Volume 16 Number 21 1988 Nucleic Acids Research

A survey on intron and exon lengths

John D.Hawkins

Department of Biochemistry, Medical College of St Bartholomew's Hospital, Charterhouse Square, London, EC1M 6BQ, UK

Received July 15, 1988; Revised and Accepted September 23, 1988

ABSTRACT
The lengths of introns and exons in various parts of genes of vertebrates, insects, plants and fungi are tabulated. Differences between the various groups of organisms are apparent. The results are discussed and support the idea that, generally speaking, introns were present in primitive genomes, though in some cases they may have been inserted into pre-existing genes.

INTRODUCTION
Since their discovery, late in the 1970s (reviewed and discussed by Witkowski (1)) the idea that eukaryotic and some viral protein encoding genes are generally interrupted by non-coding introns has been an intriguing one, and has given rise to speculation and discussion about their origin and existence (e.g. 2–5). Naora & Deacon (6), Blake (4) and Traut (7) collected some information about the lengths of introns and exons, and Senapathy (8) has shown that coding sequences of longer than 600 nt are extremely unlikely to be found among random sequences of nucleotides because of the intervention of stop codons.

I have culled data from the literature on the exon/intron structure of a large number of eukaryotic genes. In making this collection, it has become apparent that there are significant differences in the exon/intron organisation between different phyla, so it seems worthwhile to document them.

METHODS
For this survey, the structure of genes was noted and various elements defined as follows:

separate 5′-non-coding exons, separated by an intron from the exon containing the codon for the site of initiation of translation (5′-ncex):

introns wholly within the 5′-non-coding region (5′-in):

the 5′-untranslated part of the first coding exon, further subdivided according to whether or not there is a preceding untranslated exon (5′-ncexwin) or (5′ncexnin):

the coding portion of the first exon (5′-cex):

internal exons and introns:

the 3′-coding portion of the last exon (3′-cex):

the 3′-non-coding part of the last exon, with and without following non-coding exons (3′-win and 3′-nin):

introns within the 3′-untranslated part of the gene (3′-in):

separate 3′-exons with no coding information (3′-ncex).

Not all these elements, particularly the first and last and their associated introns, are always present. In many cases the extents of the 5′- and 3′-non-coding elements of the first and last coding exons have not been accurately established.

Since there seem to be differences in the size and frequency of occurrence of some of these elements in different groups of organisms, they have been separated into four groups—vertebrates, insects (mostly Drosophila species), higher plants and fungi. Genes for only a few other invertebrates and protista have been sequenced but they do not form large enough groups to show whether they fall into consistently different patterns from the groups analysed.

Some selection of which genes to include has had to be made to avoid bias. Genes specifying the same protein in different species frequently have the same numbers and sizes of exons (e.g. the α- and β-globin genes). In such cases only one example has been included, but where the sizes of the introns in such genes differ appreciably they have all been included. In the case of gene families, such as the cytochrome-P450 family, I have excluded several members whose gene structures are almost identical, though I have included some where the structure is at least partially different. Such selection is inevitably arbitrary but necessary if a useful picture is to be obtained. Collagen genes possess many exons of the same size—presumably as a result of multiplication of a basic unit—and only one exon of any one size has been included.

A number of genes show considerable heterogeneity in the point at which transcription is commenced. In these cases I have tried to choose the length of the most abundant transcript or, where this is not clear-cut, I have arbitrarily taken the longest transcript.

RESULTS AND DISCUSSION

A summary of the lengths of introns and different kinds of exons is presented in Tables I and II.

Introns (Figs. 1 & 2)

There are marked differences in the distribution of intron size among the various groups of organisms.

Fungi have the shortest introns. In this group I have not included introns in ribosomal

Table I. Average length (nt) and number of introns

	5′-introns	internal introns	3′-introns
Vertebrates	1811 (91)	1127 (1941)	681 (25)
Insecta	5507 (19)	622 (210)	
Fungi		86 (126)	
Plants		249 (200)	

The number of introns examined in each category is shown in brackets.

Table II. Average lengths (nt) and number of exons

	5′-ncex	5′-ncexwin	5′-ncexnin	5′-cex	intex	3′-cex	3′-nin	3′-win	3′-ncex
Vertebrates	94(111)	26(89)	77(193)	134(302)	137(1305)	198(291)	434(291)	39(23)	409(24)
Insecta	312 (18)	68(14)	138 (49)	223 (66)	392 (149)	505 (71)	317 (61)		
Fungi			71 (27)	133 (54)	260 (77)	418 (52)	175 (19)		
Plants			60 (45)	209 (45)	183 (149)	302 (46)	90 (33)		

For abbreviations, see Methods.
The number of exons examined in each category is shown in brackets.

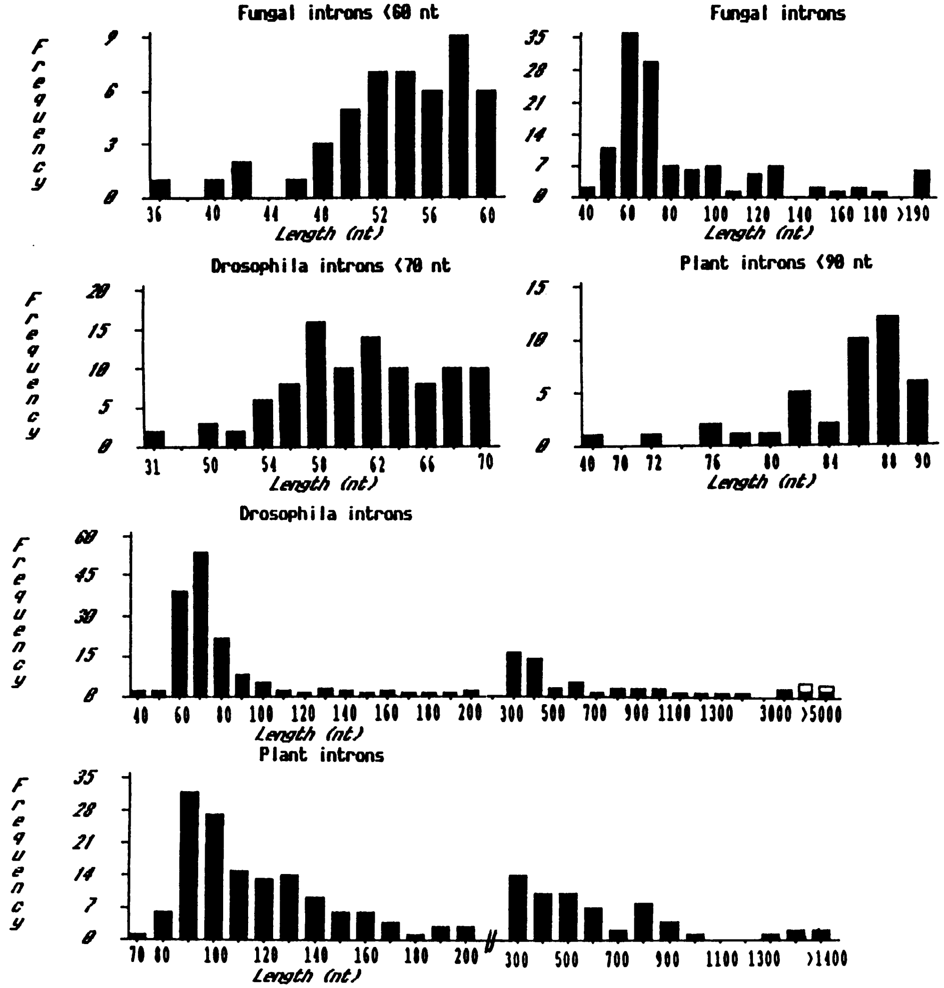

Fig. 1. Lengths of internal introns. The figures on the abscissae are the upper lengths of the bins.

protein genes, since they seem to show a fairly constant pattern which is different from the other fungal genes. Apart from these, the mean length of fungal introns is 85 nt with well over half containing less than 100 nt and none more than 520 nt.

The ribosomal protein genes of *Saccharomyces cerevisiae* that have so far been sequenced display interesting features (Table III). They nearly all have an intron, which is unusual for genes in this species, and these are all relatively long (230–513 nt). There are also two other ribosomal protein genes (in *Candida* and *Dictyostelium*) that have single introns of 356 and 389 nt. In all these genes the intron is near the 5′-end of the coding sequence. Several of the other *Saccharomyces* genes that do possess introns have relatively long ones (actin–309 nt; ubiquitins–367 & 434 nt; CRY-1–305 nt; tubulin–298 nt), but this is not a universal feature of this species.

Insects also have many short introns–again over half are shorter than 100 nt: in fact 80% are between 50 and 75 nt long. There are also a few longer ones of at least 2000 nt.

