

COL3A1 Systematic Numbering of cDNA and Protein Sequences

The reference cDNA used here is based on GenBank RefSeq NM_000090.3. The corresponding reference genomic DNA sequence is GenBank RefSeqGene NG_007404.1.

The DNA sequence is numbered from the first base of the translation initiation codon and the amino acid sequence is numbered from the first amino acid of the signal peptide. As an aid, the amino acid numbering in the triple helical region is also presented in the “historic” fashion with the first glycine of the triple helix numbered as “1”.

The type III collagen signal peptide has “historically” been 24 amino acids in length, but is presented here as 23 amino acids long. This shorter length is consistent with the feature table in NM_000090.3 and with independent analysis of the predicted cleavage point for signal peptides.

The location of the cleavage point for the C-propeptide is incorrect in the feature tables in NM_000090.3 and in RefSeq NP_000081.1. It is correct in the corresponding UniProtKB/Swiss-Prot entry P02461. The location of the cleavage point is correct in this document.

Version 1.1: August 2008

5' Untranslated Region

```

                                     GGCTGAG TTTTATGACG
                                     -117  ----- +----- -101

-100  GGCCCCGGTGC TGAAGGGCAG GGAACAACCTT GATGGTGCTA CTTTGAACCTG
      +-----+ +-----+ +-----+ +-----+ +-----+
-50   CTTTTCTTTT CTCCTTTTTG CACAAAGAGT CTCATGTCTG ATATTTAGAC
      +-----+ +-----+ +-----+ +-----+ +-----+
      -50                                     -1
  
```

Signal Peptide

```

1  ATGATGAGCTTTGTGCAAAAGGGGAGC TGGCTACTTCTCGCTCTGCTTCATCCC 54
   -----+-----+----- -+-----+-----+-----
1  MetMetSerPheValGlnLysGlySer TrpLeuLeuLeuAlaLeuLeuHisPro 18
   M M S F V Q K G S W L L L A L L H P

55  ACTATTATTTTGGCA 69
   -----+-----
19  ThrIleIleLeuAla 23
   T I I L A
  
```

N-terminal Propeptide

```

1_79 exon1 | exon2 80_282
70  CAACAGGAAGCTGTTGAAGGAGGATGT TCCCATCTTGGTCAGTCCTATGCCGAT 123
   -----+-----+----- -+-----+-----+-----
24  GlnGlnGluAlaValGluGlyGlyCys SerHisLeuGlyGlnSerTyrAlaAsp 41
   Q Q E A V E G G C S H L G Q S Y A D
  
```

124	AGAGATGTCTGGAAGCCAGAACCATGC	CAAATATGTGTCTGTGACTCAGGATCC	177
	-----+-----+-----+	-----+-----+-----+	
42	ArgAspValTrpLysProGluProCys	GlnIleCysValCysAspSerGlySer	59
	R D V W K P E P C	Q I C V C D S G S	
178	GTTCTCTGCGATGACATAATATGTGAC	GATCAAGAATTAGACTGCCCCAACCCA	231
	--+-----+-----+-----	-----+-----+-----+	
60	ValLeuCysAspAspIleIleCysAsp	AspGlnGluLeuAspCysProAsnPro	77
	V L C D D I I C D	D Q E L D C P N P	
		80_282 exon2	exon3 283_333
232	GAAATTCATTTGGAGAATGTTGTGCA	GTTTGCCACAGCCTCCAACCTGCTCCT	285
	-----+-----+-----	-+-----+-----+-----	
78	GluIleProPheGlyGluCysCysAla	ValCysProGlnProProThrAlaPro	95
	E I P F G E C C A	V C P Q P P T A P	
		283_333 exon3	exon4/5 334_447
286	ACTCGCCCTCCTAATGGTCAAGGACCT	CAAGGCCCAAGGGAGATCCAGGCCCT	339
	-----+-----+-----	-----+-----+-----	
96	ThrArgProProAsnGlyGlnGlyPro	GlnGlyProLysGlyAspProGlyPro	113
	T R P P N G Q G P	Q G P K G D P G P	
340	CCTGGTATTCTGGGAGAAATGGTGAC	CCTGGTATTCCAGGACAACCAGGGTCC	393
	+-----+-----+-----	---+-----+-----+---	
114	ProGlyIleProGlyArgAsnGlyAsp	ProGlyIleProGlyGlnProGlySer	131
	P G I P G R N G D	P G I P G Q P G S	
		334_447 exon4/5	exon4/5 334_447
394	CCTGGTTCTCCTGGCCCCCCTGGAATC	TGTGAATCATGCCCTACTGGTCTCAG	447
	-----+-----+-----	-----+-----+-----	
132	ProGlySerProGlyProProGlyIle	CysGluSerCysProThrGlyProGln	149
	P G S P G P P G I	C E S C P T G P Q	
	exon 6 448_528		
448	AACTATTCTCCC	459	
	--+-----		
150	AsnTyrSerPro	153	
	N Y S P		

