
An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs

Marilyn Kozak

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received July 6, 1987; Revised and Accepted September 23, 1987

ABSTRACT

5'-Noncoding sequences have been compiled from 699 vertebrate mRNAs. (GCC)GCC²CCATGG emerges as the consensus sequence for initiation of translation in vertebrates. The most highly conserved position in that motif is the purine in position -3 (three nucleotides upstream from the ATG codon); 97% of vertebrate mRNAs have a purine, most often A, in that position. The periodical occurrence of G (in positions -3, -6, -9) is discussed. Upstream ATG codons occur in fewer than 10% of vertebrate mRNAs-at-large; a notable exception are oncogene transcripts, two-thirds of which have ATG codons preceding the start of the major open reading frame. The leader sequences of most vertebrate mRNAs fall in the size range of 20 to 100 nucleotides. The significance of shorter and longer 5'-noncoding sequences is discussed.

INTRODUCTION

To search for signals that might influence early steps in translation, I have scrutinized the 5'-noncoding sequences of 699 vertebrate mRNAs, which are identified in the Appendix. The survey included all sequences to which I had access in the published literature except those in which the functional initiator codon had not been clearly identified or where it seemed possible that the cloned cDNA sequence fell short of the true initiator codon. To minimize redundancy, I did not enter every available sequence for large multigene families (especially globins, histones and immunoglobulins), but the sequences that were omitted were usually similar to the ones that were entered; two cases where that is not true are described in footnotes k and n. When a particular gene was sequenced from more than one organism, I entered both sequences if they differed in at least two positions near the ATG codon. Otherwise I entered only one--the one for which more accessory information was available or, arbitrarily, the human sequence. All mRNA sequences are written with T in place of U since nearly all of the sequences were determined by analyzing DNA.

RESULTS AND DISCUSSIONContext

Previous surveys of eukaryotic mRNA sequences (1, 2) revealed that the sequence flanking functional initiator codons is nonrandom: CC²CCATGG was proposed as the consensus sequence for initiation of translation in higher eukaryotes. The present survey confirms and extends that conclusion using a larger and more diversified data base.

Table 1 and Fig. 1 show a distinctive pattern over the 12 nucleotide stretch preceding the ATG initiator codon. The whole region is deficient in T resi-

Table 1. Frequency of A, C, G and T around the translational start site in vertebrate mRNAs.

POSITION:	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+4
percent A	23	26	25	23	19	23	17	18	25	61	27	15	23
percent C	<u>35</u>	<u>35</u>	<u>35</u>	26	<u>39</u>	<u>37</u>	19	<u>39</u>	53	2	<u>49</u>	55	16
percent G	23	21	22	<u>33</u>	23	20	44	23	15	36	13	21	46
percent T	19	18	18	18	19	20	20	20	7	1	11	9	15

Data were compiled from the 699 sequences listed in the Appendix. A window of 12 nucleotides preceding the initiator codon is presented, as well as one nucleotide (position +4) following the ATG codon. The most abundant nucleotide in each position is underlined. Values that are >50% or $\geq$ twice the frequency of the next most abundant nucleotide in that position are shown in boldface.

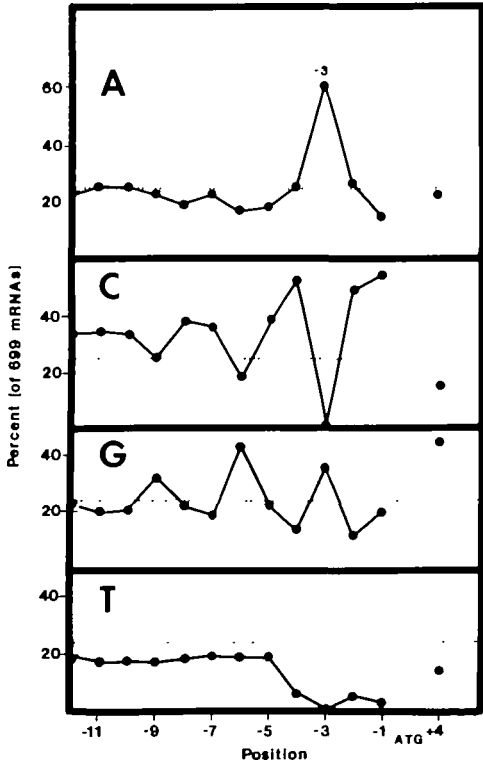


Figure 1. Frequency of A, C, G and T around the ATG initiator codon in 699 vertebrate mRNAs, which are listed in the Appendix. The dotted line across each panel shows the 25% value that would be expected on a random basis.

dues, especially in positions -1 to -4. C is the preferred nucleotide in every position except -3, -6 and -9. In positions -6 and -9, G is preferred. Position -3 shows the strongest bias: 61% of vertebrate mRNAs have an A in that position, 36% have G, and only 3% have a pyrimidine. On the 3'-side of the ATG codon G is the preferred flanking nucleotide. Thus the expanded consensus sequence for initiation in vertebrate mRNAs is (GCC)GCC^ACCATGG. Site-directed mutagenesis experiments have confirmed the contribution of every nucleotide in positions -1 to -6, as well as the G in position +4 (3, 4), but the significance of the GCC motif in positions -9 to -7 remains to be established. Because mutations in positions -3 and +4 have the strongest influence on translational efficiency, for practical purposes an initiator codon can be designated "strong" or "weak" by considering only those positions. That view is supported by the data in Table 2, in which initiator codons are grouped according to the nucleotides in positions -3 and +4. Among 699 vertebrate mRNAs, only 23 have a pyrimidine 3 nucleotides upstream from the functional initiator codon, and 17 of those "compensate" by having G in position +4. Thus only six mRNAs out of 699 lack the preferred nucleotide in both of the crucial positions. One would expect those six mRNAs to be translated very inefficiently, which is not inconsistent with the fact that four of them encode hormones or lymphokines (entries 264, 397, 407 and 650). The other two (entries 172 and 281) are members of large multigene families, for which reason one cannot assess the extent to which the particular mRNA that corresponds to the cloned cDNA is translationally active.

Considerable evidence (3) supports the idea that the ATG codon and flanking sequences function as a "stop signal" for the migrating 40S ribosomal subunit, which binds directly only at the 5'-end of the mRNA. Consequently, position as well as context determines which ATG is the functional initiator codon. In cases where there are two ATG codons in equally favorable contexts near the 5'-end of a message, it would be incorrect to conclude that either ATG is equally likely to initiate translation. Theory predicts and experiments confirm (54) that ribosomes initiate exclusively at the 5'-proximal ATG codon when it lies in a favorable context.

The repetition of G in positions -3, -6 and -9 is quite noticeable in

Table 2. Sequences flanking ATG codons in vertebrate mRNAs

Sequence	Functional initiator codons	"Nonfunctional" UPSTREAM ATG codons ^a
-3 +4		
AnnATGG	175	4
A.....A	114	5
A.....C	63	8
A.....T	73	4
G.....G	130	8
G.....A	47	7
G.....C	47	5
G.....T	27	5
C.....G	9	7
C.....A	2	8
C.....Y	4	12
T.....G	8	4
T.....A	0	13
T.....Y	0	16
Total #	699	106 ATGs in 59 mRNAs ^{b, c}

^a"Nonfunctional" is a provisional designation. Upstream ATG codons are expected to function--an expectation that has been verified with some viral mRNAs but not yet with cellular mRNAs. The more important point is that the indicated upstream ATG codons are not absolute barriers to initiating downstream: a downstream ATG codon starts the major open reading frame in these mRNAs.

^bThe tabulation of upstream ATG codons does not include the 34 oncogenes listed in the Appendix, since they comprise a separate group vis-à-vis the frequency of upstream ATG codons (see text).

^cThirty of the upstream ATG codons in this set derive from just four mRNAs: entries 73, 283, 556 and 599. Excepting those four entries and the proto-oncogenes, only 9% of vertebrate mRNAs have upstream ATG codons and they typically have only one.

Table 3. Length distribution of vertebrate mRNA leader sequences.

Length:	<10	10-19	20-29	30-39	40-49	50-59	60-69	70-79	nucleotides
# of mRNAs:	4	10	23	29	36	38	37	40	
Length:	80-89	90-99	100-199	200-299	300-399	400-499	500-599	≥600	nucleotides
# of mRNAs:	15	22	68	19	1	3	2	2	

The table is based on 346 mRNAs for which the transcriptional start site has been mapped. In the case of genes that produce multiple transcripts, only the longest leader was scored. In three cases where ribosomes initiate at the first and second ATG codons (see footnote 9 in the Appendix), two values were entered in the table; in all three cases there are fewer than 10 nucleotides between the cap and the first functional ATG codon.

Fig. 1. Trifonov (5) has pointed out that there is a strong preference for G in the first position of codons throughout the coding region of both prokaryotic and eukaryotic mRNAs, and he postulates that the periodicity of G residues helps ribosomes stay in frame during translation. An interesting possibility is that "frame monitoring" begins shortly upstream from the initiator codon in eukaryotes. In support of that idea, it has been shown by site-directed mutagenesis that correct initiation is strongly favored by placing a purine in position -3 and G in position -6, but the facilitating effect is completely lost when the purines are shifted one nucleotide to the left or right (3, 4).

In recent surveys of mRNA sequences from plants (6), *Drosophila* (7) and yeast (8), the most striking finding was conservation of a purine--usually A--in position -3. The reported values for position -3 were 53% A and 23% G in 47 plant mRNAs; 82% A and 13% G in 77 *Drosophila* mRNAs; and 81% A in 96 yeast mRNAs. Although such data encourage the idea that A or G in position -3 somehow favors initiation in all eukaryotic systems, the effect of context on initiation has yet to be tested experimentally in nonvertebrates. The overall A-richness of leader sequences in yeast and plant mRNAs somewhat diminishes the statistical significance of finding A most often in position -3. Leader sequences on *Dictyostelium* mRNAs are also notoriously A+T rich. This might be a hint that ribosomes from lower eukaryotes and plants are less able to deal with secondary structure than are metazoan ribosomes (9).

Upstream ATG codons

Three points about the occurrence of upstream ATG codons merit comment.

- (i) They are relatively rare in vertebrate mRNAs-at-large, as indicated in Table 2. The raw data from which Table 2 was compiled are given in the Appendix.
- (ii) The big exception to the foregoing generalization are proto-oncogenes (entries 454 to 487), nearly two-thirds of which produce mRNAs that have ATG codons--usually more than one--preceding the start of the major open reading frame. In view of the inhibitory effect of upstream ATG codons (3), it is probably not an accident that activation of proto-oncogenes by transduction or translocation often deletes the cumbersome leader sequence. Preliminary evidence (10, 11) encourages the hypothesis that the expression of some oncogenes is regulated in part at the level of translation.
- (iii) The context around "nonfunctional" upstream ATG codons differs strikingly from the functional initiator codons listed in Table 2. Whereas functional initiator codons are rarely preceded by a pyrimidine in position -3, upstream ATG codons often occur in that unfavorable context. The notion that scanning is "leaky" in such mRNAs--with some 40S ribosomal subunits stopping and initi-

ating at the upstream site while some reach the second ATG codon--is supported by some experimental evidence (3, 12).

Leaky scanning obviously cannot account for the ability of ribosomes to initiate downstream from a strong ATG codon, and a considerable number of the upstream ATG codons in Table 2 do occur in a favorable context. Those ATG codons are nearly always followed by a terminator codon, however, and it seems likely that--after translating the upstream "minicistron" and terminating--ribosomes reinitiate at the next ATG codon downstream (12, 13 and references therein). Given leaky scanning and reinitiation as devices with some experimental justification that permit initiation at an ATG codon that is not "first," an upstream ATG codon should pose a problem only if it occurs in a favorable context for initiation and is not followed by a terminator codon before the start of the major open reading frame. Among the 699 mRNAs listed in the Appendix, only five have that problematical structure; they are entry 120 (where the upstream ATG codon lies very close to the cap and hence might be inefficient--see below); entries 520, 553 and 599 (where the potentially-inhibitory ATG codon is preceded by a far-upstream minicistron which might have a sparing effect, as explained in reference 13); and entry 247, for which I have no excuse.

It might be noted parenthetically that upstream ATG codons seem to occur more commonly in *Drosophila* than in vertebrate mRNAs, although I cannot cite precise statistics for *Drosophila*, nor is it always certain that the ATG-burdened leader sequence belongs to a functional form of mRNA (14).