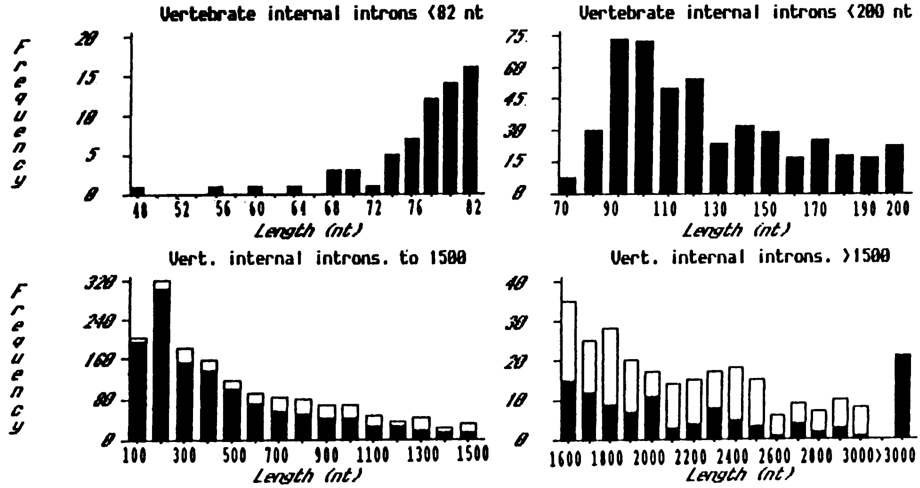

Fig. 2. Lengths of internal introns of vertebrates. The filled bars are of introns whose length is accurately known. The open bars are of introns whose length is only known approximately. The figures on the abscissae are the upper lengths of the bins. There are also 119 introns over 3000 nt long whose length is only known approximately.

As in the case of vertebrates (see below), a number of the longer ones have not been sequenced so their lengths are only known approximately. Including these, the average length is 622 nt.

Higher plants have fewer very short introns, with only a third of 100 nt or less, and none longer than 2000 nt have so far been described. Their mean length is 249 nt.

In contrast, vertebrates have introns with a wide range of lengths, although shorter ones predominate. The largest single size range is 80–99 nt–in this respect like higher plants. The precise lengths of many of the larger vertebrate introns have not been determined accurately, but are only known from measurements of either electron microscope images of mRNA-DNA duplexes or restriction endonuclease fragments separated by electrophoresis. In spite of the uncertainties arising, these estimated lengths must be included since they comprise a very considerable proportion of the longer introns. Taking them into account, 19% (out of 1941) are longer than 1600 nt, and the mean length is 1127 nt.

The shortest recorded intron is 31 nt long (found in the Drosophila genes white and for the Na^+ channel), so this may be near the minimum length required to include suitable sequences to mark the 5′- and 3′- ends, a site for lariat formation in the splicing reaction, and enough flexibility to take up a suitable conformation for this process. Drosophila and fungal genes contain a majority of introns between 50 and 75 nt long, suggesting that they have got down to very near the minimum length required for proper splicing. Bingham *et al.* (9) have recently proposed that, in some cases, longer introns in Drosophila may contain sequences that can control their own splicing so as to determine whether or not a particular protein is produced under any given physiological condition. Introns may also contain other elements, such as enhancers (10) which require them to be longer than the minimum length needed for correct splicing.

Introns outside the coding region (Fig. 3)

The vertebrate introns separating exons preceding the coding ones show a tendency to be fairly long. Only 12 out of 91 are less than 200 nt long, and their average length is 1811 nt. Similarly the 5′-introns in insects are longer than the internal ones, though

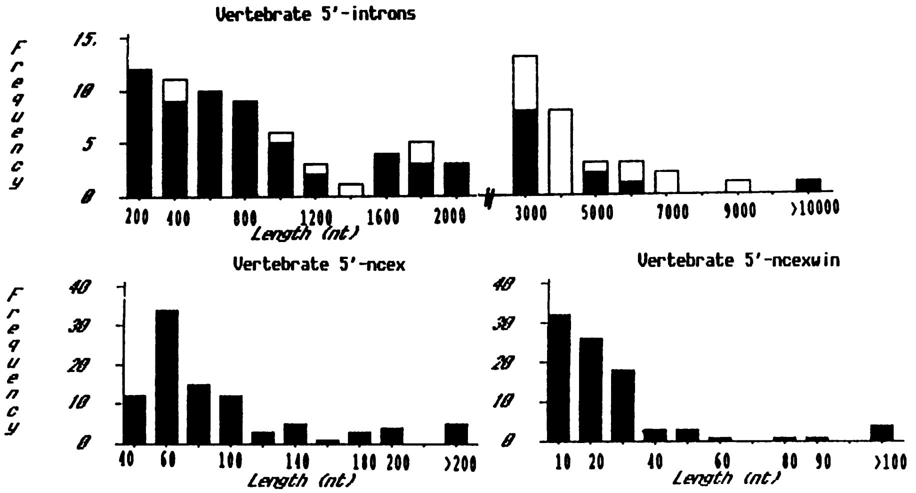

Fig. 3. Lengths of vertebrate introns and exons preceding the coding exons. The filled bars are of introns whose length is accurately known. The open bars are of introns whose length is only known approximately. The figures on the abscissae are the upper lengths of the bins. The lengths of the exons in the last bins are: Introns — 14 257; 5'-ncex — 273, 292, 706, 1121, 1317; 5'-ncexwin — 115, 149, 280, 314. For abbreviations, see Methods.

not to such a great extent, except in the case of the *antennapedia* gene where there are three enormous introns of approximately 25 000, 30 000 and 35 000 nt.

The small number of introns in the 3'-non-coding region of vertebrate genes are of fairly typical length.

Internal exons (Figs. 4 & 5)
The sizes of internal exons vary between the different groups. For vertebrates there is a fairly broad peak in the size distribution between 100 and 170 nt, with a mean length of 137 nt. Only 7 out of 1305 are over 550 nt long—the possible length of a primordial gene, according to Naora *et al.* (5). Quail troponin I contains the shortest exon so far recorded with only 7 nt. The rat troponin T gene is unusual in containing 7 exons shorter than 20 nt out of a total of 15 exons.

Exons of higher plants show a similar size distribution though the peak is less sharp. The mean length is 183 nt, perhaps because there is a slightly higher proportion of longer exons: 2.7% are over 550 nt long. Only 4 are shorter than 50 nt, and three of these (37, 39, 48 nt) are in the Alfalfa gene for glutamine synthetase.

In the fungi the most abundant exons contain less than 100 nt, but there is a large spread and the mean length is 260 nt. 7.8% are more than 550 nt long. This figure does not include the many *Saccharomyces* genes that contain no introns. Short exons are found in some tubulin genes—in *Aspergillus* the β-tubulin genes have exons of 24, 25, 26 and 27 nt, while a *Candida* tubulin gene has one of 36 nt. The *Neurospora* ribosomal protein L-29 gene has a structure completely different from the genes listed in Table III, containing exons of 39, 17, 30, 25, 28 nt.

The length of majority of exons in the insects is in the range of 100—180 nt, but there is an appreciable number of longer ones, with 15% being more than 550 nt long. The mean length in this group is 392 nt. Small exons are rather rare, but Drosophila has exons of 28, 32 and 34 nt in the myosin L-chain, protein kinase C and Rh. opsin genes respectively.

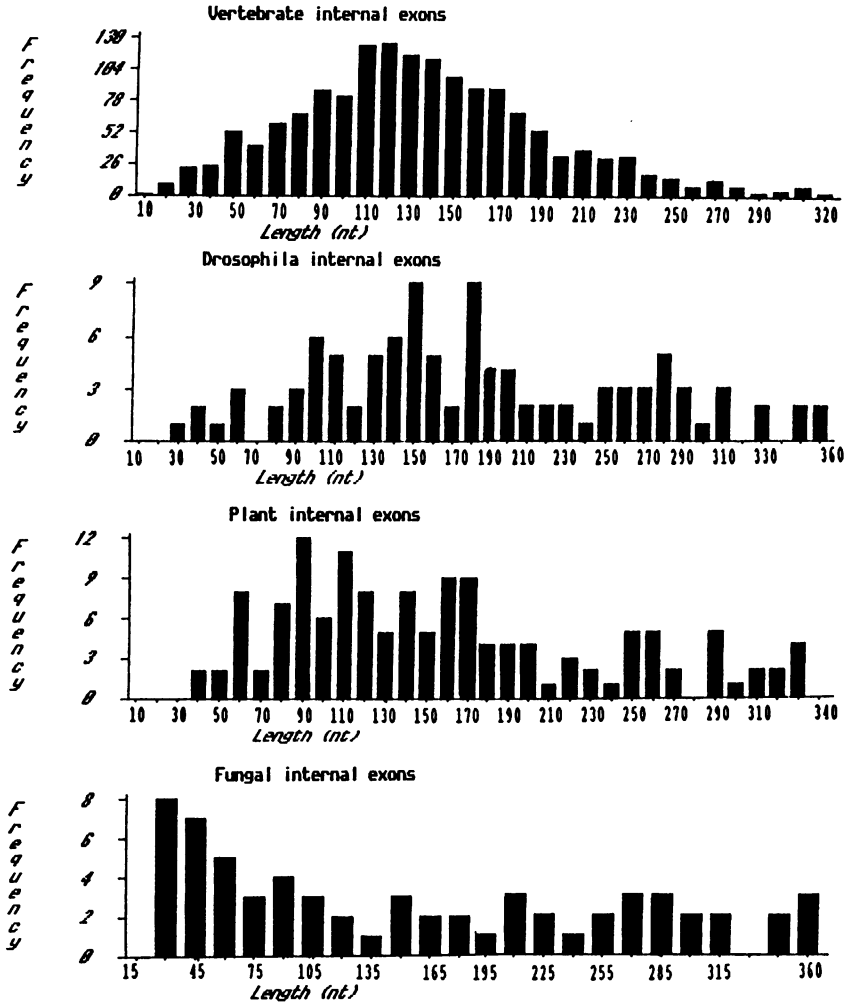

Fig. 4. Lengths of internal exons. The figures on the abscissae are the upper lengths of the bins. Only the shortest exons are shown.

