Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5618–5622, December 1977
Evolution

Molecular evolution of human and rabbit β-globin mRNAs

(mRNA sequence/codon preference/silent substitution)

FOTIS C. KAFATOS*†, ARGIRIS EFSTRATIADIS*, BERNARD G. FORGET‡, AND SHERMAN M. WEISSMAN§

* Cellular and Developmental Biology, Harvard University Biological Laboratories, Cambridge, Massachusetts 02138; † Department of Biology, University of Athens, Athens, Greece; ‡ Hematology Section, Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06510; and § Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510

Communicated by Helen M. Ranney, September 26, 1977

ABSTRACT **The primary structures of human and rabbit β-globin mRNAs are compared. Using as a standard the extent of nucleotide substitutions inferred from the hypervariable amino acid residues of fibrinopeptides A and B, which are thought to change largely by neutral evolution, we show that not all silent mutations in globin mRNA are neutral. The divergence of the sequences is limited in part by the selective usage of synonymous codons. The divergent nucleotides tend to be distributed nonrandomly: in the coding region silent substitutions are most rare in segments that are also deficient in substitutions leading to replacements.**

Until recently, the molecular evolution of nucleic acids could not be studied with the directness possible for proteins (1), because the information was based largely on reassociation and hybridization data, or on inferences from protein sequences. However, simplified DNA (2, 3) and RNA sequencing techniques are now available, and should provide a rapid increase in our direct knowledge of the primary structure of genes. For example, the entire nucleotide sequences of phage R17 and φX174 genomes have been reported (4, 5), and we have derived the complete sequence of rabbit β-globin mRNA by sequencing cloned DNA (6). Here we compare the sequence of rabbit and human β-globin mRNAs (7, 8) from the point of view of molecular evolution and mRNA function.

For clarity and convenience, we first define our use of some important terms.

Substitution. The presence of different nucleotides in equivalent positions of homologous DNA or RNA sequences. Mutation is the initial change in an individual organism, while substitution is a change in a population (9).

Replacement. The presence of different amino acids in equivalent positions of homologous proteins (9).

Silent substitution. A nucleotide substitution that does not lead to an amino acid replacement, either because of the degeneracy of the genetic code, or because the nucleotide is found in a noncoding region.

Silent substitution site. A site at which some but not necessarily all of the potential substitutions are silent in the absence of additional changes. Silent substitution sites in the coding region are the third-base positions in the codons of all amino acids except Met and Trp, and the first-base positions of some Leu and Arg codons (e.g., of CUG, because UUG is a synonym, but not of CUC, because UUC, AUC, and GUC are not synonyms).

Replacement site. A site in the coding region, at which no silent substitution is possible in the absence of additional changes in the same codon. All positions in the mRNA are considered as either replacement or silent substitution sites.

Neutral mutation. A mutation that does not result in a significant selective advantage or disadvantage. This implies that all functions of both the mRNA and the corresponding protein are essentially unaffected by the mutation.

Percent substitution. A relative index of evolutionary rate at the nucleic acid level. For any pair of known sequences or portions thereof, this value can be calculated by dividing the number of sites that show substitutions by the total number of sites and multiplying by 100. For determining absolute rates (e.g., ref. 10), percent substitution values should be corrected for multiple substitutions (11, 12) and divided by the time elapsed since divergence of the two species. However, if two pairs of sequences (e.g., β-globin mRNAs and fibrinopeptide mRNAs; see below) are being compared in the same two species, or in species that have diverged at the same time, the percent substitution values can be treated directly as nonlinear substitution rates. This is very useful, because paleontological estimates of the time of divergence are often subject to considerable uncertainty (13, 14).

Replacement-corrected percent substitution. By definition, in some silent substitution sites some substitutions may lead to amino acid replacement (e.g., U → G in CAU). To correct for possible substitutions leading to replacements, the number of silent sites can be multiplied by a "replacement correction" factor, which is the sum of possible substitutions that would be silent, divided by the sum of all possible substitutions (i.e., 3 times the number of sites).