N-terminal Telopeptide

460	CAGTATGATTCATATGATGTCAAGTCT	GGAGTAGCAGTAGGA	501
	-----+-----+-----	---+-----+---	
154	GlnTyrAspSerTyrAspValLysSer	GlyValAlaValGly	167
	Q Y D S Y D V K S	G V A V G	

Triple Helix

	448_528 exon6	exon7 529_582	
502	GGACTCGCAGGCTATCCTGGACCAGCT	GGCCCCCAGGCCCTCCCGTCCCCCT	555
	-----+-----+-----	-+-----+-----+-----	
168	GlyLeuAlaGlyTyrProGlyProAla	GlyProProGlyProProGlyProPro	185
1	G L A G Y P G P A	G P P G P P G P P	18

	529_582 exon7	exon8 583_636	
556	GGTACATCTGGTCATCTGGTTCCCCT	GGATCTCCAGGATACCAAGGACCCCCT	609
	----+-----+-----+--	-----+-----+-----	
186	GlyThrSerGlyHisProGlySerPro	GlySerProGlyTyrGlnGlyProPro	203
19	G T S G H P G S P	G S P G Y Q G P P	36
	583_636 exon8	exon9 637_690	
610	GGTGAACCTGGGCAAGCTGGTCCTTCA	GGCCCTCCAGGACCTCCTGGTGCTATA	663
	+-----+-----+-----	---+-----+-----+---	
204	GlyGluProGlyGlnAlaGlyProSer	GlyProProGlyProProGlyAlaIle	221
37	G E P G Q A G P S	G P P G P P G A I	54
	637_690 exon9	exon10 691_744	
664	GGTCCATCTGGTCCTGCTGGAAAAGAT	GGAGAATCAGGTAGACCCGGACGACCT	717
	-----+-----+-----+--	-----+-----+-----	
222	GlyProSerGlyProAlaGlyLysAsp	GlyGluSerGlyArgProGlyArgPro	239
55	G P S G P A G K D	G E S G R P G R P	72
	691_744 exon10	exon11 745_798	
718	GGAGAGCGAGGATTGCCTGGACCTCCA	GGTATCAAAGGTCCAGCTGGGATACCT	771
	--+-----+-----+-----+--	-----+-----+-----+---	
240	GlyGluArgGlyLeuProGlyProPro	GlyIleLysGlyProAlaGlyIlePro	257
73	G E R G L P G P P	G I K G P A G I P	90
	745_798 exon11	exon12 799_852	
772	GGATTCCCTGGTATGAAAGGACACAGA	GGCTTCGATGGACGAAATGGAGAAAAG	825
	-----+-----+-----+---	-+-----+-----+-----	
258	GlyPheProGlyMetLysGlyHisArg	GlyPheAspGlyArgAsnGlyGluLys	275
91	G F P G M K G H R	G F D G R N G E K	108
	799_852 exon12	exon13 853_897	
826	GGTGAAACAGGTGCTCCTGGATTAAG	GGTGAAAATGGTCTTCCAGGCGAAAAT	879
	----+-----+-----+---	-----+-----+-----+---	
276	GlyGluThrGlyAlaProGlyLeuLys	GlyGluAsnGlyLeuProGlyGluAsn	293
109	G E T G A P G L K	G E N G L P G E N	126
	853_897 exon13	exon14 898_951	
880	GGAGCTCCTGGACCCATGGGTCCAAGA	GGGGCTCCTGGTGAGCGAGGACGGCCA	933
	+-----+-----+-----+---	---+-----+-----+-----	
294	GlyAlaProGlyProMetGlyProArg	GlyAlaProGlyGluArgGlyArgPro	311
127	G A P G P M G P R	G A P G E R G R P	144
	898_951 exon14	exon15 952_996	
934	GGACTTCCTGGGGCTGCAGGTGCTCGG	GGTAATGACGGTGCTCGAGGCAGTGAT	987
	-----+-----+-----+---	-----+-----+-----+---	
312	GlyLeuProGlyAlaAlaGlyAlaArg	GlyAsnAspGlyAlaArgGlySerAsp	329
145	G L P G A A G A R	G N D G A R G S D	162
	952_996 exon15	exon16 997_1050	
988	GGTCAACCAGGCCCTCCTGGTCCTCCT	GGAAGTCCCGGATTCCTGGATCCCCT	1041
	--+-----+-----+-----+---	-----+-----+-----+---	
330	GlyGlnProGlyProProGlyProPro	GlyThrAlaGlyPheProGlySerPro	347
163	G Q P G P P G P P	G T A G F P G S P	180