Leader length

The precise length of the 5'-noncoding sequence is known for about half of the entries in the Appendix, and those mRNAs were used to compile the data in Table 3. Only one-fourth of the mRNAs that were scored have a leader sequence longer than 100 nucleotides. Thus the leader sequences on most vertebrate mRNAs fall in the range of 20 to 100 nucleotides. Note that the mRNAs derived from proto-oncogenes are again atypical, as nearly all of them have very long leader sequences. Also note some extraordinarily long leader sequences (400 to >800 nucleotides) that contain no upstream ATG codons; see entries 523, 524 and 650.

The effects of leader length on translational efficiency are just beginning to be explored. There is some evidence that an ATG codon is not recognized efficiently when it occurs close to (within 10 nucleotides of) the cap; that might explain the rare examples of cellular mRNAs in which ribosomes initiate at the first and second ATG codons (see entries 143, 297, 330 and footnote g in the Appendix). A few viral mRNAs that seem to translate efficiently have leader sequences only 9 or 10 nucleotides long (15-17), but the possibility that some ribosomes reach the second ATG in those mRNAs has not been ruled out. In the case of SV40 late 16S mRNA, the ATG codon that initiates the agnoprotein occurs 10 nucleotides down from the cap and it is clearly recognized inefficiently, despite a favorable context; lengthening the leader sequence by 33 nucleotides seems to improve initiation at the agnoprotein start site, with the result that fewer ribosomes reach the downstream VP1 start site (18). In the case of vaccinia virus, a novel transcriptional maneuver (19) adjusts both the length and the context in a way that favors the efficient translation of late mRNAs.

Whereas a leader sequence that is too short might be deleterious, there is no evidence that long leader sequences are incompatible with efficient translation provided that inhibitory features (notably secondary structure and upstream ATG codons) are avoided. Many of the naturally occurring long 5'-noncoding sequences are G+C rich, however, and therefore secondary structure might negate any advantage of length. From an opposite perspective, the presence of secondary structure in GC-rich leader sequences might necessitate that they be

50. Soares, M.B., Ishii, D.N. and Efstratiadis, A. (1985) Nucl. Acids Res. 13, 1119-1134.
51. Soares, M.B., Turken, A., Ishii, D., Mills, L., Episkopou, V., Cotter, S., Zeitlin, S. and Efstratiadis, A. (1986) J. Mol. Biol. 192, 737-752.
52. Royer-Pokora, B., Kunkel, L., Monaco, A., Goff, S., Newburger, P., Baehner, R., Cole, F.S., Curnutte, J.T. and Orkin, S.H. (1986) Nature 322, 32-38.
53. Orkin, S.H. (1987) Trends in Genetics 3, 207.
54. Kozak, M. (1983) Cell 34, 971-978.

APPENDIX

Entry No.	Messenger RNA/source ^a	Leader length ^b	Sequence flanking initiator codon	Upstream ATGs ^c			References ^d
				w	w/t	a s/t	
Acetylcholine receptors							
001	α nicotinic, hu muscle		ctccggttagccCATGg				Noda '83 Nature 305,818
002	α nicotinic, rat neural	225	cgggtcttagacATGg				Boulter '86 Nat 319,368
003	β nicotinic, bo muscle		cgcggcgccggtATGa				Tanabe '84 EJB 144,11
004	γ nicotinic, hu muscle		agctgaggcaccATGc				Shibahara '85 EJB 146, 15
005	ε nicotinic, bo muscle		ccagacagcggaATGg				Takai '85 Nat 315, 761
006 ^e	atrial muscarinic, po	270, 400	agagaacacaaaATGa		3		Peralta '87 Sci 236,600
007	cerebral muscarinic, po	>444	ccaccacgaccATGa	1	2	1	Kubo '86 Nature 323,411
008	α ₁ acid glycoprotein, hu	36	cctggtctcagTATGg				Dente '87 EMBO 6, 2289
009	α ₁ acid glycoprotein, rat	40	agtgtcttcggcATGg				Liao '85 MCB 5,3634
Actins							
010	α-skeletal, mu	70 *	aaactagacaccATGt				Hu '86 MCB 6,15
011	α-skeletal, ch	73 *	acagccagcaacATGt				Formwald '82 NAR 10,3861
012	α-skeletal, Xp	>60	ccagcctcaaacATGt				Stutz '86 JMB 187,349
013	α-cardiac, Xp	>53 *	taacctgccactATGt				"
014	α-cardiac, ch	60 *	ctatcagccaagATGt				Chang '85 NAR 13,1223
015	α-smooth muscle, hu		gcagctccagctATGt				Ueyama '84 MCB 4,1073
016	α-smooth muscle, ch	88 *	ttgacatagaagATGt				Carroll '86 JBC 261,8965
017	β-cytoplasmic, hu	84 *	cgccagctcaccATGg				Ng '85 MCB 5,2720
018	β-cytoplasmic, rat	80 *	caccagttcgccATGg				Nudel '83 NAR 11,1759
019	β-cytoplasmic, ch	96 *	ccacagccagccATGg				Kost '83 NAR 11,8287
020	γ-cytoplasmic, hu	>73	ctgcgggtcgcaATGg				Erba '86 NAR 14,5275
021	3rd cytopl isoform, ch	49 *	gcaggcccaatcATGg				Bergsma '85 MCB 5,1151
022	APRT, hu		tcttcgcacgccaATGg				Broderick '87 PNAS 84,3349
023	APRT, mu	60	acgcacgcggccATGt				Dush '85 PNAS 82,2731
024	Adenosine deaminase, hu	95	caagagggcaccATGg				Ingolia '86 MCB 6,4458
025	Adenosine deaminase, mu	~135	acgctcggaaccATGg				"
026	AdoHcy hydrolase, rat		gacttcgccagcATGg				Ogawa '87 PNAS 84,719
027	Adenylate kinase, ch	>57	cacagcagcagcATGt				Kishi '86 JBC 261,2942
028	Adipocyte P2, mu	67	aagggtttacaaaATGt				Hunt '86 PNAS 83,3786
029	Adrenodoxin, bo	>164	ccccgacaggctATGg				Okamura '85 PNAS 82,5705
030	Albumin, hu serum	39	gccttttggcacaATGa				Minghetti '86 JBC 261, 6747
031	Albumin, ch serum	41	taatctgcagccATGa				Haché '83 JBC 258,4556
032	ADH, α subunit, hu class I		gacagaatcaacATGa				Ikuta '86 PNAS 83,634
033	ADH, β subunit, hu "	70	gacagaaacgacATGa				Duester '86 JBC 261,2027
034	ADH, γ subunit, hu "	>80	gacagaatcaatATGa	1			Höög '86 EJB 159,215
035	ADH-AA, mu liver	>101	aggacagacggcATGa		1		Edenberg '85 PNAS 82,2262
036	ALDH, hu liver	>30	tagcccgctgcgATGt				Braun '87 NAR 15,3179
037	Aldolase A, rat	66, 116*	gccaccggcaccATGc				Joh '86 JMB 190,401
038	Aldolase B, rat	81 *	gtacctgtcatcATGg				Tsutsumi '85 JMB 181,153
039	Aldolase B, ch	72 *	caataagtcaccATGa				Burgess '85 JBC 260,4604
040	Aldolase C, mu	>60	acaactgtcatcATGc				Paolletta '86 EJB 156,229
041	ALP, hu intestinal		tqcccccagacATGc				Henthorn '87 PNAS 84,1234
042	ALP, hu liver/bone	>176	ttggggtgcaccATGa				Weiss '86 PNAS 83,7182
043	ALA-D, hu	>66	ctggcccccgcATGc	1			Wetmur '86 PNAS 83,7703
044	ALA-D, rat	>48	ccggcccccaccATGc				Bishop '86 NAR 14,10115
045	5-ALV synthase, ch	81	gcaggagggaaggATGg				Maguire '86 NAR 14,1379
046	α-amylase, hu salivary	220 *	cttcaaaagcaaaATGa				Nishide '86 Gene 41,299
047	α-amylase, mu salivary	95 *	cagcatagcaaaATGa	1			Hagenbüchle '81 Nat 289,643

No.			w w/t s s/t	
048	α -amylase, mu pancreatic	17	cttcaaaagcaaaATGa	Hagenbüchle '80 Cell 21,179
049	Amyloid-A4 (Alzheimer's)	>146	cgcaggggtcgagATGc	Kang '87 Nature 325,733
050	Amyloid (SAA2), mu serum	34 *	gacaccagcaggATGa	Lowell '86 JBC 261,8442
051	Androgen-BP, 45K, rat	>33	cagctgctaaactATGg	Joseph '87 PNAS 84,339
052	And.-induced RP2, mu	>40	aggacgcggcccATGa	King '86 NAR 14,5159
053	And-induced S-protein, rat	>51	ttttctggcaagATGa	McDonald '84 EMBO 3,2517
054	Angiotensinogen, rat	61 *	cacagatccgtgATGa	Tanaka '84 JBC 259,8063
055	Antifreeze protein, flounder	49	caagttctcaaaATGg	Davies '84 JBC 259,9241
056	" ocean pout	>57	tcagccacagccATGa	Li '85 JBC 260,12904
057	Arginase, hu liver	>56	aagtgtcagagcATGa	Haraguchi '87 PNAS 84, 412
058	Arginase, rat liver	>26	ccctggatgagcATGa	1 Kawamoto '87 JBC 262, 6280
059	Arginosuccinate lyase hu	>114	gaaccgcccacATGg	O'Brien '86 PNAS 83,7211
060	Arginosucc. synthase, hu	>75	atccacagcgtATGt	Boek '83 NAR 11, 6505
061	AspAT, mu mitochondrial	>88	ttaccgcccaccATGg	Obaru '86 JBC 261,16976
062	AspAT, mu cytoplasmic	>54	cattctgtcgagATGg	"
063	AspAT, ch mitochondrial		cacgtgcccaccATGg	Jausi '85 JBC 260,16060
064	ATP/ADP carrier, hu	>70	ccttctttcaacATGa	Battini '87 JBC 262, 4355
065	ATPase-Ca ²⁺ ra slow twch	>129	gccccgcagccATGg	MacLennan '85 Nat 316,696
066	ATPase-Ca ²⁺ ra fast twch	>81	gaaggagcgcaATGg	Brandl '86 Cell 44,597
067	ATPase (Na ⁺ K ⁺)- α , rat brain	>237	agcgccgccaccATGg	Shull '86 Biochem 25, 8125
068	" " α III "	>140	ggagcgcgaagATGg	"
069	" " β , human	>120	cccgccatcgccATGg	2 Kawakami '86 NAR 14, 2833
070	" " β rat kidney	>460	tgagcagacaccATGg	Young '87 JBC 262,4905
071	" " α sh kidney	>264	accacgcgcgtATGg	Shull '85 Nature 316, 691
072	" " β sh kidney	>528	tgaccgcgcaccATGg	Shull '86 Nature 321, 429
073	" (H ⁺ K ⁺), rat stomach	>206	cacctagccaccATGg	5 Shull '86 JBC 261, 16788
074	ATPase, Xp mitochondrial		caagcgcgagcATGt	4 Weeks '87 PNAS 84,2798
075	Atrial natriuretic factor	90	accacgcgcaccATGg	Argentin '85 JBC 260, 4568
076	Avidin, ch	>43	cctgtgtcagagATGg	Gope '87 NAR 15, 3595
077	BPGM, hu erythrocyte	>110	tcagccatcagtATGt	1 Joulin '86 EMBO 5, 2275
078	Bone Gla protein, rat	>48	ctagcagacaccATGa	Celeste '86 EMBO 5, 1885
079	Brain S100- α protein, bo	>89	gtaagcttcgagATGg	Kuwano '86 FEBS 202, 97
080	Brain S100- β protein, rat	>120	ggagcctccgggATGt	Kuwano '84 NAR 12, 7455
081	Brain spcif. gene 0-44, rat	66 +	taggcgcgcgagATGg	± Teou '86 MCB 6, 768
082	C-reactive protein, hu	104	caggacgtgaccATGg	Lei '85 JBC 260, 13377
083	Caerulein, Xp	>50	ccttctgaaagcATGt	Richter '86 JBC 261, 3676
084	Calcitonin, hu	>74	cagagaggtgtcATGg	Jonas '85 PNAS 82, 1994
085	Calcitonin, rat	132	cagggaggtatcATGg	Amara '84 MCB 4, 2151
086	Ca ²⁺ -binding protein, ch	>102	tgaccccccaacATGa	Hunziker '86 PNAS 83, 7578
087	Ca ²⁺ " rat brain	>130	agcgctgcaccATGg	Yamakuni '86 NAR 14, 6768
<u>Calmod. family Ca²⁺-binding proteins:</u>				
088	Calmodulin, rat	85	ttcgctcgaccATGg	Nojima '87 JMB 193, 439
089	Calmodulin pROM3, rat	>70	agcccttgacagcATGg	Nojima '87 MCB 7, 1873
090	Calmodulin, ch	91	ggcgagaccaccATGg	Putkey '83 JBC 258, 11864
091	Calmodulin, Xp	>70	aactattccgaaATGg	Chien '84 MCB 4, 507
092	Oncomodulin, rat	97	gcgggacagaaaATGa	Gillen '87 JBC 262, 5308
093	Parvalbumin, rat	73 *	ccaagttgcaggATGt	Epstein '86 JBC 261, 5886
094	Calpain, po	>90	tgagtcacagccATGt	1 Sakihama '85 PNAS 82, 6075
095	Calcyclin (S100-related) hu	65 *	cagccctcagccATGg	Ferrari '87 JBC 262, 8325