A neutrality standard for mRNA sequences

The "neutral mutation-random drift" theory (15–17) implies that some features of macromolecules are relatively unimportant, so that a significant number of neutral mutations are possible and can be fixed in the population by random drift. Neutralists agree that many mutants are deleterious and therefore nonobserved. However, sometimes they consider all or nearly all possible silent mutations as neutral (18, 12), which would imply that no appreciable selection is operating on the mRNA itself (e.g., on processing, secondary structure, stability, translatability, etc.). Comparison of the β-globin sequences can test this view, although of course not the general theory.

According to the neutralists, the rate of substitutions of neutral alleles is equal to the mutation rate (15–17). Assuming that the average rate of mutation per nucleotide is equal for different genes, a test of whether an RNA sequence evolves by neutral drift is whether its substitution rate equals that of another sequence thought to be under no selective pressure (*neutrality standard*). By contrast, it is commonly accepted that natural selection acts conservatively at the molecular level, i.e., that functionally more important molecules or parts of a molecule are constrained and evolve more slowly than relatively unimportant ones (e.g., see ref. 17). A recent detailed analysis of the rates of hemoglobin evolution (19) has suggested that replacement rates varied during the evolution of this protein, being high at the outset, when positive selection improved the

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

new protein, and decreasing more recently, when stabilizing selection acted to preserve the important features of the previously improved structure. Thus if a part of the globin mRNA sequence evolves less rapidly than the neutrality standard, we consider it functionally important and infer that a fraction of the possible mutations, being deleterious, have not become established as substitutions because they have been rejected by natural selection.

The choice of a neutrality standard is not easy. The fibrinopeptides A and B evolve very rapidly (1), and despite evidence that fibrinopeptide B is physiologically active *in vitro* (20), they are commonly considered prototypes of neutral evolution. However, some of their amino acid residues are relatively invariant (21), presumably because they are constrained by selection. Moreover, even their most variable residues are probably under some selective pressure. They evolve at a rate similar to that of the hypervariable segment of ribonuclease, for which convincing evidence of selection exists in the form of highly nonrandom usage of amino acids that are equally probable according to the genetic code (22–24). It would seem that the most rapidly evolving fibrinopeptide residues give a lower limit for the percent substitution of a true neutrality standard. The total fibrinopeptide residues would substantially underestimate that standard.

The rapidly evolving residues, positions 4–11 and 15–20 in fibrinopeptides A and B, respectively (alignments 20 and 21 of Dayhoff, ref. 1), are too few for calculation of a meaningful standard based on human and rabbit alone. However, the time separating all the major eutherian orders is comparable, though uncertain in magnitude (13, 14, 25). Thus, we have calculated the percent substitution values for replacement sites of the hypervariable regions in horse, ox, sheep, pig, dog, and cat, in 12 pair-wise comparisons relative to human and rabbit. After conversion of amino acid to nucleotide sequences, and excluding deletions and ambiguities, we have deduced that 167 sites have undergone substitutions, while 138 sites have remained unchanged. Thus, 55% substitution is a minimum neutrality standard, and we consider lower values indicative of selection. The total fibrinopeptide sites in the above animals (616, excluding ambiguities and deletions) show significantly lower percent substitution, 37% ($P < 0.01$); in human and rabbit alone (43 total sites), the same value, 37%, is observed.

General comparison of the globin mRNAs

Fig. 1 shows the rabbit β-globin mRNA sequence, together with all differences found in the human sequence; Table 1 summarizes the substitutions. Nucleotides are numbered beginning after the initiator AUG (6). Besides the 3′-terminal poly(A), each mRNA can be divided into three regions: 5′ noncoding, coding, and 3′ noncoding. It is clear that the differences are limited and not randomly distributed. Excluding three deletions in the 5′ noncoding region and a highly divergent part of the 3′ noncoding region (star in Fig. 1), a total of 66 substitutions in 553 sites is observed (11.9%). The substitutions are nearly equal in frequency in coding and noncoding regions (11.1% and 14.9%, respectively). By contrast, in the RNA phages the substitution rate in the noncoding region is 6.5-fold lower than in the coding region (26).