Naora & Deacon (6) have suggested that exons may be grouped into 3 major and 2 minor discrete groups according to length. The analysis of this larger data base (using bins of 10 as against 25 used by Naora & Deacon) suggests that in vertebrates the major peak in the distribution is at 100–120 nt, but there may be shoulders at around 50, 170 and 200–230 nt. In insects there appear to be peaks at 90–110, 140–150 and around 280 nt. In higher plants there are peaks at 80–110, 150–170 and possibly between 240–290 nt. The lengths of fungal exons have a completely different distribution with the greatest number being short–between 17 and 45 nt. However the numbers in these last three groups are rather small so it is premature to draw any definite conclusions.

Intronless genes

It is well known that the majority of genes of *Saccharomyces* have no introns, even though

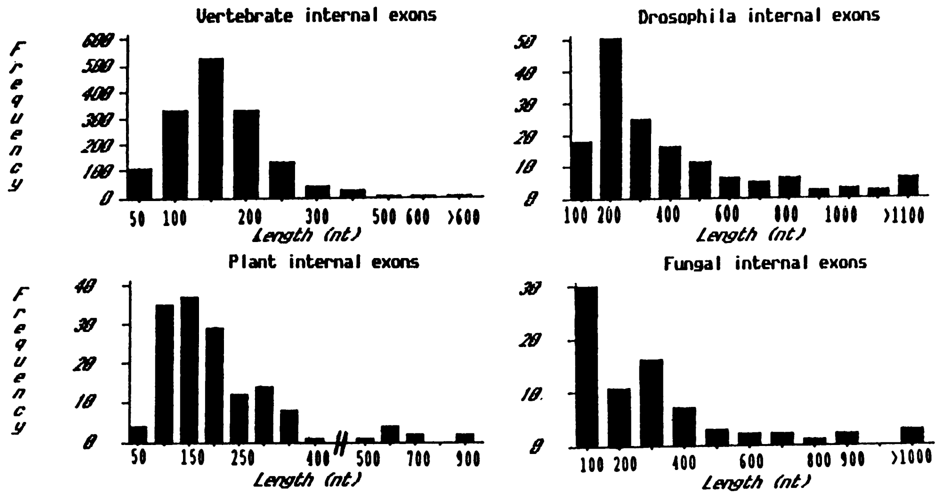

Fig. 5. Lengths of internal exons. The figures on the abscissae are the upper lengths of the bins. The lengths of the exons in the last vertebrate bin are: 690, 961, 1101, 1375, 7572. The long exons in the other graphs are listed in Table IV.

Table III. Fungal ribosomal protein genes

Ribosomal protein	Exons	Introns
S10	5′+6/705+3′	352, 394*
S16A	5′+20/415+3′	390, 551*
L3	5′+1164+3′	none
L16	35+525+90	none
L25	24+13/401+80	415
L29	35+48/402+87	510
L32	58+3/315+100	230
L34	5′+57/285+3′	349, 421*
L46	5′+6/450+3′	383
rp28	5′+112/449+3′	429, 427*
rp29	38,7+468+131	458 (5′) †
46	35+525+90	none
rp51	5′+3/408+3′	325, 398*
59	5′+7/407+3′	307
L25	25+13/416+3′	389
1024	15+12/546+35	350

*There are two copies of these genes in the genome.
†This intron interrupts the 5′-non-coding portion of the gene. The figures in the exons column are (in order): 5′-non-coding nt; 5′-coding nt; 3′-coding nt; 3′-non-coding nt.
All are genes of *Saccharomyces cerevisiae*, except the last two which are *Candida* and *Dictyostelium* respectively.

their length frequently exceeds 1000 nt. There are some cases of intronless genes in other fungi. In the series presented here 18% of recorded fungal genes have no introns and their coding lengths vary between 312 and 6351 nt. These include the exceptionally long gene of *Dictyostelium* mysoin H chain which has 6351 nt, as well as a miscellaneous collection of other genes.

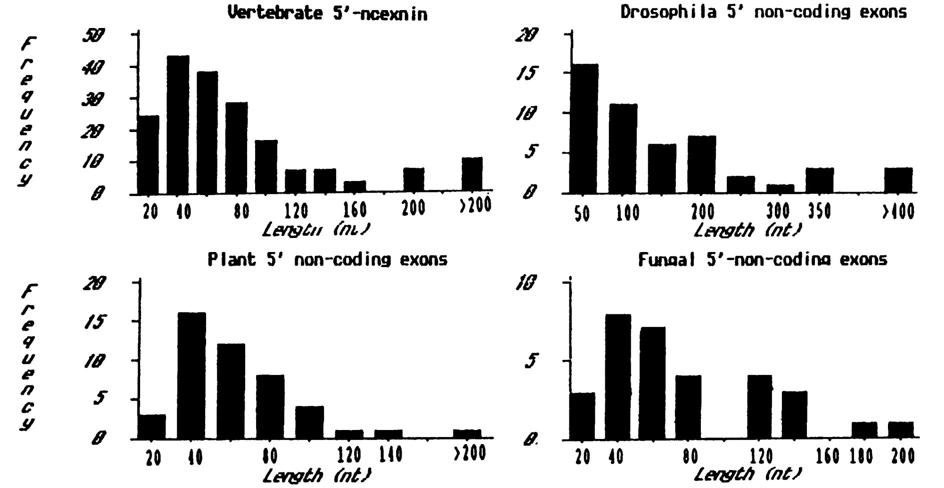

Fig. 6. Lengths of 5'-untranslated parts of 5'-exons. The figures on the abscissae are the upper lengths of the bins. The lengths of the exons in the last bins are: Vertebrates — 212, 220, 238, 333, 337, 526, 684, 843, 1022; Drosophila — 460, 489, 798; Plants — 422 nt.

Intronless genes occur with a similar frequency in higher plants (17%) and in insects (19%). Prominent among these are 3 globin genes of *Chiromonas* and 5 heat shock protein genes of *Drosophila*.

Such genes seem to be much rarer in vertebrates where only 13 out of 328 genes display this trait. One is a chicken heat shock protein (which may be homologous to a similar protein in *Drosophila*), and four, with appreciable homology to each other are the genes for the α_2-, β_1-, and β_2-adrenergic receptors and the M1-muscarinic receptor. The coding portions of these genes are all approximately 1300 nt long.

Among the few protistan genes that have been sequenced two thirds (14 out of 21) contain no introns (data not shown).

Histone genes with very rare exceptions have no introns. They encode short proteins and the genes are of such a length that it is unlikely that randomly generated polynucleotides of these lengths would contain termination codons. They are present in multiple copies in the genome, but for the purposes of this survey, I have counted the gene for each type of histone only once.

5'-non-coding exons (Fig. 6)

111 separate 5'-non-coding exons occur in 328 vertebrate genes. Very occasionally there is more than one in a single gene. They tend to be rather smaller than the internal exons. This is interesting because there are no constraints imposed by the necessity not to have in frame termination codons.

They are also found in 18 out of 80 insect genes, and again are shorter than the internal exons.

In both plants and fungi there is so far only a single recorded example of this phenomenon.

5'-untranslated parts of 5'-exons (Figs. 3 & 6)

In all groups the non-coding parts of these exons are on average shorter than the mean lengths of the internal exons, particularly where there is a preceding wholly non-coding

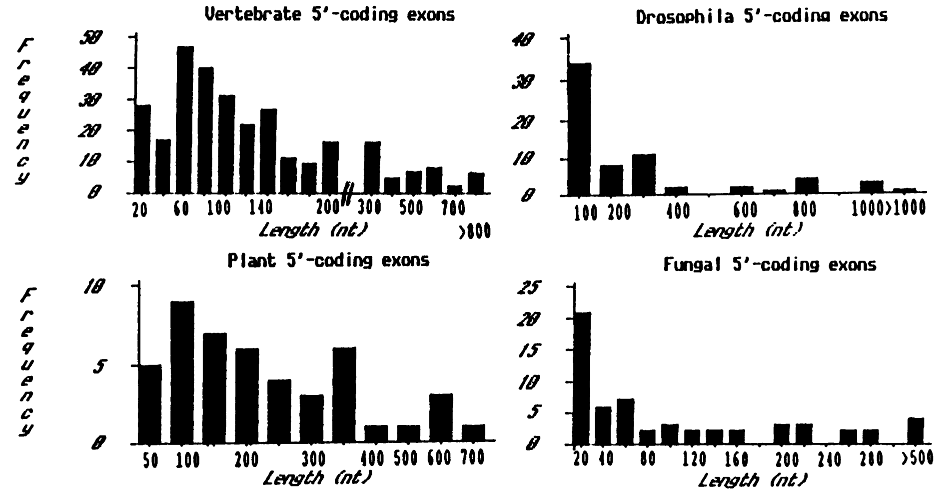

Fig. 7. Lengths of 5′-coding exons. The figures on the abscissae are the upper lengths of the bins. The lengths of the exons in the last bins are: Vertebrates — 828, 829, 831, 837, 1047, 1080; Drosophila — 1312; Fungi — 578, 651, 770, 1224 nt.