No.			w w/t	s s/t	
096	Carbamyl-P-synthetase rat	140	aacatcttcaaaATGa	1	Lagacé '87 JBC 262,10415
097	Carbonic anhydrase I, hu		cagtagaagataATGg		Barlow '87 NAR 15, 2386
098	Carbonic anhydrase II, mu	60	accggcgtgaccATGt		Venta '85 JBC 260, 12130
099	Carbonic anhydrase III, hu	>43	agggaagcgaccATGg		Lloyd '86 Gene 41, 233
100	Carbonic anhydrase II, ch	39	ggccggcgccaccATGt		Yoshihara '87 NAR 15, 753
101	Cartilage-link protein ch	>135	gtgactgtgaagATGa		Deak '86 PNAS 83, 3766
102	α -Casein, rat	62 *	atctttagcaaccATGa		Yu-Lee, '86 NAR 14, 1883
103	β -Casein, rat	52 *	gacttgacagccATGa		Jones '85 JBC 260, 7042
104	β -Casein, mu	>55	gacttgacagccATGa		Yoshimura '86 NAR 14, 8224
105	γ -Casein, rat	56 *	gatcaagtaaccATGa		Hobbs '82 NAR 10, 8079
106	Catalase, hu kidney	70	aaaccgcacgctATGg		Quan '86 NAR 14, 5321
107	Catalase, rat liver	>83	caatcctacaccATGg		Furuta '86 PNAS 83, 313
108	Cathepsin B, hu	>195	ctggcttccaacATGt	1	Chan '86 PNAS 83, 7721
109	Cathepsin D, hu	>51	gcgcgcgcgcgcATGc		Faust '85 PNAS 82, 4910
110	Cholecystokinin, hu	64 *	aatccaaaagccATGa		Takahashi '86 Gene 50, 353
111	Cholecystokinin, rat	59 *	gcacccgaagatATGa		Deschenes '85 JBC 260, 1280
112	Chromogranin, bo	180	cccggttccgccATGc		Benedum '86 EMBO 5, 1495
113	proChymosin, bo	25	cccagatccaagATGa		Hidaka '86 Gene 43, 197
114	Chymotrypsinogen-B, rat	22	ttgaccagcaccATGg		Bell '84 JBC 259, 14265
<u>Coagulation factors & modulators:</u> see also Fibrinogen, Inhibitors, PA, Thrombospondin					
115	Factor VII, hu	>35	agagatttcacacATGg		Hagen '86 PNAS 83, 2412
116	Factor VIII, hu	170	tagcaataagtcATGc	1	Gitachier '84 Nat 312,326
117	Factor X, bo	>75	aaggcccccaccATGg		Pung '84 NAR 12, 4481
118	Factor XIII-a, hu placenta	>84	gtaaagtcaaaaATGt		Grundmann '86 PNAS 83,8024
119	von Willebrand factor, hu	246	ttgcaggggaaagATGa	1	Verweij '86 EMBO 5, 1839
120	Protein C, hu	75 *	agtgcctccagaATGt	1	Plutsky '86 PNAS 83, 546
121	Protein S, hu	>108	cgccgcttcgaaATGa		Hoakins '87 PNAS 84, 349
122	Protein S, bo	>35	gcccgcttccgccATGa		Dahlbäck '86 PNAS 83, 4199
123	proCollagen, $\alpha 2$ (I), ch	133	tagcaagtagacATGc	2	Vogeli '81 PNAS 78, 5334
124	proCollagen, $\alpha 1$ (I), ch	>100	taatatattagacATGt	2	Yamada '83 JBC 258, 14914
125	proCollagen, $\alpha 1$ (II), rat	155	tcgcggtgagccATGa	1	Kohn '85 JBC 260, 4441
126	Collagenase, hu skin	>68	acaaagggccagtATGc		Goldberg '86 JBC 261,6600
<u>Complement components & modulators:</u>					
127	preClr, hu	>51	gggccttgagaaATGt		Journet '86 Bio. J. 240,783
128	preC2, hu	>36	agggaggacaccATGg		Bentley '86 Bio. J. 239,339
129	preC3, hu	>60	tgccccagcaccATGg		deBruijn '85 PNAS 82,708
130	preC3, mu	56	ttttcttctactATGg		Wiebauer '82 PNAS 79,7077
131	preC4, mu	61	gatcctccagccATGc		Nonaka '86 PNAS 83,7883
132	Factor B, hu	129	ccttccaacgccATGg	1	Wu '87 Cell 48,331
133	Factor I, hu		aacacctccaacATGa		Catterall '87 BioJ 242,849
134	Decay-accelerate factor hu	>66	accggcgccgccATGa		Caras '87 Nature 325, 545
135	C1 inhibitor, hu		gtcgccgcccagATGg		Bock '86 Biochem 25, 4292
136	Conalbumin, ch	76	ccttgcccaaacATGa		Jeltsch '82 EJB 122,291
137	Corticotropin-RF, sh	>127 *	gcgccccctaacATGc		Furutani '83 Nature 301,537
138	Creatine kinase, rat brain	>29	gcgcgcgcgcgcAUGc		Benfield '85 Gene 39, 263
139	Creatine kinase, ra muscle	>50	gacgcgcgcaccATGc		Putney '84 JBC 259, 14317
140	Creatine kinase-B, ch	>42	gtagggacagccATGc		Hossle '86 NAR 14, 1449
141	α A-Crystallin, ha	68	gccaaagaagAACATGg		Heuvel '85 JMB 185, 273
142	α B-Crystallin, ha	43	cacctaagcaccATGg		Quax-Jeuken '85 PNAS 82,5819
143	β A3/A1-Crystallin, mu	56	ccaaccaccaagATGg		Peterson '86 Gene 45,139
144	β B1-Crystallin, ch	>70	ctgaccaccgcgATGt		Hejtmancik '86 JBC 261,982

Nucleic Acids Research

No.			w w/t	s s/t	
145	8B1-Crystallin, rat	38 *	gcacagaaaccATGt		Dunnen '86 PNAS 83, 2855
146	γ-Crystallin, rat	42	caaacacagccATGg		Moormann '83 JMB 171, 353
147	δ1-Crystallin, ch	86 *	acgtcgtccgaaATGg		Hayashi '85 EMBO 4, 2201
148	Cell cycle tsll gene, hu	>78	tgcttgagtgcATGa	1	Greco '87 PNAS 84, 1565
149	Cell cycle "cdc2", hu	>140	cattgactaactATGg		Lee '87 Nature 327, 31
150	Cyclin, hu	>118	cactccgccaccATGt		Almendral '87 PNAS 84, 1575
151	Cyclophilin, hu T-cell		gtactattagccATGg		Raendler '87 EMBO 6, 947
152	Cystic fibrosis Ag, hu	51	tcogtgggcatcATGt		Dorin '87 Nature 326, 614
153	Cytochrome c, mu	*	ttagaattaaaaATGg		Limbach '85 NAR 13, 617
154	Cytochrome c, ch		ctagtactgacaATGg		Limbach '83 NAR 11, 8931
155	Cytochr. c oxidase-IV, bo	32	ggtggcagcagaATGt		Lomax '84 PNAS 81, 6295
156	NADPH-cyto P450 reductase	70	tgatcaccaacATGg		Murakami '86 DNA 5, 1
Cytochrome P-450 proteins:					
157	P ₁ -450, human	122 *	cctacactgacATGc		Jaiswal '85 NAR 13, 4503
158	P-450 _{wp} , human	>68	aggaagtagtgATGg		Beaune '86 PNAS 83, 8064
159	P-450c17, human	>41	caccagccaccATGt		Chung '87 PNAS 84, 407
160	P-450c21, human	9, 53, 118	ggcgctctcgccATGc	±	Higashi '86 PNAS 83, 2841
161	P-450 _{scc} , human	75	tggtggggacagcATGc		Chung '86 PNAS 83, 8962
162	P-450 ₁ , rabbit	>43	aggaagaagagaATGg		Johnson '87 JBC 262, 5918
163	P-450(M-1), male rat liver		gagaaggctgccATGg		Yoshioka '87 JBC 262, 1706
164	P-450 _{MC} , rat	>72	gccacctagatcATGc		Yabusaki '84 NAR 12, 2929
165	P-450 _{dbl} , rat	>51	agcaaggcagccATGg		Gonzalez '87 DNA 6, 149
166	P-450-PCN, rat	>82	agacctgcagggATGg	1	" '85 JBC 260, 7435
167	P-450 _o , rat	30	tacaccaggaccATGg		Mizukami '83 PNAS 80, 3958
168	P-450 _o , rat		tcactggccactATGc		Nagata '87 JBC 262, 2787
169	P-450 _o , ch	>39	ccctgcccaccATGg		Hobbs '86 JBC 261, 9444
170	Cytokeratin type I, bo	25	aacagcatcaccATGt		Rieger '85 EMBO 4, 2261
171	Cytokeratin 19, bo	59	tcctgcttcgccATGa		Bader '86 EMBO 5, 1865
172	Cytokeratin Endo A, mu	79	cagacttcaccaATGt		Vasseur '85 PNAS 82, 1155
173	Cytokeratin Endo B, mu	>54	tctccagacaagATGa		Singer '86 JBC 261, 538
174	Cytokeratin-8, Xp	>36	cacagctccaccATGt		Franz '86 PNAS 83, 6475
175	Desmin, ha	81	cacgcccaccATGa		Quax '85 Cell 43, 327
176	Diazepam inhibitor, rat	>117	cacctcgccagtATGt		Mocchetti '86 PNAS 83, 7221
177	DHFR, hu	71	cccgctgctgtcATGg		Chen '84 JBC 259, 3933
178	DHFR, mu	55, 115	cccgctgccatcATGg		Nunberg '80 Cell 19, 355
179	Dihydropteridine reductase	>80	ggcaggagcaggATGg	1	Lockyer '87 PNAS 84, 3329
180	Elastase I, rat pancreatic	22	tctctccacaacATGc		MacDonald '82 Biochem
181	Elastase II, " "	22	cacggacacaccATGa		21, 1453
182	Elastin, hu	>21	tttctccccgagATGg		Indik '87 PNAS 84, 5680
183	Endopeptidase, neutral, ra	>	agatttttaggtgATGg		Devault '87 EMBO 6, 1317
184	Endothelial cell GP, hu	>38	agctgctgagccATGg		Jaye '86 Sci 233, 541
185	preproEnkephalin A, hu	130 *	agcgtcaactccATGg		Noda '82 Nature 297, 431
186	preproEnkephalin B, hu		tgctgagacaggATGg		Horikawa '83 Nat 306, 611
187	preproEnkephalin, rat	156 *	accggagcccccATGg		Rosen '84 JBC 259, 14309
188	Enolase, non-neuron, rat	>110	cagaacttcaccATGt		Sakimura '85 NAR 13, 4365
189	Enolase, neuronal, rat	>68	atcccagccatcATGt		Sakimura '85 PNAS 82, 7453
190	Epidermal GP(pre), hu	>436	ctcatcaagattATGc	1	Bell '86 NAR 14, 8427
191	Epidermal GP(pre), mu	350	gctaaataaaagATGc	1	Scott '83 Sci 221, 236
192	Epididymal proteins D, E	>82	gaaaatagaaccATGg		Brooks '86 EJB 161, 13
193	Epoxide hydrolase, rat	175 *	cagtcaggagtcATGt	1	Falany '87 JBC 262, 5924
194	Erythroid membr pr 4.1, hu	>798	aaacacaggaaacATGc	3	Conboy '86 PNAS 83, 9512
195	Erythroid potentiating, hu	>72	agagaacccaccATGg		Carmichael '86 PNAS 83, 2407