5′ noncoding region

Comparison of this region in rabbit and human (6, 8, 27) reveals limited divergence. Of the 56 nucleotides in the rabbit, three correspond to deletions and seven to substitutions in the human. Thus, the total divergence is 18% and the percent substitution is 13%—both significantly below the neutrality standard. We conclude that this region is under considerable selective pressure. Its conservativeness may be related to the functional requirements for specific secondary structure (6), possibly related to ribosome binding.

Table 1. Occurrence of substitutions at selected sites and segments of β-globin mRNAs*

Segment	Silent substitution sites			Replacement sites	
	Number	%		Number	%
		Raw	Corrected		
Replacement-free†	14/103	14	22	0/185	0
Replacement-rich†	16/51	31	46	16/91	18
Total coding region	32/156	20.5	30	16/276	5.8
Sum			48/432 (11.1%)		
Total mRNA‡	50/277	18.1	22	16/276	5.8
Sum			66/553 (11.9%)		

* Shown as number of substitutions over sites considered, and as percent substitution, both raw and replacement-corrected. Nucleotides 63 and 219 (see Fig. 1) were considered only in the raw values for total coding region and total mRNA, where they bias the results against our hypothesis.

† For definitions of these segments, see text. The replacement-free segments span nucleotides 1–3, 21–56, 70–141, 174–199, 234–253, 267–327, 350–367, and 381–438. The incidence of silent substitutions in these segments was significantly lower than in the remaining replacement-rich segments according to χ^2 evaluation of binomial proportions in a contingency table ($P < 0.05$).

‡ Includes the entire 5′ and 3′ noncoding regions, except for the three 5′ sites that correspond to deletions and the region corresponding to the star in Fig. 1.

3′ noncoding region

The 3′ noncoding regions were also sequenced and compared by Proudfoot (28). They show considerable divergence, much of it due to unequal lengths: including the termination codon, they consist of 95 nucleotides in rabbit and 135 nucleotides in human. The divergent nucleotides are distributed very nonuniformly. The most distal part, encompassing the 65 nucleotides adjacent to the poly(A), shows only 10 somewhat clustered substitutions (15%); thus we conclude that this part is functionally important and is stabilized by selection. By contrast, the proximal part nearest the coding region shows marked divergence. It includes 30 nucleotides in the rabbit (positions 439–468) and 70 in the human. Several alignments are possible (e.g., see ref. 28), depending on the number of insertions and deletions that are allowed (3 minimum). No more than 22 nucleotides can be matched by any reasonable alignment, and the divergence is so extensive that if it proves to be generally applicable, it may help explain the somewhat limited cross-hybridization values for mammalian globin mRNAs (29–31). Proudfoot (28) has pointed out that the difference in the proximal part may have originated as a partial duplication event. No conclusions can be drawn as yet concerning the functional significance of this region.

Coding region: Replacement sites

Because of uncertainties in the human sequence, only 432 of the 438 nucleotides in the coding region can be compared, of which only 48 (11.1%) are different (Table 1). This is fairly close to the difference at the amino acid level (14 replacements in 146 residues or 9.6%).

Of the 48 substitutions, 16 are found in replacement sites (shaded in Fig. 1), and 2 additional ones lead to replacement although they are found in potentially silent substitution sites