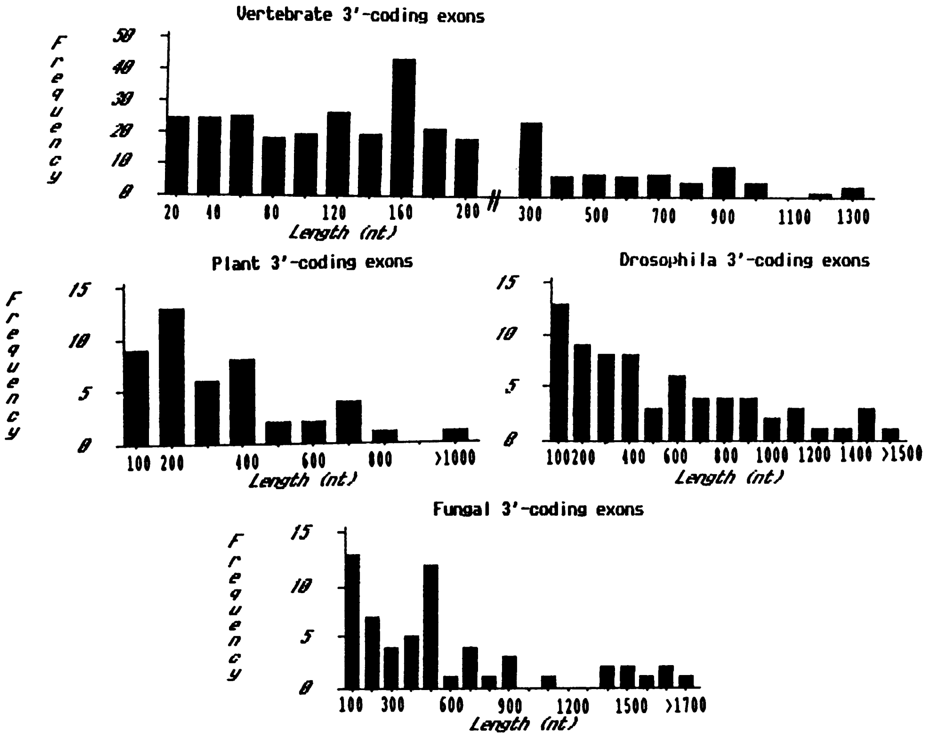

Fig. 8. Lengths of 3′-coding exons. The figures on the abscissae are the upper lengths of the bins. The lengths of the exons in the last bins are: Vertebrates — 1259, 1546, 1802; Drosophila — 2523; Fungi — 2331; Plants — 1732 nt.

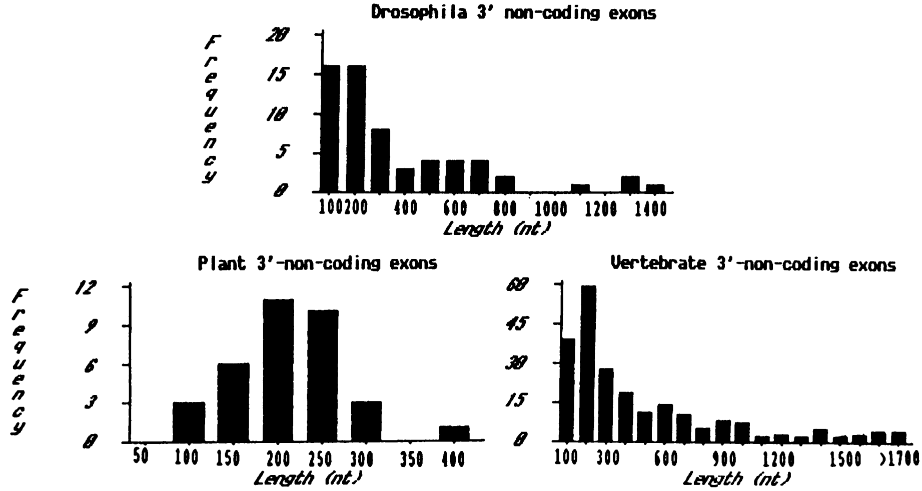

Fig. 9. Lengths of 3′-untranslated parts of 3′-exons. The figures on the abscissae are the upper lengths of the bins. The lengths of the exons in the last bins are: Vertebrates — 1815, 1970, 2413, 2499 nt.

Table IV. Genes with long exons

	5′	internal	3′
VERTEBRATES			
Rat			
angiotensinogen	829		
apo-AIV			1000
cytochrome P-450d	828		
α-fibrinogen			1140
acyl-CoA dehydrogenase			1259
Human			
cytochrome P-450-4	837		
clotting factor VII			1802
clotting factor VIII		3106	
apo-B		7572	
pre-EGF			1062
mid-size neurofilament pr.	1080		1546
elastin	950		
preproenkephalin B			975
Mouse			
Cytochrome-P1-450	837		
68 kD neurofibrillar pr.	1047		
Chicken			
β-tubulin			1061
AcCh receptor		1375, 961	
PLANTS			
Zea hsp 70			1733
Hordeum amylase		814	
Pisum amylase		891	

FUNGI			
Neurospora ATPase			2192
ATP/ADP carrier			810
Glu dehydrogenase			1043
Saccharomyces actin			1436
nin 28			896
CDC 17			1615
TUB 1			1319
tubulin		886	
Aspergillus acetamidase			821
aldehyde dehydr.		1202	
Candida β-tubulin			1422
Dictyostelium DG17	1224		
Mucor TEF2		1090	
Drosophila			
yellow			1388
dopa decarboxylase			1350
per	949	824	
yolk protein 1			1100
yolk protein 2			1094
notch		6147	
zest			818
actin	920		
tubulins			1337
			814
eve			990
Rh2 opsin		915	
chorion protein			813
Kruppel			1364
srg			984
I(2)gl		1191	
glued		1304	
tny		1314, 2649	
Na channel		1031	
engrailed	1312		
chaoptin		911	
Gart		1051	1289
abl		1635	
nina c		864, 938	
Draf-1	906		
metallothionein			1011
glucose-6-P dehydrogenase			1075

exon. Again, there are no constraints on the nucleotide sequence except for the requirement for a preferred sequence immediately preceding the initiation codon (11). Perhaps there is a preference for only a short sequence of nucleotides immediately before the initiation codon to ease scanning mechanisms for the correct positioning of the mRNA on the ribosome prior to translation.

5′-coding exons (Fig. 7)

The 5′-coding parts of the first exons in vertebrates and plants are, on average, about the same size as the internal exons. However, their length is rather more variable, with a few as short as only 3 nt, coding just for the initiating methionine residue. There are a few long ones, some of which clearly belong to protein gene families, such as cytochrome P-450, neurofilament proteins and keratins, so they may be somewhat over-represented in the sample chosen. In insects and fungi the 5′-coding part of the first exon is, on average, shorter than the length of the internal exons, but this may be a consequence of the greater length of the internal exons in these groups.

3′-coding exons (Fig. 8)

On average, these are longer than the internal exons in each group, though there is no obvious explanation for this.

3′-non-coding exons (Fig. 9)

In vertebrates, on average, these are more than twice as long as the 3′-coding exons. There is no obvious reason for this, though perhaps the chance of generating the poly-A addition signal AAATAA from a random sequence is low and therefore not likely to occur within a short distance of the termination codon.

In other groups, the 3′-non-coding exons tend to be shorter than the 3′-coding ones. In the fungi only a small number have been reported.

Separate 3′-non-coding exons

These are rare – only 23 examples have been reported in vertebrates, one in a plant and 3 in insects. The non-coding end of the 3′-coding exon is less than 45 nt in 21 out of these 23 cases and the lengths of the separate 3′-non-coding exons are generally longer than most other vertebrate exons. 9 are longer than 300 nt. A plausible mechanism for their origin is the insertion of a non-coding intron into an already existent long exon.

Long exons (Table IV)

Naora *et al.* (5) have calculated that there is only a very low chance of a random sequence of nucleotides exceeding 550 nt without encountering an in frame termination codon, while Senapathy (8) puts 600 nt as a likely upper limit for the length of exons for the same reason. It may be objected that protein-coding exons are not random sequences of nucleotides, but they must surely have originated from random processes. A small but significant number of exons are over 800 nt in length and some of these may obviously have arisen by chance but this will have been a rare occurrence. A more probable explanation of their origin is their incorporation into the genome following the action of a reverse transcriptase on an mRNA transcribed from a gene that already contained introns.

In vertebrates these long exons occur in only 6% of recorded genes and are commonest in the first and last coding exons. Several of the long exons which form the first coding exon are found in families of genes such as the cytochrome P-450 family and genes coding for structural proteins. In these cases they have probably arisen as a single event and spread through the families concerned during their evolution.

The genes for Factor VIII and lipoprotein apo-B have exceptionally long exons of 3106 and 7572 nt respectively.