No.			w w/t	s s/t	
196	Erythropoietin, hu	>180	ccaggcgcgagATGg	1	Jacobs '85 Nature 313, 806
197	Fatty acid binding protein	48	ctcattgccaccATGa		Sweetser '86 JBC 261, 5553
198	" rat liver				
198	" rat intestine		acagctgacatcATGg		Alpers '84 PNAS 81, 313
199	FA thioesterase, rat	>310	tcaactcacagaATGg		Safford '87 Biochem 26, 1358
200	FA thioesterase, duck	>160	aattattcaggaaATGg		Poulose '85 JBC 260, 15953
201	Ferritin, L-chain, hu	181	tgccaaccaaccATGa		Santoro '86 NAR 14, 2863
202	Ferritin, L-chain, rat	200	tccggatcagccATGa		Leibold '87 JBC 262, 7335
203	Ferritin, H-chain, hu	216	ttagtcgcccgcATGa		Hentze '86 PNAS 83, 7226
204	Ferritin, H-chain, ch	151	cccgccagcgccATGg		Stevens '87 MCB 7, 1751
205	Ferritin, bullfrog	145	caaaccgctgaaATGg		Didsbury '86 JBC 261, 949
206	α -Fetoprotein, hu	44	taactagcaaccATGa		Gibbs '87 Biochem 26, 1332
207	preFibrinogen, Aa, hu	>57	cccttagaaaaagATGt		Kant '83 PNAS 80, 3953
208	preFibrinogen- γ , hu	51	cactcagacatcATGa		Rixon '85 Biochem 24, 2077
209	preFibrinogen- α , rat	58	caatcagaaactATGc	1	Crabtree '85 JMB 185, 1
210	" - β , rat		aaatctgaaaccATGa		Fowlkes '84 PNAS 81, 2313
211	" - γ , rat	44	caccagacactATGa		Morgan '87 NAR 15, 2774
212	Fibroblast growth factor	>320	cccgaggggaccATGg		Abraham '86 EMBO 5, 2523
213	Fibronectin, hu	267	caccgtctcaacATGc		Dean '87 PNAS 84, 1876
214	FSH, β -chain, bo	>48	caagtgcgccaggATGa		Each '86 PNAS 83, 6618
<u>Guanine nucleotide binding proteins:</u>					
215	G _{0q} , bovine retina	>196	gaagggggccaccATGg		Van Meurs '87 PNAS 84, 3107
216	G _{1a} , rat brain	>41	gcggacggcaggATGg		Itoh '86 PNAS 83, 3776
217	G _{2a} , rat brain	>246	cccgccgcccgcATGg		"
218	G _{2a} , bovine retina	>93	cctgccaggaccATGg		Medynski '85 PNAS 82, 4311
219	β_1 subunit, bovine retina	>94	taagattagaagATGa		Pong '86 PNAS 83, 2162
220	γ -subunit, " "	>67	gcttagcagaagATGc		Yatsunami '85 PNAS 82, 1936
221	preproGalanin, po	>225	cgtgcgtcaagATGc	1	Rokaeus '86 PNAS 83, 6287
222	Gap junction protein, rat	>31	gaatgaggcaggATGa	1	Paul '86 J Cell Biol 103, 123
223	GAP-43, rat neuronal	>60	gaagataccaccATGc		Karns '87 Sci 236, 597
224	preproGastrin, hu	65 *	tctgcagacgagATGc		Ito '84 PNAS 81, 4662
225	Gastrin-releasing, hu	>55	cccgtcgggaccATGc		Spindel '84 PNAS 81, 5699
226	Gelsolin, hu		cgtgtgccaccATGg		Yin '86 Nature 323, 455
<u>α-Globin family:</u>					
227	human adult	37	agagaaccaccATGg		Baralle '77 Cell 12, 1085
228	human embryonic (ζ)	55	caccctgccgccATGt		Proudfoot '82 Cell 31, 553
229	baboon β 1	193	tccagcgcggaATGg		Shaw '87 Nature 326, 717
230	mouse adult	32	caggaagaaaccATGg		Baralle '78 Nature 274, 84
231	rabbit adult	36	gaaggaaccaccATGg		Baralle '77 Nature 267, 279
232	goat embryonic (ζ)	46	tcagctgccaccATGt		Wernke '86 JMB 192, 457
233	duck adult, major (α^A)	36	ggagctgcaaccATGg		Erbil '82 Gene 20, 211
234	chicken embryonic	55	ctctcctgcacaATGg		Engel '83 PNAS 80, 1392
<u>β-Globin family:</u>					
235	human fetal (γ)	53	agtccagacgccATGg		Slightom '80 Cell 21, 627
236	human embryonic (ϵ)	53	aggcctggcatcATGg		Baralle '80 Cell 21, 621
237	rabbit adult	53	aaaacagacagaATGg		Baralle '77 Cell 10, 549
238	rabbit embryonic (β 3)	62	agaccagacatcATGg		Hardison '81 JBC 256, 11780
239	chicken adult	77	ccaaccgcccgcATGg		Dolan '83 JBC 258, 3983
240	chicken embryonic (ρ)	45	cccgctgccaccATGg		Beninson '81 PNAS 78, 4782
241	Xenopus adult, major	46	tcaactttggccATGg		Patient '83 JBC 258, 8521
242	Xenopus larval		tctacagccaccATGg		Banville '83 JBC 258, 7924
243	$\alpha_2\mu$ Globulin, rat salivary	70	ttccctaccaacATGa		Laperche '83 Cell 32, 453
244	preproGlucagon, hu pancr.	105 *	aagacagcagaaATGa		White '86 NAR 14, 4719
245	preproGlucagon, rat	101 *	cagataaaaaaATGa		Heinrich '84 JBC 259, 14082
246	preproGlucagon, anglerfish	>58	acggttgtaaacATGa		Lund '82 PNAS 79, 345

No.			w	w/t	s	s/t	
247	Glucose-6-P-dehydrog., hu	70 *	atattcatcatcATGg		1		Martini '86 EMBO 5,1849
248	Glucose-regl'd 78K prot, rat	206	agcgccggcaagATGa				Chang '87 PNAS 84, 680
249	Glucose transporter, hu	>180	cgagcgctgccATGg				Mueckler '85 Sci 229, 941
250	Glucose transporter, rat	>207	cgagcgcgccATGg	?			Birnbaum '86 PNAS 83, 5784
251	β -Glucuronidase, hu	>26	ggaccgggaagcATGg				Oshima '87 PNAS 84, 685
252	γ -Glutamyl transpeptidase	>228	actggacgagccATGa				Laperche '86 PNAS 83, 937
253	Glutathione peroxidase, mu	37	aacatctccagtATGt				Chambers '86 EMBO 5, 1221
254	Glutathione-S-transferase; -subunit 2, hu	>55	gagactgctatcATGg				Board '87 PNAS 84, 2377
255	-Ya subunit, rat	64 *	acagttgctgctATGt		1		Pickett '86 PNAS 83, 9393
256	-Yb1 subunit, rat	>37	agcgccagaaccATGc				Ding '85 JBC 260, 13268
257	-Yc subunit, rat	>42	gcaattgctgccATGc				Pickett '85 JBC 260, 5820
258	GST-placental, rat	70	tacgcagcagctATGc				Okuda '87 JBC 262, 3858
259	GAPDH, hu	>75	cgctcagacaccATGg				Tso '85 NAR 13, 2485
260	GAPDH, rat	71	ctcatagacaagATGg				Fort '85 NAR 13, 1431
261	GAPDH, rat	57 *	tataaaggcgagATGg				Stone '85 PNAS 82, 1628
262	Glycero-P-dehydrogenase, mu	21	gcaagcagcaccATGg				Ireland '86 JBC 261, 11779
263	Glycogen phosphorylase, hu	>113	gccgccccagccATGg				Newgard '86 PNAS 83, 8132
264	Gonadotropin- β , salmon		tgagtccttcagATGt				Trinh '86 EJB 159, 619
265	Growth hormone, hu	60	cacctagctgcaATGg				DeNoto '81 NAR 9, 3719
266	Growth hormone, rat	60	cactgagtgccgATGg				Page '81 NAR 9, 2087
267	Growth hormone, salmon	>64	ttaagagtaaaaATGg				Sekine '85 PNAS 82, 4306
268	GH-releasing factor, hu	91 *	cccggtggaaggATGc				Mayo '85 PNAS 82, 63
269	Haptoglobin, hu	30	agaccaaccaagATGa				Bensi '85 EMBO 4, 119
270	Heat shock 70K, hu	212	gcagggaaccgcATGg				Hunt '85 PNAS 82, 6455
271	Heat shock 70K, hu	119	gaagcttcagccATGc				Voellmy '85 PNAS 82, 4949
272	Heat shock 27K, hu	>40	gagtcagccagcATGa				Hickey '86 NAR 14, 4127
273	Heat shock 73K, rat	80 *	acgcaagcaaccATGt				Sorger '87 EMBO 6, 993
274	Heat shock 70K, ch	111	gaatctatcatcATGt				Morimoto '86 JBC 261, 12692
275	Heat shock 108K, ch	101	ggccgcggcatcATGa				Kulomas '86 Biochem 25, 6244
276	Heat shock 70K, trout	>60	ttattcggtaacATGt				Kothary '84 MCB 4, 1785
277	Heat shock 70K, Xp	124	aggagcgcaaatATGg				Bienz '84 EMBO 3, 2477
278	Helix-destabilizing, rat	>28	catcctaccgctATGt				Cobianchi '86 JBC 261, 3536
279	Heme oxygenase, rat	128	cccagtcgccgATGg				Muller '87 JBC 262, 6795
280	β -Hexosaminidase, α -chain	>168	gaccagcgggccATGa				Myerowitz '85 PNAS 82, 7830
281	HMG-14, hu	>145	ccccccgccagATGc				Landsman '86 JBC 261, 16082
282	HMG-17, hu	>88	gcgcgccaccATGc				Landsman '86 JBC 261, 7479
283	His-rich glycoprot., hu	>120	tggtttaacaaaATGa	3	2		Koide '86 Biochem 25, 2220
<u>Histocompatibility antigens (MHC):</u>							
284	Class I (hu): HLA-Bw58		tcagacgcccagATGc				Ways '85 JBC 260, 11924
285	Class II (hu): DR(α -chain)	64	cccaagaagaaaATGg				Schamboeck '83 NAR 11, 8663
286	Class II (hu): DR(β -chain)	>30	ctgttctccagcATGg				Tieber '86 JBC 261, 2738
287	Class II (hu): DC- β	58	ttcgttcaattATGg				Boss '84 PNAS 81, 5199
288	Class II (hu): SB(DP)-a		agacccacacaacATGc				Lawrance '85 NAR 13, 7515
289	Class I (mu): H-2K ^b		tcagtcgtcagcATGg				Kimura '86 Cell 44, 261
290	Class I (mu): H-2K ^d	>27	ccgaccagtcgATGg				Lalanne '83 NAR 11, 1567
291	Class I (mu): Tla gene T3 ^b		gatttccttaacATGa				Pontarotti '86 PNAS 83, 1782
292	Class II (mu): A(β -chain)	>36	tgtgccttagagATGg				McDevitt '87 PNAS 84, 2435
293	Class II (mu): SB(β -chain)		gtccctccagcATGg				Peterson '85 JBC 260, 14111
294	Class II (mu): E(α -chain)	48	cccaagaagaaaATGg				Mathis '83 Cell 32, 745
295	Class II (mu): E(β -chain)	52	ctctcctgcagcATGg				Saito '83 PNAS 80, 5520
296	Class II (mu): E(β 2-chain)		ctctcctcaagcATGg				Braunstein '86 EMBO 5, 2469
2979	Class II-associated Ia (In), hu	58	cagaagccagtcATGg				Strubin '86 Cell 47, 619