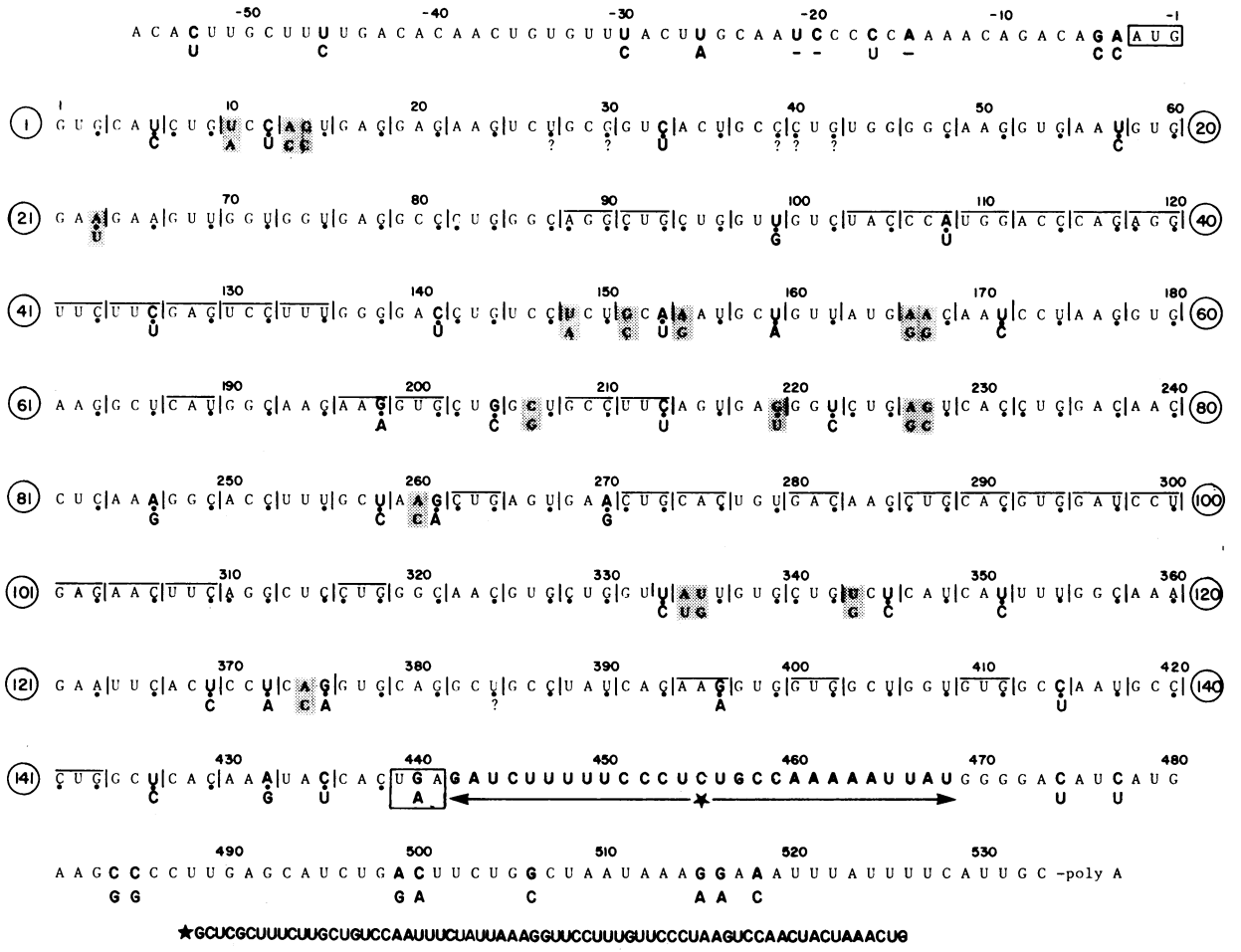

FIG. 1. Differences between the sequences of rabbit (upper line) and human (lower line) β-globin mRNAs. Circled numbers on the left and right indicate the amino acid positions. The nucleotides of the rabbit sequence are numbered from the first position following the initiator AUG (6). The initiation and termination codons are boxed. Triplets of the translated region are separated by vertical lines. Nucleotide substitutions leading to amino acid replacements are shadowed, and all other differences are indicated by bold face. The region immediately following the termination codon is very divergent, and the longer human sequence (★) is shown separately. Unidentified residues are shown by ?s, and deletions by −s. Dots indicate the silent substitution sites in the coding region. Overlining indicates the codons of the functionally most important amino acid residues (19). Data are from refs. 6, 7, and 8.