In plants there are only 3 exons more than 800 nt long. In insects long exons are more common as 31% of sequenced genes contain them, most frequently in the 3′-coding position or internally. Fungal genes also contain an appreciable proportion of longer exons. They are found in 21% of the sequenced genes, most frequently in the 3′-coding position. These facts, combined with the comparatively large proportion of small introns in these groups (see above), suggest that there may be selective pressure to reduce the sizes of their genomes by removing introns and making those that are left as small as possible.

CONCLUSIONS

This sŭrvey confirms the idea that exons are rarely over 800 nt long, except in some organisms—particularly *Saccharomyces*, some other fungi and *Drosophila*. A plausible explanation is that introns were probably present in primitive organisms. New proteins could have arisen by splicing out useless coding information between blocks of coding sequences thereby creating proteins containing more than one domain, each of which might have had distinctive binding or other functions. There is some evidence that certain exons do code for domains with specific functions, but this is by no means a universal rule (4). In some species there may have been pressures to contract the size of the genome so that introns would have been lost. This has been carried to extremes in *Saccharomyces*. *Drosophila* may represent a part way house in this direction. Prokaryotes have gone even further down this road with the virtually complete exclusion of introns from their genomes.

However, there may be some situations in which introns have been inserted into genes. This is the most satisfactory explanation for the presence of introns in the 5′- and 3′-non-coding parts of genes, though it is also possible that exonic sequences around them could have mutated to give sequences which no longer have coding functions.

REFERENCES

1. Witkowski, J.A. (1988) Trends Biochem. Sci. 13, 110−113.
2. Gilbert, W. (1978) Nature 271, 501−503.
3. Blake, C.C.F. (1978) Nature 273, 267−269.
4. Blake, C.C.F. (1983) Nature 306, 535−537.
5. Naora. H., Miyahara, K. & Curnow, R.N. (1987) Proc. Natl. Acad. Sci. USA 84, 6195−6199.
6. Naora,H. & Deacon, N.J. (1982) Proc. Natl. Acad. Sci. USA 85, 6196−6200.
7. Traut, T.W. (1988) Proc. Natl. Acad. Sci. USA 85, 2944−2948.
8. Senapathy, P. (1986) Proc. Natl. Acad. Sci. USA 83, 2133−2137.
9. Bingham, P.M., Chou, T-B., Mims, I. & Zacher, Z. (1988) Trends Genet. 4, 134−138.
10. Gillies, S.D., Morrison, S.L., Oi, V.T. & Tonegawa, S. (1983) Cell 33, 717−728.
11. Kozak, M. (1987) Nucl. Acids Res. 15, 8125−8145.

APPENDIX

INSECTS	REFERENCES	
Drosophila melanogaster		
Myosin H chain (part)	Bernstein	MCB 86, 6, 2511
Myosin H chain (part)	Wassenberg	JBC 87, 262, 10741
Myosin L chain	Falkenthal	P. 85, 82, 449
ftz	Laughon	N. 84, 310, 25
Alcohol dehydrógenase	Kreitman	N. 83, 304, 414
Alcohol dehdrogenase related	Schaeffer	G. 87, 117, 61
antennapedia	Schneuwly	EMBO 86, 5, 734
yellow	Geyer	EMBO 86, 5, 2657
Dopa decarboxylase	Eveleth	EMBO 86, 5, 2665
engrailed	Kassis	EMBO 86, 5, 3583
per	Jackson	N. 86, 320, 187
Tropomyosin	Basi	JBC 86, 261, 829
Amylase	Boer	NAR 86, 14, 8399
Yolk protein 1	Hung	NAR 81, 9, 6410
Yolk protein 2	Hung	JMB 83, 164, 481
Yolk protein 3	Yan	NAR 87, 15, 67
dunce	Chen	P. 86, 83, 9313
dunce	Chen	N. 87, 329, 721 ì
Protein kinase C	Rosenthal	EMBO 87, 6, 433
Opsin (nina E)	O'Tousa	Cell 85, 40, 839
Rh2 opsin	Cowman	Cell 86, 44, 705
Opsin	Fryxell	EMBO 87, 6, 443
notch	Kidd	MCB 86, 6, 3094
Ribosomal protein 49	O'Connell	NAR 84, 12, 5495
Ribosomal protein A1	Qian	NAR 87, 15, 987
zeste	Pirrotta	EMBO 87, 6, 791
cad	Mlodzik	Cell 87, 48, 465
Actin	Sanchez	JMB 83, 163, 533
Heat shock proteins (4)	Southgate	JMB 83, 165, 35
Heat shock protein	Hacket	NAR 83, 11, 7016
Heat shock protein	Delaney	JMB 86, 189, 1
heat shock protein	Pauli	JMB 88, 200, 47
α-tubulins (3)	Theurkauf	P. 86, 83, 8477
β-tubulins (2)	Rudolph	MCB 87, 7, 2231
eve	Frasch	EMBO 87,6, 749
Dfd	Regulski	EMBO 87, 6, 767
Cuticle proteins (3)	Snyder	Cell 82, 29, 1047
Cuticle protein	Henikoff	Cell 86, 44, 33
Gart; Cuticle protein	Henikoff	G. 87, 117, 711
Contractile protein	Karlik	Cell 84, 37, 469
P element	Laski	Cell 86, 44, 7
H and L	Snyder	JMB 83, 166, 101
dash	Hoffman	Cell 83, 35, 393
Chorion genes (2)	Spradling	EMBO 87, 6, 1045
kruppel	Rosenberg	N. 86, 319, 336
white	O'Hare	JMB 84, 180, 437
srg (3)	Vincent	JMB 85, 186, 149
rudimentary	Freund	JMB 86, 189, 25
epi 28/29	Cherbas	JMB 86, 189, 617
sgs 3,7,8	Garfinkel	JMB 83, 168, 765
sgs 5	Shore & Guild	JMB 86, 190, 149
hunchback	Tautz	N. 87, 327, 383
Calmodulin	Smith	JMB 87, 196, 471
l(2)gl	Jacob	Cell 87, 50, 215
Dint-1	Rijsewijk	Cell 87, 50, 649
tra	Boggs	Cell 87, 50, 739
shaker	Baumann	EMBO 87, 6, 3419
glued	Swaroop	P. 87, 84, 6501
Metallothioneine	Gustavo	G. 86, 112, 493
rosy	Keith	G. 87, 116, 67
Na^+ channel	Salkoff	NAR 87, 15, 8569
sup-white-apricot	Chou	EMBO 87, 6, 4095
Cu-Zn O dismutase	Seto	NAR 87, 15, 10601
spalt	Frei	EMBO 88, 7, 197
chaoptin	Reinke	Cell 88, 52, 291
mst(3)gl-9	Kuhn	EMBO 88, 7, 447
AcCh receptor	Nef	EMBO 88, 7, 611
Ac Ch receptor related	Wadsworth	MCB 88, 8, 778
c-abl	Henkemeyer	MCB 88, 8, 843
ninaC (2)	Montell	Cell 88, 52, 757
Draf-1	Nishida	EMBO 88, 7, 775
zipper	Zhao	EMBO 88, 7, 1115
Glucose-6-P dehydrogenase	Fouts	Gene 88, 63, 261
Chiromonas Globin	Antoine	Gene 87, 56, 41
Globin	Trewitt	NAR 87, 15, 5494
Hyalophora Cecropin B	Xanthopoulus	EJB 88, 172, 371