No.			w/w/t	s/s/t
298	Histone H1, hu		tttcttgcaccATGt	Carozzi '84 Sci224, 1115
299	" H2a, hu	46	tcagaagtagttATGt	Zhong '83 NAR 11, 7409
300	" H2b, hu	40	ctagacagtgctATGc	"
301	" H3, hu	37	tgtggttttgccATGg	"
302	" H3, hu	27	gcagttctgcgaATGg	Clark '81 NAR 9, 1583
303	" H3.3, hu	111 *	ggtctctgtaccATGg	Wells '87 NAR 15, 2871
304	" H4, hu	36	ttgctcgtcgtcATGt	Heintz '81 Cell 24, 661
305	" H2b, mu	41	tctctgttccactATGc	Sittman '83 NAR 11, 6679
306	" H3.1, mu	28	cggttacttgccATGg	"
307	" H3.2, mu	21	ttcttttagaaaATGg	"
308	" H1, ch embry.	38	acgtccgctcaccATGt	Sugarman '83 JBC 258, 9005
309	" H2A.1, ch	146	tcagtcgctgcgATGt	D'Andrea '81 NAR 9, 3119
310	" H2A.F, ch embry	70	ggcggcgccaccATGg	Harvey '83 PNAS 80, 2819
311	" H2B, ch	36	ggagagttcgacATGc	Grandy '82 JBC 257, 8577
312	" H3.3A, ch	*	gtcggcagcagcATGg	Brush '85 MCB 5, 1307
313	" H3.3B, ch	>105 *	aagtgagaaaaaATGg	Dodgson '87 NAR 15, 6294
314	" H4, ch embry.	26	caggctctcgccATGt	Sugarman '83 JBC 258, 9005
315	" H5, ch erythrocyte	109	gaagcggcgccATGA	Krieg '83 NAR 11, 619
316	" H1, Xp	28	tttacttcaaagATGA	Turner '83 NAR 11, 4093
317	" H2A, Xp	47	agcacagtaaatCATGt	Moorman '82 FEBS 144, 235
318	" H2B, Xp	35	agcagcacaaattATGc	"
319	" H3, Xp		aactgatacactATGg	Moorman '81 FEBS 136, 45
320	" H4, Xp	28	gctcaagaaagaATGt	"
321	Hydroxyindole O-MeTr'ase, bo	>120	cccagaaggaagATGt	1 Ishida '87 JBC 262, 2895
322 ^h	HMG-CoA reductase, hu	73 to 105*	tctgtagctacaATGt	Luskey '87 MCB 7, 1881; and " '85 JBC 260, 10271
323	HMG-CoA synthase, hu	63 or 122	tgctctttcaccATGc	Gil '87 PNAS 84, 1863
324	HPRT, hu	100 to 170	gcgcggctccggtATGg	Patel '86 MCB 6, 393
325	HPRT, mu	90, 118	accggctcccgctATGc	Melton '84 PNAS 81, 2147
326	Ig L-chain, kappa-II, hu		caccttctcacaATGA	Klobeck '85 NAR 13, 6499
327	Ig L-chain, kappa-III, hu		cccagaggaaccATGg	"
328	Ig L-chain, kappa-IV, hu	25	aggggcagcaagATGg	Marsh '85 NAR 13, 6531
329	Ig H-chain, IgE, hu	>56	ccgttctctaccATGg	Kenten '82 PNAS 79, 6661
330 ^g	Ig L-chain, kappa, mu	18	catcacaccagcATGg	Kelley '82 Cell 29, 681
331	Ig L-chain, lambda ₁ , mu	40	ggtttgtgaattATGg	Picard '83 PNAS 80, 417
332	Ig H-chain, mu	>45	agtcctgtcactATGA	Early '80 Cell 19, 981
333	Ig L-chain, ch		tggaattccgcccATGg	Reynaud '87 Cell 48, 379
334	Inhibin, A-subunit, hu	>144	gccagggtgagctATGg	1 1 Mayo '86 PNAS 83, 5849
335	Inhibin, A-subunit, bo	>60	gccaggggagctATGt	Porage '86 PNAS 83, 3091
	Inhibitors: see also EPA, Kininogen, Lipocortin, Macro- and Microglobulins			
336	anti-Protein C, hu	>46	agaacatccaccATGc	Suzuki '87 JBC 262, 611
337	α1-antiChymotrypsin, hu		gcagagttgagaATGg	Chandra '83 Biochem22, 5055
338	anti-Elastase, hu	17	ctctgccttcaccATGA	Statler '86 NAR 14, 7883
339	anti-placental PA, hu	>55	cagattgaacaATGg	Ye '87 JBC 262, 3718
340	antiThrombin III, hu	~70	agattagcgcccATGt	Bock '82 NAR 10, 8113
341	{α1-antitrypsin, hu liver	49 *	agtgaatcgacaATGc	(1) (1) Ciliberto '85 Cell 41, 531
342	macrophage	520 *		Perlino '87 EMBO 6, 2767
343	Insulin, hu	59 *	tgctcttctgccATGg	Bell '80 Nature 284, 26
344	Insulin-I, rat	57 *	catgtttccaacATGg	Lomedico '79 Cell 18, 545
345 ^c	Insulin, gp	60 *	catcctttcatcATGg	Chan '84 PNAS 81, 5046
346	Insulin, anglerfish	>85 *	ccccagctcatcATGg	Perler '80 Cell 20, 555
347	Insulin, salmon	>71	ttctactgcagcATGg	Hobart '80 Sci 210, 1360
			cctaccatcaccATGg	Sorokin '82 Gene 20, 367
348	Insulin-like GF I, hu	>180	acttcagaagcaATGg	1 1 Rotwein '86 PNAS 83, 77
349 ^j	Insulin-like GF II, rat	100 *	cttcagggtaccaATGg	Soares '86 JMB 192, 737
350	Integrin, ch	96	cggcgcgcgccATGg	Tamkun '86 Cell 46, 271
351 ^b	Interferon-α2, hu(LeIP A)	~70	gcaacatctacaATGg	Lawn '81 PNAS 78, 5435
352	Interferon-α, mu	~70	gcaacatcaccATGg	Shaw '83 NAR 11, 555

No.			w/w/t	s/s/t	
353	Interferon- α , rat	~72	gccacatttgccATGg		Dijkema '84 NAR 12, 1227
354	Interferon- α , bo	~70	tcaaggtccccgATGg		Capon '85 MCB 5, 768
355	Interferon- β , hu fibroblast	75	cgtgttgccaacATGa		Ohno '81 PNAS 78, 5305
356	Interferon- β_2 , hu	63	aggagcccgactATGa		Haegeman '86 EJB 159, 625
357	Interferon- γ , hu immune	130	ctctcggaacgATGa		Gray '82 Nature 298, 859
358	Interferon- γ , rat	110	agctctgagacaATGa		Dijkema '85 EMBO 4, 761
359	IFN-induced gene 6-16, hu	106	gcgcgcgcaccATGc		Kelly '86 EMBO 5, 1601
360	IFN-induced ISG-54K, hu	75	tgcagtgcaccATGa		Levy '86 PNAS 83, 8929
361	IFN-induced 15K prot. hu	>75	cagccacagccATGg		Blomstrom '86 JBC 261, 8811
362	Involucrin, hu	62 *	gtagcttctaagATGt		Eckert '86 Cell 46, 583
364	Keratin-I, 50K, hu	60	ctctctgcaccATGa		Marchuk '85 PNAS 82, 1609
365	Keratin, 67K, hu	47	ctctaagtcacATGa		Johnson '85 PNAS 82, 1896
366	Keratin-II, 56K, hu		atctctgggaaccATGg		Tyner '85 PNAS 82, 4683
367	Keratin-I, 47K, mu	>116	cctctctcagccATGa	1	Knapp '86 NAR 14, 751
368	Keratin-I, 59K, mu	25	cactacaccaccATGt		Krieg '85 JBC 260, 5867
369	Keratin B2A, sh (see also #170-174)	50	atcctctgacaccATGg		Powell '83 NAR 11, 5327
370	protein Kinase C, α/β , ra	>221	ccgcgcgcgaagATGg		Ohno '87 Nature 325, 161
371	protein Kinase C, γ , ra	>204	ttgggggggaccATGg		"
372	protein Kinase, cAMP-alt	>50	atcgccccagtcATGg		Showers '86 JBC 261, 16288
373	" II(Ca ²⁺), rat	~200	atcgccaccgcgATGg		Bennett '87 PNAS 84, 1794
374	preKininogen, hu	130, 154, 184	gattgttagatcATGa		Kitamura '85 JBC 260, 8610
375	α -Lactalbumin, hu	~26	ggggtagccaaaATGa		Hall '87 Bioch. J. 242, 735
376	LDH-A, hu	>99 *	tcacaggtccaatATGg		Taujilbo '85 EJB 147, 9
377	LDH-C, mu	>54	gtaaggctcaacATGt		Sakai '87 Bioch. J. 242, 619
378	Lamin C, hu	>200	aacctgccggccATGg		Fisher '86 PNAS 83, 6450
379	L-C acyltransferase, hu	24	accagggtctggaATGg		McLean '86 NAR 14, 9397
380	Lens MIP, bo	>50	atccccctgccATGt		Gorin '84 Cell 39, 49
381	Leukocyte adhesion pr. β , hu	>72	acaccgagggacATGc		Kishimoto '87 Cell 48, 681
382	Lipase, rat hepatic	>15	aagacgagagacATGg		Komaromy '87 PNAS 84, 1526
383	Lipase, rat lingual	48	tagcagtacaagATGt		Docherty '85 NAR 13, 1891
384	Lipoprotein lipase, hu	>174	acgcgccccgagATGg		Wion '87 Sci 235, 1638
385	Lipid binding protein, mu	65	aaggtttacaaaATGt		Phillips '86 JBC 261, 10821
386	Lipocortin-II, hu	>50	gttctcttcaaaaATGt		Huang '86 Cell 46, 191
387	apoLipoprotein A-I, rat	40 *	acatctctcaggATGa		Haddad '86 JBC 261, 13268
388	apoLipoprotein A-I, ch		ttcagcgcggaagATGa		Lewis '87 JBC 262, 7058
389	apoLipoprotein A-II, hu	58 *	actgttaccacATGa		Shelley '85 JMB 186, 43
390	apoLipoprotein A-II, mu	>41	tagtctgcccacATGa		Kunisada '86 NAR 14, 5729
391	apoLipoprotein II, VLD, ch	77 *	taccaacaaaccATGg		AB '83 NAR 11, 2529
392	apoLipoprotein A-IV, mu	91	tggaggagccaggATGt		Williams '86 MCB 6, 3807
393	apoLipoprotein B, hu	128	cgcagctggcgATGg		Protter '86 PNAS 83, 1467
394	apoLipoprotein C-II, hu	38 *	tctctggacactATGg		Wei '85 JBC 260, 15211
395	apoLipoprotein E, rat	65 *	acaactgggaagATGa		Fung '86 JBC 261, 13777
396	Luteinizing hormone (β), rat	7	atcaagaATGg		Jameson '84 JBC 259, 15474
397	<u>Lymphokines:</u>				
397	CSF-1, hu macrophage	178	ccagctgcccgtATGa		Ladner '87 EMBO 6, 2693
398	GM-CSF, mu	35	gtcctgaggaggATGt		Miyatake '85 EMBO 4, 2561
399	multi-CSF, hu (IL-3)	>38	gccgatccaaacATGa		Dorssers '87 Gene 55, 115
400	" mu	29	cagaacgagacaATGg		Miyatake '85 PNAS 82, 316
401	deleted				
402	Interleukin-1 α , hu	>45	aaagaagtcaagATGg		March '85 Nature 315, 641
403	Interleukin-1 β , hu	87 *	ctctgaagcagccATGg		Clark '86 NAR 14, 7897
404	BSP-1, hu (IL-4)	>63	cgacacctattaATGg		Yokota '86 PNAS 83, 5894
405	BSP-1, mu (20K)	63	acagagctattgATGg		Otsuka '87 NAR 15, 333
406	BSP-2, hu		aggagccagctATGa		Hirano '86 Nature 324, 73