(shaded and dotted in Fig. 1). The replacement sites have a 5.8% substitution value which, as expected, is considerably lower than that in the fibrinopeptide standard. Clearly, a large proportion of the possible mutants are rejected by selection at the protein level. Whether the replacements that are observed are neutral or have selective value remains to be determined. Not all replacements are associated with the minimum possible number of substitutions (see residues 5 and 76).

Like most structural genes for which data are available (32, 33), the β-globin gene shows polymorphism in both rabbit (34, 35) and human (36). In the rabbit, a common and phenotypically normal variant with 4 amino acid differences exists. The four nucleotides that must be affected (positions 154, 167, 227, and 334) also show substitutions between human and rabbit, and indeed have tolerated extensive divergence during vertebrate evolution (1).[¶] In human populations there is no evidence for common and phenotypically normal variants, but polymorphism is manifested by the many abnormal variants (36). These variants may be used to infer independently the nucleotides present in some of the silent substitution sites of normal human β-globin mRNA, although such an analysis leads to contradictory predictions for one site (codon 67) and a mistaken prediction for a second (codon 50; ref. 7).

Coding regions: Silent substitution sites

Of the 156 silent substitution sites, 32 or 20.5% show substitutions, 30 of them silent. Thus, these sites are significantly ($P < 0.01$) more conservative than the fibrinopeptide standard. Even when replacement-corrected so as to consider exclusively possible substitutions that are silent (297 of a total of 462), the percent substitution (30.3%) differs significantly from the fibrinopeptide standard ($P < 0.01$). We conclude that silent mutations in the coding as in the noncoding regions are not necessarily neutral, and that some stabilizing selection operates on the mRNA itself. This selection is not as intense as that which operates on the protein, because the substitution value is much lower at the replacement sites (5.8%). A somewhat different analysis of partial sequence data (10, 38) had previously tended to support the notion that silent mutations are neutral.

The suggestion that silent mutations are not necessarily neutral is strongly supported by the nonrandom use of synon-

[¶] Browne *et al.* (37) have reported a partial sequence of cloned β-globin DNA, apparently derived from the alternative allele of rabbit β-globin. Their sequence (nucleotide positions 277 to 445) indicates that the allelic replacement at amino acid position 112 is due to a single nucleotide substitution. The alleles also seem to differ by a silent substitution at position 421. Approximately three-quarters of that region (nucleotide positions 319 to 445) has also been sequenced by Proudfoot (28); his sequence is identical to the one we have reported (6).

Table 2. Codon usage in β-globin mRNA

1 \ 2	U			C			A			G			2 / 3
U	Phe	UUU	3/5 8	Ser	UCU	3/0 3	Tyr	UAU	1/2 3	(Cys)	UGU	1/2 3	U
		UUC	5/3 8		UCC	3/2 5		UAC	2/1 3		UGC	0	C
	Leu	UUA	0		UCA	0	(Term)	UAA	0/1 1	(Term)	UGA	1/0 1	A
		UUG	0		UCG	0		UAG	0	(Trp)	UGG	2/2 4	G
C	Leu	CUU	0	Pro	CCU	3/5 8	His	CAU	4/2 6	Arg	CGU	0	U
		CUC	2/3 5		CCC	0		CAC	5/7 12		CGC	0	C
		CUA	0		CCA	1/2 3	Gln	CAA	0		CGA	0	A
		CUG	16/14 30		CCG	0		CAG	4/3 7		CGG	0	G
A	(Ile)	AUU	1/0 1	Thr	ACU	2/3 5	Asn	AAU	4/1 5	Ser	AGU	4/2 6	U
		AUC	0		ACC	2/3 5		AAC	4/5 9		AGC	0	C
		AUA	0		ACA	0/1 1	Lys	AAA	3/3 6	Arg	AGA	0	A
	(Met)	AUG	1/1 2		ACG	0		AAG	9/8 17		AGG	3/3 6	G
G	Val	GUU	4/3 7	Ala	GCU	7/4 11	Asp	GAU	1/5 6	Gly	GGU	4/4 8	U
		GUC	2/2 4		GCC	6/7 13		GAC	3/2 5		GGC	6/8 14	C
		GUA	0		GCA	1/1 2	Glu	GAA	4/2 6		GGA	0	A
		GUG	12/13 25		GCG	1/0 1		GAG	6/6 12		GGG	1/1 2	G