FUNGI		
Neurospora		
ATPase	Hager	P. 86, 83, 7695
ATP/ADP carrier	Arends	EMBO 84, 3, 379
Histones H3 & H4	Woudt	NAR 83, 11, 5347
Glutamate dehydrogenase	Kinnaird	Gene 83, 26, 253
Reiske Fe-S protein	Harnisch	EJB 85, 149, 95
Ribosomal protein L-27	Kreader	NAR 87, 15, 9027
pyrB	Newbury	Gene 86, 43, 51
Metallothioneine	Munger	EMBO 85, 4, 2665
cyt-21	Kuiper	JBC 88, 263, 2840
cyt-18	Akins	Cell 87, 50, 331
Aspergillus		
Glucoamylase	Boel	EMBO 84, 3, 1581
Triose P isomerase	McKnight	Cell 86, 46, 143
argB	Upshall	MGG 86, 204, 349
Alcohol dehydrogenase	Gwynne	Gene 87, 51, 205
Aldehyde dehydrogenase	Pickett	Gene 87, 51, 217
Acetamidase	Corrick	Gene 87, 53, 63
β-tubulin	May	Gene 87, 55, 231
Histone H2A	May	Gene 87, 58, 59
pyrG	Oakley	Gene 87, 61, 385
pyrB	Wilson	NAR 88, 16, 2339i
Alcohol dehydrogenase	McKnight	EMBO 85, 4, 2093
Phosphoglycerate kinase	Clements	CG 85, 9, 293
Trichoderma		
endoglucanase	Penttila	Gene 86, 45, 253
endoglucanase II	Teeri	Gene 87, 51, 43
Saccharomyces		
Ribosomal protein S10	Leer	NAR 82, 10, 5869
Ribosomal protein 51	Teem	P. 83, 80, 4403
Ribosomal protein L-29	Kaufer	NAR 83, 11, 3123
Ribosomal protein S33	Leer	NAR 83, 11, 7759
Ribosomal protein L3	Schultz	JB 83, 155, 8 i
Ribosomal protiens S16A,28,L46	Leer	NAR 84, 12, 6685 i
Ribosomal protein 59	Teem	NAR 84, 12, 8295
Ribosomal protein 29	Mitra	JBC 84, 259, 9218
Ribosomal protein L16	Leer	FEBS 84, 175, 371 i
Ribosomal protein L32	Dabeva	JBC 87, 262, 16055 i
Actin	Ng	P. 80, 77, 3912
kin 28	Simon	EMBO 86, 5, 2697
MAT 1	Miller	EMBO 84, 3, 1064
Ubiquitin	Ozkaynak	EMBO 87, 6, 1431
CRY1	Larkin	MCB 87, 7, 1764
TUB1	Schatz	MCB 86, 6, 3711
Calmodulin	Takedo	P. 87, 84, 3580
CDC2	Hindley	Gene 84, 31, 129
Topoisomerase	Vemura	NAR 87, 15, 9727
β-tubulin (NDA3)	Hiraoka	Cell 84, 39, 349
DNA ligase	Barker	EJB 162, 659
Dictyostelium		
Myosin H chain	Warrick	P. 86, 83, 9433
M4	Kimmel	NAR 80, 8, 5599
UDPG phosphorylase	Ragheb	NAR 87, 15, 3891
Ubiquitin (2)	Giorda	MCB 87, 7, 2097
DG17	Driscoll	MCB 87, 7, 4482
Ribosomal protein 1024	Steel	NAR 87, 15, 10285
PYR 5–6	Jacquet	MGG 88, 211, 441
Schizophyllum IG2	Dons	EMBO 84, 3, 2101
Physarum β-Tubulin	Monteiro	JMB 87, 193, 432
Histone H4	Wilhelm	NAR 87, 15, 5478
Ligninase	Smith	NAR 88, 16, 1219
Candida β-Tubulin	Smith	Gene 88, 63, 53
Ribosomal protein L-25	Woudt	CG 87, 12, 193
Podospora OMPase	Turcq	Gene 87, 53, 201
Colletotrichum Cutinase	Ettinger	BC 87, 26, 7883
Mucor TEF2	Sundstrom	NAR 87, 15, 9997
Thermomyces Actin	Wildeman	NAR 88, 16, 2553·
Ascobolus Met2	Goyonne	Gene 88, 63, 297
PLANTS		
Glycine		
Albumin	Higgins	JBC 86, 261, 11124 i
Uricase II	Nguyen	P 85, 82, 5040
Leghaemoglobin	Brisson	P. 82, 79, 4055 i
Leghaemoglobin	Ostergaard Jensen	N. 81, 291, 677
Actin	Shah	P. 82, 79, 1022 i
Nodulin	Katinakis	P. 85, 82, 4157
Nodulin (2)	Sandal	NAR 87, 15, 1507
Conglycinin	Schuler	NAR 82, 10, 8234
Sb PRP1	Hong	JBC 87, 262, 8367
Glycinin	Fukazawa	NAR 87, 15, 8117
Heat shock protein	Czarnezka	MCB 88, 8, 1113 i
Lipoxygenase 3	Yenofsky	MGG 88, 211, 215
Zea		
A1	Schwarz-Semmer	EMBO 87, 6, 287
Triose-P isomerase	Marchionni	Cell 86, 46, 133
Alcohol dehydrogenase	Dennis	NAR 85, 13, 727
waxy	Klosgen	MGG 86, 203, 237 i
Heat shock protein	Rochester	EMBO 86, 5, 451
c1	Paz-Ares	EMBO 87, 6, 3553
Zein Z4	Hu	EMBO 82, 1, 1337
Rubisco	Lebrun	NAR 87, 15, 4360
Solanum		
Inhibitor IIK	Thornburg	P. 87, 84, 744
Patatin	Rosahl	MGG 86, 203, 214
Patatin (2)	Pikaard	NAR 86, 14, 5564
Patatin	Mignery	Gene 88, 62, 27 i
Light inducible gene	Eckes	MGG 86, 205, 14
Pisum		
Lectin	Gatehouse	NAR 87, 15, 7642
Amylase	Baulcombe	MGG 87, 209, 33
ELIP	Kolanus	MGG 87, 209, 234
Rubisco	Coruzzi	EMBO 84, 3, 1671

Legumin	Lycett	NAR 84, 12, 4493
Nicotiana		
Rubisco	Mazur	NAR 85, 13, 2373
ATP synthase	Boutry	EMBO 85, 4, 2159
Pathogenesis related protein	Cornelissen	NAR 87, 15, 6799
Arabidopsis		
Alcohol dehydrogenase	Chang	P. 86, 83, 1408 i
EPSP synthetase	Klee	MGG 87, 210, 437
α-tubulin	Ludwig	P. 87, 84, 5831
β-tubulin	Oppenheimer	Gene 88, 63, 87
Acetolactate synthase	Mazur	PP 87, 85, 110
Hordeum Aleurin	Whittier	NAR 87, 15, 2515
Amylase	idem	ibid
Vicia Legumin B	Baumlein	NAR 86, 14, 2707
Vicilin	Weschke	NAR 87, 15, 10065
Convicilin	Bown	BJ 88, 251, 717
Triticale Gliadin	Rafalski	EMBO 84, 3, 1409
Glutelin	Thompson	NAR 85, 13, 6835
Carboxypeptidase	Baulcombe	JBC 87, 262, 13726
Petunia Glycine-rich protein	Condit	N. 86, 323, 180
Rubisco	Turner	NAR 86, 14, 3325
Napin Napin	Josefsson	JBC 87, 262, 12196
Napin	Scofield	JBC 87, 262, 12202
Alfalfa Glutamine synthetase	Tischer	MGG 86, 203, 221
Rice Glutelin	Takaiwa	FEBS 87, 221, 43
Parasponia **Leghaemoglobin**	Landsmann	N. 86, 324, 166
Heliotropium **Albumin**	Allen	MGG 87, 210, 211
Daucus **Extensin**	Chen	EMBO 85, 4, 2145
Lycopersicum **Protease inhibitor**	Lee	P. 86, 83, 7279
GTOMA	Holdsworth	NAR 87, 15, 10600
Phaseolus **Phaseolin**	Sleightom	P. 83, 80, 1897
Antirhinum **Chalcone synthetase**	Sommer	MGG 86, 202, 429
VERTEBRATES		
Human		
Globin β-chain	Proudfoot	Cell 80, 21, 537
Globin α-chain	Poncz	JBC 83, 258, 11599
Prolactin	Truong	EMBO 84, 3, 429
Myoglobin	Waller	EMBO 84, 3, 439
N-ras	Brown	EMBO 84, 3, 1321
R-ras	Lowe	Cell 87, 48, 137
c-sis	Chiu	Cell 84, 37, 123
c-fos	Roebroek	EMBO 85, 4, 2897
mis	Cate	Cell 86, 45, 685
HLA-Cw3	Sodoyer	EMBO 84, 3, 881
DC-3	Bass	P. 84, 81, 5199
SB	Lawrence	NAR 85, 13, 7515
HLA-DZ	Trowsdale	EMBO 85, 4, 2231
HLA-Bw58	Ways	JBC 85, 260, 11924
HLA-RS5	Srivastava	P. 87, 84, 4224
β2-microglobulin	Gussow	JI 87, 139, 3132
HLA-II γ-chain	O'Sullivan	P. 86, 83, 4484
Dihydrofolate reductase	Chen	JBC 84, 259, 3933
T cell growth factor	Holbrook	P. 84, 81, 1634
Insulin	Bell	N. 80, 284, 26
IGF-II	Dall	N. 84, 310, 777
IGF-I	Rotwin	JBC 86, 261 4828
Gastrin	Ito	P. 84, 81, 4662
Rod rhodopsin	Nathans	P. 84, 81, 4851
Blue & red-green opsins	Nathans	S. 86, 232, 193
Plasminogen	Ng	P. 84, 81, 5355
α-subunit glycopr. hormones	Fiddes	RPHR 84, 40, 43
Chorionic SMT	Selby	JBC 84, 259, 13131
Atrial natriuretic factor	Greenberg	N. 84, 312, 656
Ac Ch receptor	Shibahara	EJB 85, 146, 25
Tumour necrosis factors α & β	Nedwin	NAR 85, 13, 6361
Erythropoietin	Jacobs	N. 85, 313, 806
Growth hormone releasing factor	Mayo	P. 85, 82, 63
50 kD keratin	Marchuk	P. 85, 82, 1609
67 kD keratin	Johnson	P. 85, 82, 1896
Fibroblast 26kD protein	Haegeman	EJB 86, 159, 625
GK-6 keratin	Rosenberg	MCB 88, 8, 722
Urokinase	Riccio	NAR 85, 13, 2759
LDL receptor	Sudhof	S. 85, 228, 815
Lipoprotein E	Paik	P. 85, 82, 3445
Lipoprotein apo C-II	Wei	JBC 85, 260, 15211
Lipoprotein apo A-II	Tsao	JBC 85, 260, 15222
Lipoprotein apo B	Blackhart	JBC 86, 261, 15366
Lipoprotein apo C-III	Shelley	JMB 85, 186, 43
Haptoglobin	Maeda	JBC 85, 260, 6698
Kininogen	Kitamura	JBC 85, 260, 8610
T cell antigen receptor-β	Tunnacliffe	P. 85, 82, 5068
T cell antigen receptor-α	Yoshikai	N. 85, 316, 837
T cell antigen receptor-γ	Lefranc	P. 86, 83, 9596
CD 3	Tunnacliffe	EMBO 87, 6, 2953
Vasopressin & oxytocin	Sausville	JBC 85, 260, 10236
β-actin	Nakajima-Iijima	P. 85, 82, 6133
Interleukin-2 receptor	Ishida	NAR 85, 13, 7579
Prealbumin	Tsuzuki	JBC 85, 260, 12224
Pancreatic polypeptide	Leiter	JBC 85, 260, 13013
C reactive protein	Lei	JBC 85, 260, 13377
Phosphoglycerate kinase	Michelson	P. 85, 82, 6965
Pulmonary surfactant protein	White	N. 85, 317, 361
Ferritin H chain	Costanzo	NAR 86, 14, 721
Protein C	Plutzky	P. 86, 83, 547
Alcohol dehydrogenase	Duester	JBC 86, 261,2027
Cytochrome P-450 (C21)	Higashi	P. 86, 83, 2841
Cytochrome P450–4	Quattrochi	P. 86, 83, 6731
Salivary Pro-rich protein	Kim	JBC 86, 261, 6712
Serum albumin	Minghetti	JBC 86, 261, 6747
Tissue plasminogen activator	Freizner Degen	JBC 86, 261, 6972
GM-CSF	Miyatake	EMBO 85, 4, 2561