No.			w/w/t	s/s/t
407	Lymphotoxin (TNF- β), hu	>79	ttggtttctccccATGa	Gray '84 Nature 312, 721
408	Lysophospholipase, rat	21	cagacactcactATGg	Han '87 Biochem 26, 1617
409	Lysozyme, ch	29	gacactggcaacATGa	Jung '80 PNAS 77, 5759
410	α_2 -Macroglobulin, hu	>43	tctttctgcaacATGg	Kan '85 PNAS 82, 2282
411	α_2 -Macroglobulin, rat	>63	cttttcgcagcATGg	Gehring '87 JBC 262, 446
412	Malate dehydrogenase, mu	>50	cccgccctagccATGc	Joh '87 Biochem 26, 2515
413	Malic enzyme, mu	>65	ccggtgccagccATGg	Bagchi '87 JBC 262, 1558
414	Malic enzyme, rat		acggtgctggccATGg	Magnuson '86 JBC 261, 1183
415	Mn superoxide dismutase, mu	>55	taaacctcaataATGt	Hallewell '86 NAR 14, 9539
416	Melanoma Ag p97, hu	>60	cccgacggcgccATGc	Rose '86 PNAS 83, 1261
417	Menadione reductase, rat	>74	acttctggagccATGg	Robertson '86 JBC 261, 15794
418	Metallothionein-I _A , hu	73	ccgcggctcgaaATGg	Richards '84 Cell 37, 263
419	" -I _B , hu	69	cttggtccacaaATGg	Hoguy '86 MCB 6, 2149
420	" -I _F , hu	71	cttcggcttgcaATGg	Varshney '86 MCB 6, 26
421	" -II, hu	69	cttcagctcgccATGg	Karin '82 Nature 299, 797
422	" -Ia, sh	72	cttttctccaaATGg	Peterson '86 EJB 160, 579
423	α_1 Microglobulin, hu	>72	gagcccatagccATGa	Traboni '86 NAR 14, 6340
424	β_2 Microglobulin, mu	>52	tcagtctgcagcATGg	Daniel '83 EMBO 2, 1061
425	Mullerian inhibiting subst. 10		agcaccacgATGc	Cate '86 Cell 45, 685
426	Multidrug resistance, hu	140	cgcgaggtcgggATGg	Ueda '87 JBC 262, 505
427	Mx protein, mu	>213	gagagccagacgATGg	1 Staeheli '86 Cell 44, 147
428	Myelin basic protein, mu	47	ggcttggtgtgATGg	1 Takahashi '85 Cell 42, 139
429	Myelin P2 protein, mu	>44	aaggtttacaaaATGt	Bernlohr '84 PNAS 81, 5468
430	Myelin P ₀ (peripheral), rat	>31	cctaccaccagctATGg	Lemke '85 Cell 40, 501
431	Myelin-assoc."MAG", rat	>130	ttgctggacaagATGa	Arquint '87 PNAS 84, 600
432	Myeloperoxidase, hu	>163	aggagaagagagATGg	1 Morishita '87 JBC 262, 3844
433	Myoglobin, hu	70	tcagactgcgccATGg	Weller '86 MCB 6, 4539
434	Myoglobin, mu	55	ttgaagccaccATGg	Blanchetot '86 EJB 159, 469
Myosins:				
435	H-chain, rat embry. skel.	90 **	tcagccaacactATGa	Strehler '86 JMB 190, 291
436	H-chain(fast), ch. adult	60 *	gtgagcgcgccATGg	Gulick '85 JBC 260, 14513
437	H-chain(fast), ch.embry.	101 **	taaacagcgacgATGg	"
438	L-chain 1, mu	125	cttttaatacaaaATGg	Robert '84 Cell 39, 129
439	L-chain 3, mu	94	tagaactccatcATGt	"
440	L-chain 2, rat skel.	56	aggatctaagacATGg	Nudel '84 NAR 12, 7175
441	L-chain 1, ch skel.	123	aagcaacacaaaATGg	Nabeshima '84 Nat 308, 333
442	L-chain 3, ch skel.	71	caactctcaatcATGt	" '82 NAR 10, 6099
443	L-chain 2A, ch cardiac		ctctgcgaagacATGg	Winter '85 JBC 260, 4478
444	Neurofilament p68, mu		ccggccgccaccATGa	Lewis '86 MCB 6, 1529
445	Neuroleukin, mu	>52	gggtccctcgccATGg	Gurney '86 Sci 234, 566
446	Neural cell adhesion, ch	>215	ccgcggctgcgATGc	Edelman '87 Sci 236, 799
447	Neural cell adhesion, mu	161	cggcagttttacaATGc	Barthels '87 EMBO 6, 907
448	Neuropeptide Y, hu	86 *	gcgccagccaccATGc	Minth '86 JBC 261, 11974
449	Nerve GF (a) mu	42	acacctgttaccATGt	Evans '85 EMBO 4, 133
450	" " (β) submax. gland	99 *	ctctcagtgaaATGc	Selby '87 MCB 7, 3057
451	" " (γ) mu	42	acacctgtcaccATGt	Evans '85 EMBO 4, 133
452	Nuclear prot. N1/N2, Xp	>64	gggttgctgatcATGg	Franke '86 EMBO 5, 3547
453	Nucleoplasm, Xp	>113	tatctacgtgacATGg	2 Dingwall '87 EMBO 6, 69

No.			W	w/t	S	S/T
	proto-Oncogenes:					
454	{ c-abl, mu, type I mRNA	>93	ggccacgggaccATGt			Ben-Neriah '86 Cell 44,577
	{ type IV mRNA	>200	tattattgctttATGg	1	1	
455	{ c-bcl-2, hu, 5.5 kb mRNA	>1000	cctctgggaaggATGg	6	2	Croce '86 PNAS 83, 5214
	{ 3.5 kb mRNA	>150		1	1	"
456	c-bcr, hu	>534	gccggccgcccATGg	1	1	Adams '87 EMBO 6, 115
457	c-erb-A, hu	>300	acccccaacagTATGa	4	2	Evans '86 Nature 324, 641
458	c-erb-A, ch	>288 *	gaattgcggtgaaATGg	2	1	Sap '86 Nature 324, 635
459	c-fes/fps, fe	*	gcggacgggcactATGg			Roebroek '87 JV 61, 2009
460	c-fms, hu	>300	cccaccgagggccATGg			Coussens '86 Nature 320, 277
461	c-hat, hu	>238	ctctgggcccggATGt	1		Taira '87 PNAS 84, 2980
462	c-int-1, mu	184 +	gccaggcagggccATGg		1	Varmus '85 MCB 5, 3337
463	c-int-2, mu	>326	cgcgatgccgggATGg	1		Moore '86 EMBO 5, 919
464	pp56-LSTRA, mu	>193	ccgggaggggatcATGg			Sefton '86 Nature 319,682
465	c-lyn, hu	297	cgagcgggaaatATGg			Yamanashi '87 MCB 7, 237
466	{ c-mos, mu, ovarian mRNA	~70	cttgaggggtgtaATGc	2		Propst '87 MCB 7, 1629
	{ testicular mRNA	280		4		
467	c-myb, hu	>113	gcccgccgcccATGg			Majello '86 PNAS 83, 9636
468	c-myb, mu	200 to 680	gcccgccctggccATGg	+		Watson '87 EMBO 6, 1643
469	c-myc, hu	400, 570*	ctcccgccgacgATGc		+	Saito '83 PNAS 80, 7476
470	c-myc, fe	400, 587*	gcaggcgccgagATGc			Stewart '86 Virol 154,121
471	c-neu, hu (HER2)	178 +	gcagtgagcaccATGg		1	Tal '87 MCB 7, 2597
472	c-pim-1, mu	~400	ctggaggtggggATGc			Selton '86 Cell 46, 603
473	c-raf-1, hu	>129 *	taagctgcatcaATGg	1		Bonner '86 NAR 14, 1009
474	A-raf-1, hu	>194	atctaaggctccATGg	1	1	Beck '87 NAR 15, 595
475	c-ral, simian		ctgtgacacgagATGg			Chardin '86 EMBO 5, 2203
476	c-Ha-ras-1, hu	69 to 332 *	ccctgaggagcgATGa	(1) (2)		Honkawa '87 MCB 7, 2933
477	c-Ha-ras-1, rat	~175 *	ccctgagaagcgATGa	1		Damante '87 PNAS 84, 774
478	c-Ki-ras, mu	200 to 250 *	ggcctgctgaaaATGa			Hoffman '87 MCB 7, 2592
479	N-ras, hu	~245	rtctgggtgtgaaATGa	2		Hall '85 NAR 13, 5255
480	R-ras, hu	65	agcgggtggcgacATGa			Lowe '87 Cell 48, 137
481	rho (ras-related)	>159	gttgcttgagcagATGg			Yeremian '87 NAR 15, 1869
482	c-sis, hu (PDGF2)	1022	cccggagtcggcATGa	2	1	Rao '86 PNAS 83, 2392
483	c-src, ch	>100 *	cagcccaaccaccATGg			Takeya '83 Cell 32, 881
484	c-src, Xp	>58	caacaggacaagATGg	1		Steele '85 NAR 13, 1747
485	c-syk, hu ("c-slk")	>589	ggaatttagataATGg	1	2	Seemba '86 PNAS 83, 5459
486	c-yes, hu	>162	gcagatttgataATGg			Sukegawa '87 MCB 7, 41
487	p53, hu	138, 230*	cggttcactgccATGg	(2)	(1)	Lamb '86 MCB 6, 1379
488	preproOpiomelanocortin, bo	129 *	cctgcctggaagATGc			Nakanishi '81 EJB 115,429
489	preproOpiomelanocortin, Xp	62 *	tccagtcctgaaATGt			Martens '87 EJB 165, 467
490	Ornithine ATase, hu	>54	ttgaaggacacaATGt			Inana '86 PNAS 83, 1203
491	Ornithine ATase, rat	>50	aggaccacacacaATGc			Mueckler '85 JBC 260, 12993
492	Ornithine decarboxylase	>300	acatcgagaaaccATGa	1		Gupta '85 JBC 260, 2941
493	OTCase, mu	136	agcaaaaagaagATGc			Veres '86 JBC 261, 7588
494	Ovalbumin, ch	64	tcagagttcaccATGg			McReynolds '78 Nat 273, 723
495	Ovoinhibitor, ch		agggtgctctgccATGa			Scott '87 JBC 262, 5899
496	ovomucoid, ch	53	cagtacctcaccATGg			Catterall '80 JCellB 87,480
497	3-Oxoacyl-CoA thiolase, rat	>100	ctgagcttcgtcATGg	7		Arakawa '87 EMBO 6, 1361
498	preproOxytocin, bo	33	cgcgctctgcaccATGg			Ruppert '84 Nature 308, 554
499	preproOxytocin, rat	40	aacaccaacgccATGg			Ivell '84 PNAS 81, 2006
500	Pancreatic polypeptide, hu	>50 *	tctggactccggATGg			Leiter '85 JBC 260, 13013
501	Parathyroid hormone, hu	>70 *	ttgtatgtgaaATGa	1		Vasicek '83 PNAS 80, 2127
502	Parotid secretory protein mu	55 *	agcaaaccaagATGt			Poulsen '86 EMBO 5, 1891
503	Pepsinogen, hu	54	ccgggaagaaccATGa			Sogawa '83 JBC 258, 5306
504	Pepsinogen, rat	>60	caaaccggcatcATGa			Ichihara '86 EJB 161, 7
505	Peroxi. enoyl-CoA hydratase	24	taccttgagaaaATGg			Ishii '87 JBC 262, 8144