The frequency of use of each codon is shown for rabbit (upper left), human (lower left), and the sum of the two (right); amino acid positions 9, 10, 13, 14, and 128 of the human sequence are excluded. Term, termination codon. Underlined amino acids are those for which codon usage in both species combined is significantly nonrandom according to a χ^2 test ($P < 0.05$, even after Yates's correction for small numbers); codon usage cannot be evaluated statistically for the cases shown in parentheses, and is statistically random for the rest. *Strongly favored codons* (58% of the translated region): CUG, GUG, UCU, UCC, CCU, CCA, ACU, ACC, GCU, GCC, CAG, AAG, AGU, AGG, GGU, GGC. *More or less avoided codons* (10% of the translated region): CUC, GUU, GUC, ACA, GCA, GCG, AAA, GGG. *Indifferent codons* (29% of the translated region): UUU, UUC, UAU, UAC, CAU, CAC, AAU, AAC, GAU, GAC, GAA, GAG. Assignments of the codons to these three categories are based on usage in both rabbit and human mRNA, relative to the usage of synonymous codons.

ymous codons (see below). It is further supported by the nonrandom distribution of observed substitutions: silent and nonsilent substitutions tend to be clustered together, suggesting that the evolutionary constraints operating at the mRNA and protein levels tend to coincide.

The functionally most important and evolutionarily most conservative amino acids in globin are clustered. For example, overlining in Fig. 1 indicates the largely clustered distribution of the codons for the 35 most important residues, i.e. those involved in heme interactions, Bohr effect sites, and $\alpha_1\beta_2$ contact sites (19). Because of the clustering, it is possible to define "replacement-rich" and "replacement-free" segments of the mRNA by arbitrary but consistent rules. As defined in Table 1, replacement-rich segments are all those that include a substitution leading to replacement plus 6 nucleotides upstream and 6 downstream from that substitution; replacement-free are all segments excluded by this definition. The replacement-free segments also prove significantly deficient in *silent* substitutions (by a factor of two, relative to replacement-rich segments; Table 1). This deficiency is also observed if the replacement-rich segments are defined somewhat differently (e.g., as extenting 4 or 8 nucleotides from each substitution leading to replacement). The most dramatically conservative long segment, centered at nucleotide 297, includes 14% of the coding region, 34% of the functionally most important amino acids, and only 1 silent substitution.

What might be the meaning of these correlations? Whenever selection operates, deleterious mutations create a "genetic load" for the organism. Suppose that an mRNA segment must be base-paired for optimal mRNA function; its nucleotides are then constrained against substitution. If some of these nucleotides are also constrained by coding for functionally important amino acid residues, the overall genetic load would be reduced relative to that which would have been required if the segments important for mRNA and protein functions were noncoincident. Coincidence, then, might be viewed as "coadaptation" of protein and mRNA functions.

Possible significance of the limited divergence: Codon selection

As previously noted (6, 7), and as shown in Table 2, synonymous codons are used quite nonrandomly, and similarly in both β-globin mRNAs. This implies selection of the nucleotides occupying the silent substitution sites. There are 54 possible codons corresponding to amino acids abundant enough in β-globin to permit statistical evaluation of possible preferential codon

usage. Of these, 16 are strongly favored, 8 are more or less avoided, 18 are not used at all, and 12 are indifferent, i.e., they are not used at a frequency significantly different from that of their synonyms (Table 2). The preferential codon usage in β-globin mRNA is positively correlated with the relative abundance of the respective isoacceptor tRNAs in rabbit reticulocytes (D. Hatfield, R. C. Matthews, and M. Caicuts, unpublished data; see also ref. 6 for a review of the older literature). This helps explain the preponderance of C ↔ U transitions (23 of 66, or 17 of 30 silent substitutions in the coding region); in general these transitions do not change the tRNA species needed for translation, as a result of wobble. Selective codon usage has also been noted in a variety of eukaryotic and prokaryotic mRNAs, including RNA phages, and is specific for different mRNAs (e.g., refs. 4, 6, 39–41).