G-CSF	Nagata	EMBO 86, 5, 575
Interleukin-1α	Furutani	NAR 86, 14, 3167
Interleukin-1β	Clark	NAR 86, 14, 7897
Interleukin-6	Yasukawa	EMBO 87, 6, 2939
Interleukin-5	Campbell	P. 87, 84, 6629
Heat shock protein	Hickey	NAR 86, 14, 4127
Aldolase B	Tolan	MBM 86, 3, 245
Catalase	Quan	NAR 86, 14, 5327
Preproglucagon	White	NAR 86, 14, 4719
T3	Tunnacliffe	EMBO 86, 5, 1245
Glucose-6-P dehydrogenase	Martini	EMBO 86, 5, 1849
Osteocalcin	Celeste	EMBO 86, 5, 1885
Furin	Roebroek	EMBO 86, 5, 2199
Factor VIII	Gitschier	N. 84, 312, 328 ì
Factor IX	Yoshitake	BC 85, 24, 3736
Factor X	Leytus	BC 86, 25, 5098
Factor XI	Asakai	BC 87, 26, 7221
Factor VII	O'Hara	P. 87, 84, 5158
Factor XII	Cool	JBC 87, 262, 13662
Prothrombin	Friezner Degen	BC 87, 26, 6165 ì
Neuropeptide Y	Minth	JBC 86, 261, 11974
γ-crystallin	Den Dunnen	Gene 85, 38, 197
β-crystallin	Hogg	JBC 86, 261, 12420
Epidermal growth factor	Bell	NAR 86, 14, 8439
Phenylalanine hydroxylase	DiLella	BC 86, 25, 743
SLPI	Stetler	NAR 86, 14, 7883
CANP	Miyake	NAR 86, 14, 8809
LCAT	McLean	NAR 86, 14, 9397
Adenosine deaminase	Wiginton	BC 86, 25, 8234
Myelin proteolipid	Diehl	P. 86, 83, 9807
Metallothioneine-1F	Varshney	MCB 86, 6, 26
Ribosomal protein S-14	Rhoads	MCB 86, 6, 2774
Lactalbumin	Hall	BJ 87, 242, 738
α-foetoprotein	Gibbs	BC 87, 26, 1332
pS2	Jeltsch	NAR 87, 15, 1401
PNP	Williams	JBC 87, 262, 2332
Tumour growth factor	Derynck	NAR 87, 15, 3188
β-hexosaminidase	Proia	JBC 87, 262, 5677
β_2-tubulin	Lewis	JMB 85, 182, 11
Midsize neurofilament subunit	Myers	EMBO 87, 6, 1617
β_2-adrenergic receptor	Kobilka	JBC 87, 262, 7321
β_1-adrenergic receptor	Frielle	P. 87, 84, 7920l
α_2-adrenergic receptor	Kobilka	S. 87, 238, 650 1
M1 AcCh receptor	Allard	NAR 87, 15, 10604
hsc-71	Dworniczak	NAR 87, 15, 5181
Calcyclin	Ferrari	JBC 87, 262, 8325
Salivary amylase	Nishida	Gene 86, 41, 299
Thyroglobulin	Parma	JMB 87, 196, 769
Thymidine kinase	Flemington	Gene 87, 52, 267
Phosphlipase A2	Seilhamer	DNA 86, 5, 519
Elastin	Indik	P. 87, 84, 5680
Antitrypsin S	Long	BC 84, 23, 4828
Fc receptor	Suter	NAR 87, 15, 7295
IP-10	Luster	MCB 87, 7, 3723
ISG-15	Reich	P. 87, 84, 6394
Thrombomodulin	Jackman	P. 87, 84, 6425
Tyrosine hydroxylase	O'Malley	BC 87, 26, 6910
Pepsinogen	Sogawa	JBC 83, 258, 5306
Pepsinogen C	Hayano	JBC 88, 263, 1382
hst	Yoshida	P. 87, 84, 7305
Complement factor B	Campbell	P. 83, 80, 4464
Myeloperoxidase	Molishita	JBC 87, 262, 15208
Preproenkephalin B	Horikawa	N. 83, 306, 611
ERC-1	Van Duin	NAR 87, 15, 9195
Medullasin	Nakamura	NAR 87, 15, 9601
Alkaline phosphatase	Millan	NAR 87, 15, 10599
Leucocyte common antigen	Streulin	JEM 87, 166, 1548
Fatty acid binding protein	Sweetser	JBC 87, 262, 16060
CD1a	Martin	P. 87, 84, 9189
Elastase III	Tani	JBC 88, 263, 1231
Creatine kinase B	Daouk	JBC 88, 263, 2442
Rat		
Lactalbumin	Qasba	N. 84, 308, 377
Somatostatin	Montminy	P. 84, 81, 3337
Angiotensinogen	Tanaka	JBC 84, 259, 8063
Myosin L-chain	Nadel	NAR 84, 12, 7175
Myosin L-chain	Periasamy	JBC 84, 259, 13595
Embryonic myosin H-chain	Periasamy	JBC 85, 260, 15856
Trypsin	Craik	JBC 84, 259, 14255
Chymotrypsin	Bell	JBC 84, 259, 14265
Elastase	Swift	JBC 84, 259, 14271
Luteinising hormone β-chain	Jameson	JBC 84, 259, 15474
Growth hormone releasing factor	Mayo	N. 85, 314, 464
Atrial natriuretic factor	Argentin	JBC 85, 260, 4568
Cytochrome P-450d	Sagawa	JBC 85, 260, 5026
Cytochrome P-450b	Suwa	JBC 85, 260, 7980
Cytochrome P450 IIE1	Umeno	JBC 88, 263, 4956
Cholecystokinin	Deschenes	JBC 85, 260, 1280
α1 acidic glycoprotein	Reinke	JBC 85, 260, 4397
β-casein	Jones	JBC 85, 260, 7042
PEP carboxykinase	Beale	JBC 85, 260, 10748
Retinol binding protein	Laurent	JBC 85, 260, 11476
Retinol binding protein	Demmer	JBC 87, 262, 2453
Asialoglycoprotein receptor	Leung	JBC 85, 260, 12523
Fatty acid binding protein	Sweetser	JBC 86, 261, 5553
β-crystallin	den Dunnen	P. 86, 83, 2857
Interleukin-3	Cohen	NAR 86, 14, 3641
Lipoproteins apo A-I, A-IV, C-II	I Haddad	JBC 86, 261, 13268
Lipoprotein apo-E	Fung	JBC 86, 261, 13778
IGF II	Frunzio	JBC 86, 261, 17138
GSH S-transferase	Telakowski-Hopkins	P. 86, 83, 9393
Aldolase B	Tsutsumi	JMB 85, 181, 153
Aldolase A	Joh	JMB 86, 190, 404
c-H-ras	Ruta	MCB 86, 6, 1706
Calmodulin	Nojima	JMB 87, 193, 439
Androgen sensitive protein	Delaey	NAR 87, 15,1631
hsc 73	Sorger	EMBO 87, 6, 993
MBP A	Drickamer	JBC 87, 262, 2582
Fibrinogen-α	Crabtree	JMB 85, 185, 1
Fibrinogen-γ	Morgan	NAR 87, 15, 2774
GSH S-transferase	Okuda	JBC 87, 262, 3858
Tropomyosin-α	Opazo	JBC 87, 262, 4755
Mast cell protease II	Benfey	JBC 87, 262, 5377
Epoxide hydratase	Falany	JBC 87, 262, 5924
Neuropeptide Y	Larhammar	P. 87, 84, 2068
Heme oxygenase	Muller	JBC 87, 262, 6795
Troponin T	Breitbart	JMB 86, 188, 313
Ferritin L-chain	Leibold	JBC 87, 262, 7335
Acyl-CoA oxidase	Osumi	JBC 87, 262, 8138
Enoyl-CoA hydratase	Ishii	JBC 87, 262, 8144
Parvalbumin	Berchtold	JBC 87, 262, 8696
Pyruvate kinase	Cognet	JMB 87, 196, 11
MAP-I	Fung	JBC 87, 262, 9298
Glycine N-methyl transferase	Ogawa	EJB 87, 168, 141
Ornithine transcarbamylase	Takiguchi	P. 87, 84, 6136
p9Ka	Barraclough	JMB 87, 198, 13
Enolase	Sakimura	Gene 87, 60, 103
Calcium binding protein	Perret	EJB 88, 172, 43
Arginase	Ohtake	JBC 88, 263, 2245
Pancreatic polypeptide	Yonekura	JBC 88, 263, 2990
GdX	Toniolo	P. 88, 85, 851
Creatine kinase B	Benfield	Gene 88, 63, 227
Neurotensin	Kislauskis	JBC 88, 263, 4963
Mouse		
Renin	Holm	EMBO 84, 3, 557
Hypoxanthine-P-ribosyl-transferase	Melton	P. 84, 81, 2147
Immunoglobulin-Cγ3	Wels	EMBO 84, 3, 2041
Immunoglobulin λ-5	Kudo	EMBO 87, 6, 103
Tumour antigen p53	Bienz	EMBO 84, 3, 2179
Ribosomal protein L-32	Dudov	Cell 84, 37, 457
Myosin L-chain	Robert	Cell 84, 39, 129
Interleukin 3	Miyatake	P. 85, 82, 316
Interleukin 4	Otsuka	NAR 87, 15, 333
H-2D	Sher	P. 85, 82, 1176
H-2-Tl	Obata	P. 85, 82, 5475
Cytochrome P1−450	Gonzalez	JBC 85, 260, 5040
Cytochrome P450 (C21)	Chaplin	P. 86, 83, 9601
GFAP	Balcarek	NAR 85, 13, 5527
Adenine-P-ribosyl-transferase	Dush	P. 85, 82, 2731
Lactate dehydrogenase A	Fukasawa	G. 87, 116, 99
Urinary protein	Clark	EMBO 85, 4, 3159
Nerve growth factor	Evans	EMBO 85, 4, 133
H-2 A β-chain	Larhammar	JBC 85, 260, 14111
Pro-rich salivary protein	Ann	JBC 85, 260, 15863
Serum amyloid A III	Stearman	NAR 86, 14, 797
Serum amyloid 1	Lowell	JBC 86, 261, 8442 1
Immunoglobulin J chain	Matsuuchi	P. 86, 83, 456 ì
Renal kallikrein	van Leeuwen	JBC 86, 261, 5532
Thy-1	Ingraham	JI 86, 136, 1482
int-2	Moore	EMBO 86, 5, 919
Adipocyte P2	Hunt	P. 86, 83, 3786
pim-1	Selten	Cell 86, 46, 603
GSH peroxidase	Chambers	EMBO 86, 5, 1221
3-glycero-P dehydrogenase	Phillips	JBC 86, 261, 10821
3T3 serine protease	idem	ibid
Thymidylate synthetase	Deng	JBC 86, 261, 16000
Erythropoietin	McDonald	MCB 86, 6, 842
Creatine kinase	Jaynes	MCB 86, 6, 2855
α-actin	Hu	MCB 86, 6, 15
68 kD neurofibrillar Pr.	Lewis	MCB 86, 6, 1529
3/10	Gewert	EMBO 87, 6, 651
G-CSF	Tsuchiya	EJB 87, 165, 7
T49	Koyama	P. 87, 84, 1609
Ii	Koch	EMBO 87, 6, 1678
Lyt 2−2	Nakauchi	NAR 87, 15, 4337
Band 3 protein (erythrocyte)	Kopito	JBC 87, 262, 8035
Protamine I	Pes	P. 87, 84, 5316
L-myc	Legouy	EMBO 87, 6, 3359
Glutamate-oxaloacetate TA.	Tsuzuki	JMB 87, 198, 21
Urinary plasminogen activator	Friezner Degen	BC. 87, 26, 8270
Nucleolin	Bourbon	JMB 88, 200, 627
Multifinger mKr2	Chowdhury	EMBO 88, 7, 1345
Chicken		
Ovalbumin	Woo	BC 81, 20, 6437
Cytochrome c	Limbach	NAR 83, 11, 8931
Myosin L-chain	Nabeshima	N. 84, 308, 333
Myosin H-chain	Molina	JBC 87, 262, 6478
Thymidylate kinase	Kwak	NAR 84, 12, 3963
Ac Ch receptor-γ,-δ	Nef	P. 84, 81, 7975
Ac Ch receptor	Nef	EMBO 88, 7, 595
Glyceraldehyde-P dehydrogenase	Sto	P. 85, 82, 1628
δ-crystallin 1	Ohn	NAR 85, 13, 1593
αA-crystallin	Thompson	Gene 87, 56, 173
Aldolase	Burgess	JBC 85, 260, 4604
α-aminolaevulinate synthetase	Maguire	NAR 86, 14, 1379
α-actin (cardiac)	Eldridge	Gene 85, 36, 55
β-actin (smooth muscle)	Carroll	JBC 86, 261, 8965
β 5-tubulin	Sullivan	MCB 86, 6, 4409
β 3-tubulin	Sullivan	JBC 86, 261, 13317
α-tubulin	Pratt	EMBO 88, 7, 931
Carbonic anhydrase II	Yoshihara	NAR 87, 15, 753
N-CAM	Owens	P. 87, 84, 294
Ferritin H-chain	Wilkins	MCB 87, 7, 1751
Ubiquitin	Bond	MCB 86, 6, 4602
Heat shock protein 70	Morimoto	JBC 86, 261, 12692
c-erb A	Zahraoui	EJB 87, 166, 63