No.			w/t	s/t	
506	Phenylalanine hydroxylase	>222	cggggagccagcATGt		Kwok '85 Biochem 24, 556
507	Phosphate carrier prot. bo	>62	cttagggagaagATGt		Runswick '87 EMBO 6, 1367
508	Phosphodiesterase, cyclic	>59	ttcttcgcgaaaATGt		Kurihara '87 JBC 262, 3256
509	PEP carboxykinase, rat	143 *	accattgcaagaATGc		Beale '85 JBC 260, 10748
510	PEP carboxykinase, ch	166, 246 *	gcagctgcagtaATGg	(1) (1)	Cook '86 PNAS 83, 7583
511	PGK-1 (X-linked), hu	94	tgtatttccaaaATGt		Riggs '84 Gene 32, 409
512	PGK-2, mu testicular	>20	cataccatcaagATGg		Boer '87 MCB 7, 3107
513	PGK, mu X-linked		ggtcttgccaaaATGt		Mori '86 Gene 45, 275
514	γ-Phosphor-kinase, mu		atccacgtgaccATGa		Caskey '87 PNAS 84, 2886
515	Phosphorylase, purine, hu	>109	gtctgcgagaccATGg		Williams '84 NAR 12, 5779
516	Pituitary hormones, α, hu	100 *	gaaaggagcgccATGg		Fiddes '81 JMAG 1, 3
517	" : α, mu	100	tgcaagaagctATGg		Chin '81 PNAS 78, 5329
518	Plasma cell glycoprotein	>111	cagagcggggcgATGg		vanDriel '87 JBC 262, 4882
519	u-Plasminogen activator hu	119	gacctcgccaccATGa		Riccio '85 NAR 13, 2759
520	PDGF, A-chain, hu	>387	cctcgggacgcgATGa	1 1	Betscholtz '86 Nat 320, 695
521	Platelet factor 4, rat	73	caccctcttgacATGa		Doi '87 MCB 7, 898
522 ^m	Polymerase-β (DNA), rat	51	gtccccgcaccATGa		Yamaguchi '87 MCB 7, 2012
523	Polymerase II (RNA), mu	406	gcctgcctcgccATGc		Ahearn '87 JBC 262, 10695
524	Poly(A) binding protein, hu	>502	agccgtgccgagATGa		Grange '87 NAR 15, 4771
525	Porphobilinogen deaminase	>83	aacagcccaagATGa		Raich '86 NAR 14, 5955
526	Prealbumin, hu	26	attcttggcaggATGg		Sasaki '85 Gene 37, 191
<u>Prolactin growth hormone family:</u>					
527	prePlacental lactogen, hu	62	cacctagtggcaATGg		Saunders '83 JBC 258, 3787
528	prePlacental lactogen, mu	>59	aactcctcagagATGa		Jackson '86 PNAS 83, 8496
529	Proliferin, mu	68	gactctgcagagATGc		Linzner '87 EMBO 6, 2281
530	Prolactin, hu	57	agatcacgaaATGg		Truong '84 EMBO 3, 429
531	Prolactin, rat	>51	gtgggtcatcaccATGa		Cooke '80 JBC 255, 6502
532	Prolactin, bo	>67	atcatcaccaccATGg		Sawavage '82 JBC 257, 678
533	Proline-rich (acidic) ha	33	gcctcctccaagATGc		Ann '87 JBC 262, 3958
534	" (glycosylated)	>34	gcctccagcgagATGc		Maeda '85 JBC 260, 11123
535	Prostatic BP, C2 chain, rat	41	aaactgagcaccATGa		Delaey '87 NAR 15, 1627
536	Protamine 1, mu	92	caagccagcaccATGg		Peachon '87 PNAS 84, 5316
537	Protamine, trout	14	ccatcaatcacaATGc		Gregory '82 NAR 10, 7581
Proteases: see also 108, 109, 114, 126, 363, 503-4, 519					
538	-batroxobin, snake	179	agagttgaagctATGg		Itoh '87 JBC 262, 3132
539	-ser protease, mu adipose	19	cctgctgtcagaATGc		Min '86 NAR 14, 8879
540	-Ca ²⁺ protease, hu	105-155 *	tgagtcgcagccATGt		Miyake '86 NAR 14, 8805
541	-Ca ²⁺ protease, ra	>150	tgagccgcagccATGt	2	Emori '86 JBC 261, 9472
542	-Ca ²⁺ protease, ch	37	cagtacgcagctATGa		Ohno '84 Nature 312, 566
543	-cys protease, mu	>59	ggtgtttgaaccATGa		Portnoy '86 JBC 261, 14697
544	-ser protease, EGF binding		acacctgttaccATGt		Lundgren '84 JBC 259, 7780
545	-mast cell protease, rat	35	accactggcacaATGc		Benfey '87 JBC 262, 5377
546	-ser p'ase, cytotoxic T cells	>111	cttcgggggaagATGa		Brunet '86 Nature 322, 268
547	PDI (disulphide isomerase), rat		ccgacgtccgacATGc		Edman '85 Nature 317, 267
548 ^m	Proteoglycan 19 (chondroitin)		gagctggtcaggATGc		Bourdon '86 JBC 261, 12534
549	" 38K core protein, hu	>91	atgagataaatcATGa	1	Krusius '86 PNAS 83, 7683
550	Proteolipid protein, mu	162	agtccaaagacATGg		Hudson '87 PNAS 84, 1454
551	Pulmonary surfactant, hu	53	ggacccagagccATGt		White '85 Nature 317, 361
552	Pyruvate kinase, ch	80 *	actccagtaaccATGt		Lonberg '83 PNAS 80, 3661
553	Quinone reductase, rat	>113	ttcaactatgccATGa	1 1	Bayney '87 JBC 262, 572

No.	w w/t s s/t		
554	snRNP-B* antigen (U2), hu	>125	tttaacacaaacATGg Habets '87 PNAS 84, 2421
555	RNP-C protein, hu	>122	ccatcaaaacacgATGg Swanson '87 MCB 7, 1731
556	U1-RNA-associated 70K, hu	>680	ggcgagacgaagATGg 5 4 Theissen '86 EMBO 5, 3209
Receptors:			
557	β_2 -adrenergic, hu	190	agactgcgcgccATGg 1 Kobilka '87 PNAS 84, 46
558	β -adrenergic, turkey	69	cgccccgcagccATGg Yarden '86 PNAS 83, 6795
559	-for asialo-GP, L1 chain	65	cccagtgctatcATGg Leung '85 JBC 260, 12523
560	-for " L2 chain, rat	>153	cctagggccatcATGg McPhaul '87 MCB 7, 1841
561	-for EGF, hu (HER1)	100	cggggagcagcgATGg Ishii '85 PNAS 82, 4920
562	-for estrogen, hu	232	cgggcaccggaccATGg 1 Green '86 Nature 320, 134
563	-GPIIIa (rel. to integrin), hu		gaggcgagcagATGg Fitzgerald '87 JBC 262, 3936
564	-for IgE, hu	213	agcaggaccgccATGg Ikuta '87 PNAS 84, 819
565	-for IgA & IgM (epithel)	>123	cagccaccagccATGg Mostov '84 Nature 308, 37
566	-for insulin, hu		gtccccgcagccATGg 1 Ullrich '85 Nature 313, 756
567	-for insulin-like GF-I, hu		caataaaaggaATGg Ullrich '86 EMBO 5, 2503
568	-for interleukin-2(a), hu	159, 217	agggtcaggaagATGg 2 Leonard '85 Sci 230, 633
569	-for LDL, hu	>80	gaggctgcgagcATGg Sudhof '85 Sci 228, 815
570	-for nerve growth fact. hu	>113	gggagggcgggcATGg Johnson '86 Cell 47, 545
571	-for PDGF, mu	>138	agcccgagacccATGg Yarden '86 Nature 323, 226
572	-for progesterone, ra	>125	gttcaggctgacATGg Loosfelt '86 PNAS 83, 9045
573	-for SRP		cctgctgcccATGc Lauffer '85 Nature 318, 334
574	preproRelaxin, rat	>60	ccccagaccggaATGt Hudson '81 Nature 291, 127
575	preproRenin, hu	44	actgagggagcATGg Fukamizu '86 Gene 49, 139
576	Retinol-Binding protein, hu	86 *	ttcctgggcaagATGc Cortese '85 EMBO 4, 1981
577	Retinol-BP, rat	>94	tctgtccccaaaATGc Sherman '87 PNAS 84, 3209
578	Retinol-BP-II, rat	>55	gagggcccatcATGc Li '86 PNAS 83, 5779
579	Retinol-BP, Xp	>40	ttgtgaagaagATGg McKearin '87 JBC 262, 4939
580	Rhodopsin, bo	96	agggccgcagccAUGa Nathans '83 Cell 34, 807
581	Ribonuclease, panc. rat	75	agcaaaagccatcATGg MacDonald '82 JBC 257, 14582
582	Ribonucl. reduct. M1, mu	>242	cttctagcggcgATGc Caras '85 JBC 260, 7015
583	" " M2, mu	>62	ccctggttcgccATGc Thelander '86 MCB 6, 3433
584	Ribophorin II, hu	>268	ctgctcgaggagATGg 1 Crimmaudo '87 EMBO 6, 75
Ribosomal proteins:			
585	rp S14, hu	40 *	cgacgtgcagaaATGg Rhoads '86 MCB 6, 2774
586	rp S16, mu	52	gtgctcgagctATGc Wagner '85 MCB 5, 3560
587	rp S19, Xp	>46	atagccggcaagATGc Amaldi '82 Gene 17, 311
588	rp L1, Xp	40	agcagcgaggagATGg Loreni '85 EMBO 4, 3483
589	rp L14, Xp	39	acagccgccatcATGg Beccari '87 NAR 15, 1870
590	rp L27, mu		tctgcccaccgtATGc Belhumeur '87 NAR 15, 1019
591	rp L30, mu	32 *	taaggcgaggaagATGg Wiedemann '84 MCB 4, 2518
592	rp L31, rat	>25	ggggccggcagaATGg Tanaka '87 EJB 162, 45
593	rp L32, mu	51 *	tcaaaaggcatcATGg Dudov '84 Cell 37, 457
594	rp L44, hu	>83	cctgctgcaagATGg Davies '86 Gene 45, 183
595	RSV-induced 9E3 protein, ch	77	acactcctaaccATGc Sugano '87 Cell 49, 321
596	Scrapie PrP27-30, ha	90 *	agatcagccatcATGg Basler '86 Cell 46, 417
597	Secretogranin I, hu	>112	ccgagcggggccATGc Benedum '87 EMBO 6, 1203
598	Seminal vasc. prot IV, rat	22	ttttctggcagATGc Kandala '83 NAR 11, 3169
599	Sodium ch. prot I, rat	>251	caggatgacaagATGg 5 2 Noda '86 Nature 320, 188
600	Somatostatin-I, hu	105	cgcgccggcagATGc Shen '84 Sci 224, 168
601	" rat	100	gaggcagggagATGc Dixon '84 JBC 259, 11798
602	" -II, angfish	>59	ccagcagacagtATGc Hobart '80 Nature 288, 137
603	" -22, catfish		gctaccagaagATGt Dixon '82 PNAS 79, 5152
604	Sorcin/V19, ha		gtagtcttccaccATGg Borst '86 EMBO 5, 3201
605	SPARC, mu embry. endoderm	90	gttcccagcatcATGc Mason '86 EMBO 5, 1465
606	Stearyl CoA desaturase, rat	>102	ccgacagccacgATGc Thiede '86 JBC 261, 13230