The adaptive value of codon selection is as yet unknown. The general bias against C-G doublets in eukaryotic DNA (42) is also manifested by the nonuse of all four CGX codons, and of all but one XCG codons (Table 2). It may be speculated that some isoacceptor tRNAs modulate the rate of translation (43–46). An alternative explanation of codon selection is the "adaptation" of the tRNA pattern to the demands of translation (47).

Possible significance of the limited divergence: Secondary structure

Selection at the mRNA level may operate not only through codon preference, but also through requirements for specific three-dimensional features, including base-paired regions. Spatial features may be important for interaction with any one of many possible factors affecting mRNA maturation, processing, stability, translation and degradation. Physical studies reveal that globin mRNAs (48, 49), like RNA phages (4, 50), have considerable secondary structure. According to a plausible secondary structure model of phage MS2 RNA (4), double-stranded regions diverge at a rate 2.5-fold lower than single-stranded regions (26). For proper evaluation of the possible role of base pairing in limiting the evolutionary divergence of β-globin mRNAs, it is necessary to construct and experimentally test thermodynamically optimized secondary structure models (G. Pavlakis, N. Vamvakopoulos, and J. N. Vournakis, unpublished).

We thank F. Baralle for permitting us to use his data on the human 5′ noncoding region; E. Zuckerkandl, R. C. Lewontin, T. Maniatis, M. Goodman, M. McKenna, and D. Lindsley for helpful discussions; W. Salser for sending us manuscripts prior to publication; M. Koehler for the figures; and M. J. Randell for expert secretarial assistance. Work in our laboratories has been supported by grants from the National Science Foundation, National Institutes of Health, and Hellenic National Science Foundation; A.E. was supported by a fellowship of the Harvard Society of Fellows.

1. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, DC), Vol. 5.
2. Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94,** 441–448.
3. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
4. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G. & Ysabaert, M. (1976) *Nature* **260,** 500–507.
5. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. V., Slocombe, D. M. & Smith, M. (1977) *Nature* **265,** 687–695.
6. Efstratiadis, A., Kafatos, F. C. & Maniatis, T. (1977) *Cell* **10,** 571–585.
7. Marotta, C. A., Wilson, J. T., Forget, B. G. & Weissman, S. M. (1977) *J. Biol. Chem.* **252,** 5040–5053.
8. Baralle, F. (1977) *Cell*, in press.
9. Fitch, W. M. (1976) in *Molecular Evolution*, ed. Ayala, F. J. (Sinauer Associates, Sunderland, MA), pp. 160–178.
10. Salser, W., Bowen, S., Browne, D., Adii, F. E. Federoff, N., Fry, K., Heindell, H., Paddock, G., Poon, R., Wallace, B. & Whitecome, P. (1976) *Fed. Proc.* **35,** 23–25.
11. Moore, G. W., Goodman, M., Callahan, C., Holmquist, R. & Moise, H. (1976) *J. Mol. Biol.* **105,** 15–37.
12. Kimura, M. (1977) *Nature* **267,** 275–276.
13. McKenna, M. (1976) in *Phylogeny of the Primates*, eds. Luckett, W. P. & Szalay, F. S. (Plenum Publishing Corp., New York), pp. 21–46.
14. Romero-Herrera, A. E., Lehmann, H., Joysey, K. A. & Friday, A. E. (1973) *Nature* **246,** 389–395.
15. Kimura, M. (1968) *Nature* **217,** 624–626.
16. Kimura, M. & Ohta, T. (1971) *Nature* **229,** 467–469.
17. Kimura, M. & Ohta, T. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 2848–2852.
18. King, J. L. & Jukes, T. H. (1969) *Science* **164,** 788–798.
19. Goodman, M., Moore, G. W. & Matsuda, G. (1975) *Nature* **253,** 603–608.
20. Osbahr, A. J., Gladner, J. A. & Laki, K. (1964) *Biochim. Biophys. Acta* **86,** 535–542.
21. Ohta, T. & Kimura, M. (1971) *Nature* **233,** 118–119.
22. Corbin, K. W. & Uzzell, T. (1970) *Am. Nat.* **104,** 37–53.
23. Barnard, E. A., Cohen, M. S., Gold, M. H. & Kim, J. (1972) *Nature* **240,** 395–398.
24. Zuckerkandl, E. (1975) *J. Mol. Evol.* **7,** 1–57.
25. Dickerson, R. E. (1971) *J. Mol. Evol.* **1,** 26–45.
26. Min Jou, W. & Fiers, W. (1977) *J. Mol. Biol.* **106,** 1047–1060.
27. Baralle, F. E. (1977) *Cell* **10,** 549–558.
28. Proudfoot, N. J. (1977) *Cell* **10,** 559–570.
29. Gummerson, K. S. & Williamson, R. (1974) *Nature* **247,** 265–267.
30. Housman, D., Forget, B. G., Skoultchi, A. & Benz, E. J. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 1809–1813.
31. Kacian, D. L., Gambino, R., Dow, L. W., Grossbard, E., Natta, C., Ramirez, E., Spiegelman, S., Marks, P. A. & Bank, A. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 1886–1890.
32. Lewontin, R. C. (1974) *The Genetic Basis of Evolutionary Change* (Columbia University Press, New York).
33. Selander, R. K. (1976) in *Molecular Evolution*, ed. Ayala, F. J. (Sinauer Associates, Sunderland, MA), pp. 21–45.
34. Gallizi, A. (1971) *Nature New Biol.* **229,** 142–143.
35. Garrick, M. D., Hafner, R., Bricker, J. & Garrick, L. M. (1974) *Ann. N.Y. Acad. Sci.* **241,** 436–438.
36. Stamatoyannopoulos, G. (1972) *Annu. Rev. Genet.* **6,** 47–70.
37. Browne, J. K., Paddock, G. V., Liu, A., Clarke, P., Heindell, H. C. & Salser, W. (1977) *Science* **195,** 389–391.
38. Salser, W. & Isaacson-Strommer, J. (1976) *Prog. Nucleic Acid Res. Mol. Biol.* **19,** 205–220.
39. Berger, E. (1977) *J. Mol. Evol.*, in press.
40. Elton, R. A., Russell, G. J. & Subak-Sharpe, J. H. (1976) *J. Mol. Evol.* **8,** 117–135.
41. Fitch, W. M. (1976) *Science* **194,** 1173–1174.
42. Russell, G. J., Walker, P. M. B., Elton, R. A. & Subak-Sharpe, J. H. (1976) *J. Mol. Biol.* **108,** 1–15.
43. Ames, B. N. & Hartman, P. E. (1963) *Cold Spring Harbor Symp. Quant. Biol.* **28,** 349–356.
44. Itano, H. A. (1965) in *Abnormal Hemoglobins in Africa*, ed. Jonxis, J. H. P. (Blackwell, Oxford), pp. 3–16.
45. Zuckerkandl, E. & Pauling, L. (1965) *J. Theor. Biol.* **8,** 357–366.
46. Min Jou, W., Van Montagu, M. & Fiers, W. (1976) *Biochem. Biophys. Res. Commun.* **73,** 1083–1093.
47. Garel, J. (1974) *J. Theor. Biol.* **43,** 211–225.
48. Holder, J. W. & Lingrel, J. B. (1975) *Biochemistry* **14,** 4209–4215.
49. Favre, A., Morel, C. & Scherrer, K. (1975) *Eur. J. Biochem.* **57,** 147–157.
50. Min Jou, W., Haegeman, G., Ysebaert, M. & Fiers, W. (1972) *Nature* **237,** 82–88.