygen–histidine hydrogen bond in oxymyoglobin

Simon E. V. Phillips*† & Benno P. Schoenborn†

* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
† Department of Biology, Brookhaven National Laboratory, Upton, New York 11973, USA

Myoglobin (Mb) reversibly binds molecular oxygen in vertebrate muscle and consists of a polypeptide chain of 153 residues and one haem, which closely resembles one subunit of a haemoglobin (Hb) tetramer. In oxygenated myglobin (oxyMb) the iron atom is coordinated by four porphyrin nitrogen atoms, N^{ε} of the invariant 'proximal' histidine (F8), and molecular oxygen[1]. The oxygen molecule lies in a tight pocket, bounded by two hydrophobic groups (Phe CD1 Val E11) and the side chain of the 'distal' histidine (E7). This histidine is present in Hb and Mb of many different organisms, with substitution by glutamine or leucine found in only a few cases. The function of the residue is not clear, although it does present steric hindrance to linear ligands such as carbon monoxide and favours 'bent' ones, such as O_2. We report here that the imidazole stabilizes bound molecular oxygen with a hydrogen bond, as revealed by neutron diffraction analysis.

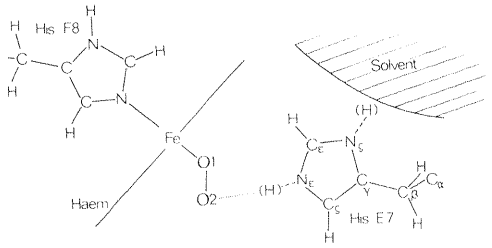

Fig. 1 Arrangement of proximal (F8) and distal (E7) histidines in oxyMb. At *p*H 8.4, nitrogen-bound hydrogen on E7 imidazole can be bonded to either N^{ε} (the naturally predominant form), with a hydrogen bond to O-2 (dotted line), or to N^{δ}, where it projects into the solvent surrounding the molecule.

Pauling[2] first proposed that the imidazole may form a hydrogen bond to the terminal oxygen atom (O-2 in Fig. 1), which carries a formal negative charge in his view of the electronic structure of the FeO_2 complex. Evidence suggesting such a bond comes from electron paramagnetic resonance and oxygen affinity data on cobalt-substituted Hb and Mb[3–5]. The *p*K of the distal histidine is ~5.5 (ref. 4). At physiological *p*H (and the *p*D of the crystals used here) the histidine may have hydrogen-bonded to either N^{ε} or N^{δ} (see Fig. 1), and interaction with O-2 may therefore be by a hydrogen bond, or a simple van der Waals contact. X-ray crystallography of protein crystals cannot distinguish between these alternatives as hydrogen atoms scatter X rays only weakly, and are not normally visible in electron density maps. Neutrons, however, are scattered as strongly by hydrogen and deuterium as C, N, O, S and Fe atoms, and well-ordered H and D atoms may be observed in neutron density maps of proteins[6,7].

Crystals of oxyMb were prepared from frozen sperm-whale skeletal muscle[1]. Large crystals (8 mm^3) were transferred to deuterated mother liquor (*p*D 8.4) at 20 °C 3 months before data collection, because hydrogen gives strong incoherent scattering of neutrons which increases the background level in diffraction data collection. Replacement of H_2O solvent in crystals with D_2O, and subsequent replacement of exchangeable H atoms with D in the protein, alleviates this problem and improves the signal-to-noise ratio of the data. It also allows exchangeable H atoms to be identified in the density map, as H scatters out of phase with respect to D.

Neutron diffraction data were collected using the protein crystallography station of the High Flux Beam Reactor at Brookhaven National Laboratory. The diffractometer was equipped with a two-dimensional multiwire proportional counter[8] and a cooling device to maintain the crystals at −5 °C and retard oxidation of the haem iron. Two crystals were used for data collection, each being exposed to the neutron beam for 21 days. No radiation damage or oxidation was observed; 88% of available data to 2 Å resolution was collected, together with further data between 2 and 1.5 Å, giving 14,411 independent reflections. The merging *R* factor between crystals was 14.1% on intensities.

Calculated phases and amplitudes for the neutron data were computed from the coordinates of all C, N, O, S and Fe atoms in the refined X-ray structure[1], including 60 ordered H_2O molecules. A difference density map (coefficients $|F_o|-|F_c|$: crystallographic *R* factor 35% for 10,152 reflections with $I > 1.5\sigma(I)$) showed clear peaks for 40% of the missing H and D atoms. Small peaks were visible at both N^{ε} and N^{δ} of His E7. H and D atoms observed in the map were added to the model, together with unobserved ones whose positions were known from stereochemistry (for example, most C—H groups), but ring nitrogen-bound H or D atoms for histidines were omitted—this reduced *R* to 33%. A second difference map failed to resolve the ambiguity at His E7, and combined crystallographic and conformational energy refinement was initiated, using methods described for X-ray refinement of oxyMb, but modified for use with neutron data. Seven cycles of coordinate refinement were carried out, with three cycles on individual atomic thermal

0028–0836/81/270081—02$01.00