Vitellogenin	van het Schip	JMB 87, 196, 245
Calmodulin	Simmen	JBC 87, 262, 4928
Ovo-inhibitor	Scott	JBC 87, 262, 5899
Vimentin	Zehner	JBC 87, 262, 8112
Link protein	Kiss	P 87, 84, 6399
B-L	Bourlet	EMBO 88, 7, 1031
Calbindin	Wilsome	JMB 88, 200, 615
HMG-17	Landsman	JBC 88, 263, 3917
Hamster		
HMG Co-A reductase	Ham	Cell 84, 38, 275
HMG Co-A synthase	Gil	JBC 86, 261, 3717
Desmin	Quax	Cell 85, 43, 329
β-crystallin	Quax-Jenken	P. 85, 82, 5819
α A-crystallin	van den Heuvel	JMB 85, 185, 273
Pro-rich salivary protein	Ann	JBC 87, 262, 3958
Rabbit		
MHC Class I Antigen	Marche	IG 85, 21, 71
Tumour necrosis factor	Ito	DNA 86, 5, 519
Synovial collagenase	Fini	BC 87, 26, 6156
Phosphofructokinase	Lee	JBC 87, 262, 4195
Calsequestrin	Zarain-Herzberg	JBC 88, 263, 4807
Ca^{2+}-ATPase	Korczak	JBC 88, 263, 4813
Cow		
Preprotachykinin A	Nawa	N. 84, 312, 729
Preprotachykinin B	Kotani	N. 86, 83, 7076 i
Cytokeratin	Reiser	EMBO 85, 4, 2261
Cytokeratin 19	Bader	EMBO 86, 5, 1865
Steroid C-21 hydroxylase	Chung	P. 86, 83, 4243
Thyroglobulin	Parma	JMB 87, 196, 769
Prochymosin	Hidaka	Gene 86, 43, 197
Xenopus		
Ribosomal protein L1	Loreni	EMBO 85, 4, 3483
TF IIIA	Tso	NAR 86, 14, 2187
Actin (cardiac)	Mohun	EMBO 86, 5, 3185
POMC	Martens	EJB 87, 165, 467
Vitellogenin	Gerber-Huber	NAR 87, 15, 4737
XIH box 2	Wright	EMBO 87, 6, 4083
α-tubulin	Smith	BJ 88, 249, 465
α-globin (larval)	Knochel	NAR 88, 16, 1625
Flounder Anit-freeze protein	Davies	JBC 84, 259, 9241
Guinea pig Insulin	Chan	P. 84, 81, 5046
Quail Troponin I	Baldwin	P. 85, 82, 8080
Sheep Metallothioneine	Peterson	EJB 86, 160, 579
Sheep Wool proteins (2)	Kuczek	EJB 87, 166, 79
Mole rat α-crystallin	Hendriks	P. 87, 84, 5320
Pig Growth hormone	Vize	Gene 87, 55, 339

Only the first-named author is given.
Abbreviations for journals: P. Proc. Natl. Acad. Sci.; NAR Nucleic Acids Res;
EJB Europ. J. Biochem.; JBC J. biol. Chem.; S. Science; N. Nature;
MGG Mol. Gen. Genetics; CG Curr. Genetics; MCB Mol. Cell. Biol;
BC Biochemistry; PP Plant Physiol.; G. Genetics; BJ Biochem J.
JB J. Bact.;MBM. Mol. Biol. & Med.; IG Immunogenetics; RPHR Recent Progr.
Hormone Res.
Since no sequences were published before the twentieth century, the dates of all references omit the first two figures.