No.		w w/t	s s/t
607	proSuccrase-isomaltase, ra	caatgaaataagATGg	1 Hunziker '86 Cell 46, 227
608	2-5A Synthetase, mu	tccagacttagcATGg	Ichii '86 NAR 14, 10117
609	Synthetase, his-tRNA, ha 77	ttggcagccaggATGg	Tsui '87 NAR 15, 3349
610	t complex protein 1, mu >60	cgtttctctgaagATGg	Willison '86 Cell 44, 727
<u>T-cell antigen receptor:</u>			
611	Ti α-chain, hu >170	cactgctcagccATGc	Yanagi '85 PNAS 82, 3430
612	Ti β-chain, hu 52	tctcactctgccATGg	Smith '87 NAR 15, 4991
613	Ti γ-chain (CD3), hu >140	agagagggaaggcATGc	Littman '87 Nature 326, 85
614	T- α-chain, mu >55	caaggctcagccATGc	Saito '84 Nature 312, 36
615	T- V _β chain (3H.25), mu	ttctaagccaccATGg	Goverman '85 Cell 40, 859
616	T- V _β chain (5.1), mu	ctgagagggaagcATGt	Chou '87 PNAS 84, 1992
617	T- V _γ chain, mu	ctacagcagaccATGa	Garman '86 Cell 45, 733
618	T3-antigen, γ-chain, hu	acagagactgacATGg	Krissansen '86 EMBO 5, 1799
619	T3-antigen, δ-chain, hu 95	ttccgctgcgagATGg	Tunnaclyffe '86 EMBO 5, 1245
620	T3-antigen, ε-chain, hu	catgaacacaagATGc	Gold '86 Nature 321, 431
<u>T-cell differentiation antigens:</u>			
621	CD2 (human T11)	ccaacccctaagATGa	Sayre '87 PNAS 84, 2941
622	CD4 (human T4) >75	ggcaaggccacaATGa	Maddon '85 Cell 42, 93
623	CD4 (rat) >53	aagcaggccaccATGt	Clark '87 PNAS 84, 1649
624	CD8 (human Leu-2/T8) >115	ggggagcgcgtcATGg	Sukhatme '85 Cell 40, 591
625	CD8, α-chain (Lyt-2, mu) 333	ggagagccaccATGg	Nakauchi '87 NAR 15, 4337
626	CD8, 37K-chain (rat)	aagagcgccaagATGc	Johnson '86 Nature 323, 74
<u>Other T-cell proteins:</u>			
627	cytotoxic PT49 protein, mu	cgactgcaggATGa	Koyama '87 PNAS 84, 1609
628	16K MAL protein, hu >55	cagcagccgcgtcATGg	Alonso '87 PNAS 84, 1997
629	preLy-6, mu 90	ccttctctgaggATGg	LeClair '86 EMBO 5, 3227
630	IgE binding factor (soluble) >93	gtaaagtggaaaATGg	Martens '85 PNAS 82, 2460
631	Tachykinin (neuromedinK) bo 144 *	tccacaggcatcATGc	Kotani '86 PNAS 83, 7074
632	Tachykinin (substnce P), rat 99 *	gcacaaatccaacATGa	Krause '87 PNAS 84, 881
633	Thrombospondin, hu >120	aacagctccaccATGg	Dixit '86 PNAS 83, 5449
634	Thy-1, mu 78 *	actcttggcaccATGa	Giguere '85 EMBO 4, 2017
635	Thy-1-related MRC OX-2, rat	cccagagcaaggATGg	Clark '85 EMBO 4, 113
636	Thymidine kinase, hu 60	cccgaggcgcaATGa	Kreidberg '86 MCB 6, 2903
637	" " ha	cgcacagccgcaATGa	Lewis '86 MCB 6, 1998
638	" " ch	agcggcgcgaaacATGa	Herrill '84 MCB 4, 1769
639	Thymidylate synthase, hu >90	gcccgcgcgcgcATGc	Takeishi '85 NAR 13, 2035
640	" " mu 24, 34, 51	gaactgtccgttATGc	Deng '86 JBC 261, 16000
641	proThymosin-α, hu >177	gcgtgcgccaccATGt	Berger '86 PNAS 83, 9403
642	Thymosin-β ₄ , rat	cttcacagcaaccATGt	Horecker '84 PNAS 81, 2295
643	Thyroglobulin, hu 41	agggccaggaaaaATGg	Christophe '85 NAR 13, 5127
644	Thyroglobulin, bo 41	aaggctcccaagATGg	Mercken '85 Nature 316, 647
645	Thyrotropin, β-subunit, mu >89	gttggttcaagcATGa	Gurr '83 PNAS 80, 2122
646	" -releasing hormone Xp>109	acagcaggaaagATGg	Richter '84 EMBO 3, 617
647	Thyroxine-binding globulin, hu	cttctctccaaaATGt	Flink '86 PNAS 83, 7708
648	Transferrin, hu 50	cgcaccgcgaagATGa	Lucero '86 NAR 14, 8692
649	TGF-α, hu	ccgcgccgtaaaaATGg	Derynck '84 Cell 38, 287
650	TGF-β1, hu >841	gcgcgctccccccATGc	Derynck '85 Nature 316, 701
651	Transin, rat	aagccagtggaaATGa	Breathnach '87 NAR 15, 1139
652	Transin-2, rat	aaggctgtctctATGg	"
<u>Translation factors:</u>			
653	Elongation factor 1α, hu >53	ctaaaagccaaaATGg	Brands '86 EJB 155, 167
654	Elongation factor 2, ha >77	ccatccgcactATGg	Kohno '86 PNAS 83, 4978
655	eIF2, rat	atacacttcagaATGc	Ernst '87 JBC 262, 1206
656	cap binding protein, hu 18	gatcgatcttaagATGg	Rychlik '87 PNAS 84, 945

No.			w w/t	s s/t	
657	Transcription fact. TFIIIA	50	gctgaaggagagATGg		Tao '86 NAR 14, 2187
658	Triose-P-isomerase, hu	34	ctcggctcgccATGg		Brown '85 MCB 5, 1694
659	" ch	52	gtcgctcggccATGg		Straus '85 MCB 5, 3497
660	Tropomyosin, hu fibroblasts	118	ccaccgcaggccATGg		MacLeod '85 PNAS 82, 7835
661	TM30, pl	>50	gcgctccgcccATGg		MacLeod '87 JMB 194, 1
662	TM-1, rat fibroblasts	>61	cccaccgcagccATGg		Helfman '85 JBC 260, 14440
663	α -Tropomyosin, rat muscle	78	gccaccgcaccATGg		Ruiz-Opazo '87 JBC 262, 4755
664	Troponin-I, fast muscle	82 *	atctaaagcaagATGt		Baldwin '85 PNAS 82, 8080
665	Troponin-T, rat	79 *	cccaccgccactATGt		Breitbart '86 JMB 188, 313
666	Troponin-C, slow muscle, ch		ccctgcccggccATGg		Putkey '87 MCB 7, 1549
667	Trypsinogen, anionic, ca	14	acttctgccatcATGa		Pineky '85 MCB 5, 2669
668	" cationic, ca	29	caggagcaaccATGa		
669	α -Tubulin I, CHO	>100	cccgtagctaccATGc		Elliott '86 MCB 6, 906
670	α -Tubulin II, CHO	>100	aaagcagcaaccATGc		"
671	α -Tubulin III, CHO	>130	ttctctagaccaccATGc		"
672 ⁿ	α -Tubulin, hu	211	ttccgggaaaacATGc	1	Hall '85 NAR 13, 207
673	β -Tubulin, hu	72	gcgcgcgccatcATGa		Lewis '85 JMB 182, 11
674	β -Tubulin, hu	159	taaatcttaaccATGa		Lee '83 Cell 33, 477
675	β -Tubulin, ch	>87	gacaccggccatcATGc		Cleveland '81 Nat 289, 650
676 ^o	β 3-Tubulin, ch	40 to 50	gccgaagccatcATGa	+	Sullivan '86 JBC 261, 13317
677	β 4-Tubulin, ch	74	ttccggccgcaccATGa		Sullivan '86 MCB 6, 4409
678	β 5-Tubulin, ch	48	cgggacagcgccATGa		"
679	Tyrosinase, mu	>174	gcttcgagaagaATGa	2	Shibahara '86 NAR 14, 2413
680	Tyr aminotransferase, rat	97 *	gcttcgagagggATGg		Grange '85 JMB 184, 347
681	Tyr hydroxylase, hu	>30	ccacactgagccATGc		Grima '87 Nature 326, 707
682	Tyr hydroxylase, rat	35	ccagcttgccactATGc		Harrington '87 NAR 15, 2363
683	Ubiquitin, hu	100 *	taacaggtaaaaATGc		Baker '87 NAR 15, 443
684	Ubiquitin, ch	63 *	ggagacgttaaacATGc		Bond '86 MCB 6, 4602
685	UDP-glucuronosyl-tr'ase, hu		cattgcatcaggATGt		Jackson '87 BiochJ. 242, 581
686	" " steroid-induced, rat>75		cttgatttttaagATGc		Harding '87 NAR 15, 3936
687	" " 3-MC induced, rat >124		ctctctgaaaggATGg	1	Iyanagi '86 JBC 261, 15607
688	Uncoupling prot.(brown fat)177		ctccgagccaaagATGg		Ricquier '86 JBC 261, 1487
689	major Urinary pr(MUP), mu	60	ctccctaccaaaATGa		Shahan '87 MCB 7, 1938
690	Uroporphyrinogen decarb. hu		agacagctgaccATGg		Goossens '86 JBC 261, 9825
691	Urotensin-I, carp	>75	cctgtgtccagcATGa		Ishida '86 PNAS 83, 308
692	Uteroglobin, ra	47	cattctgcccaccATGa		Suske '83 NAR 11, 2257
693	Valosin "precursor", po	>143	gaagcgcgccccATGg		Koller '87 Nature 325, 542
694	Vasoactive intest. peptide	174 *	agaggcacagaaATGg		Linder '87 PNAS 84, 605
695	Vasopressin-neurophysinII	48	accgctgccaggATGc		Ruppert '84 Nature 308, 554
696	Vimentin, ha	135	gctctccaaaccATGt		Quax '83 Cell 35, 215
697	Vinculin, ch	>246	ccgcctgcggccATGc		Price '87 Bioch.J. 245, 595
698	Vitellogenin-II, ch	13	ttcaccttcgctATGa		Geiser '83 JBC 258, 9024
699	Vitellogenin, Xp	13	ttcgccatcaccATGa		Walker '83 EMBO 2, 2271
700	preWhey acidic protein, rat	33	gcgcgcgacaccATGc		Campbell '84 NAR 12, 8685
701 ^p	preXenopsin, Xp	>62	catttggaaggATGt		Sures '84 PNAS 81, 380

^a A two-letter abbreviation indicates the source of each mRNA: bovine; canine; chicken; feline; gp, guinea pig; hamster; human; murine; porcine; rabbit; sheep; Xp, Xenopus.

^b The number of nucleotides comprising the 5'-untranslated sequence is indicated. In most cases this was determined by primer extension and/or S1 mapping. An entry is marked > if the cDNA included a considerable portion of the 5'-noncoding

sequence but was probably not complete. There is no entry in this column if the cDNA included little of the leader sequence or if the 5'-end of the mRNA was not mapped on the genomic sequence. If a second form of mRNA was detected but its leader sequence not precisely mapped, the entry is marked +. An asterisk indicates that the 5'-noncoding sequence is interrupted by an intron.

- ^cUpstream ATG codons are designated strong (s) or weak (w) according to context (see text). They are listed according to whether or not the reading frame established by the upstream ATG codon terminates (t) before the start of the major open reading frame. If a gene produces two major transcripts, only one of which has upstream ATG codons, the upstream ATG codons are listed in parentheses. [±] means that upstream ATG codons occur in only a minor species of mRNA. If an upstream ATG codon lies very near the error-prone 5'-end of a cDNA clone and has not yet been confirmed by sequencing either the gene or a second cDNA, I have temporarily entered a question mark in this column.
- ^dBibliographic data are given in condensed form: first author, year, journal, volume, and first page.
- ^eAlternative splicing produces transcripts with three different leader sequences, only one of which is listed. The second form of mRNA also has three upstream ATG codons, while the third has a single weak ATG codon upstream.
- ^fEight of the upstream ATG codons lie in the same reading frame, constituting an upstream cistron with the potential to encode a 40 amino acid peptide. In view of the pattern of codon usage, Shull and Lingrel postulate that this peptide is made.
- ^gIt is likely that ribosomes initiate at the first and second ATG codons in these mRNAs, producing long and short forms of the encoded polypeptide. In each case the 5'-proximal ATG codon occurs unusually close to the cap and in a sub-optimal context for initiation; it is not known which of those features accounts for the "leakiness." The distance from the cap to the first ATG codon is 5 nucleotides for #143, 7 n for #297 and 3 n for #330.
- ^hUnlike the human HMG-CoA reductase mRNA sequence which is entered in the table, a subset of transcripts from the corresponding hamster gene have upstream ATG codons.
- ⁱThe sequence of the chicken insulin gene has a weak ATG codon upstream from the translational start site, but it is not known whether the 5'-end of the transcript includes that ATG codon.
- ^jThere is a developmentally regulated switch in the promoter for IGF-II in rats. The longer transcript introduces an upstream in-frame ATG codon, and initiation at that site would add eleven amino acids to the N-terminus of IGF-II. Only the shorter transcript is represented in the table because the functional initiator codon in the longer transcript has not been verified experimentally.
- ^k α -Interferons A and D are the most highly expressed in humans and both of those mRNAs have A in position -3, as shown. Other human α -IFN genes that are expressed at lower levels have C in position -3, but it is not known if that substitution accounts for their lower expression.
- ^lIn the major form of NGF mRNA in mouse submaxillary glands, the indicated initiator codon is the first ATG triplet in the message. The exon that carries that ATG codon is spliced out of the major NGF transcript in other tissues and an ATG codon that lies farther downstream is thereby activated.
- ^mThe first cDNA that was cloned fell short of the real initiator codon, resulting in misidentification of an internal ATG codon as the translational start site.

ⁿWhereas the context around the initiator codon is standard in the α - and β -tubulin sequences shown in the table, one α - and one β -tubulin mRNA have been described in which the ATG codon lies in a poor context for initiation (Cowan '83 MCB 3,1738; Lee '84 NAR 12, 5823). It is not known whether or how well those particular mRNA species are translated.

^oThe chicken β 3-tubulin gene produces mRNAs with very heterogeneous 5'-ends, some of which would lack the upstream ATG codon.

^pEntries 363 and 401 have been deleted, leaving 699 sequences on which the calculations in Tables 1 and 2 are based. This includes all published sequences to which I had access as of May 31, 1987, in which the functional initiator codon has been clearly identified. Another 110 sequences were excluded because of uncertainty about which ATG codon initiates translation; that list was made available to the editor during the review of this manuscript.