997_1050 exon16 | **exon17 1051_1149**

1042 GGTGCTAAGGGTGAAGTTGGACCTGCA GGGTCTCCTGGTTCAAATGGTGCCCT 1095
 -----+-----+----- -+-----+-----+-----
 348 GlyAlaLysGlyGluValGlyProAla GlySerProGlySerAsnGlyAlaPro 365
 181 G A K G E V G P A G S P G S N G A P 198

1051_1149 exon17

1096 GGACAAAGAGGAGAACCTGGACCTCAG GGACACGCTGGTGCTCAAGGTCCCTCCT 1149
 ----+-----+-----+--- -----+-----+-----
 366 GlyGlnArgGlyGluProGlyProGln GlyHisAlaGlyAlaGlnGlyProPro 383
 199 G Q R G E P G P Q G H A G A Q G P P 216

exon18 1150_1194 | **1150_1194 exon18** | **exon19 1194_1293**

1150 GGCCCTCCTGGGATTAATGGTAGTCCT GGTGGTAAAGCGAAATGGGTCCCGCT 1203
 +-----+-----+----- -+-----+-----+-----
 384 GlyProProGlyIleAsnGlySerPro GlyGlyLysGlyGluMetGlyProAla 401
 217 G P P G I N G S P G G K G E M G P A 234

1204 GGCATTCTGGAGCTCCTGGACTGATG GGAGCCCGGGTCTCCAGGACCAGCC 1257
 -----+-----+-----+----- -+-----+-----+-----
 402 GlyIleProGlyAlaProGlyLeuMet GlyAlaArgGlyProProGlyProAla 419
 235 G I P G A P G L M G A R G P P G P A 252

1194_1293 exon19 | **exon20 1294_1347**

1258 GGTGCTAATGGTGCTCCTGGACTGCGA GGTGGTGAGGTGAGCTGGTAAGAAT 1311
 --+-----+-----+----- -+-----+-----+-----
 420 GlyAlaAsnGlyAlaProGlyLeuArg GlyGlyAlaGlyGluProGlyLysAsn 437
 253 G A N G A P G L R G G A G E P G K N 270

1294_1347 exon20 | **exon21 1348_1455**

1312 GGTGCCAAAGGAGAGCCCGGACCACGT GGTGAACCGGTTGAGGCTGGTATTTCCA 1365
 -----+-----+----- -+-----+-----+-----
 438 GlyAlaLysGlyGluProGlyProArg GlyGluArgGlyGluAlaGlyIlePro 455
 271 G A K G E P G P R G E R G E A G I P 288

1366 GGTGTTCCAGGAGCTAAAGGCGAAGAT GGCAAGGATGGATCACCTGGAGAACCT 1419
 ----+-----+-----+--- -----+-----+-----
 456 GlyValProGlyAlaLysGlyGluAsp GlyLysAspGlySerProGlyGluPro 473
 289 G V P G A K G E D G K D G S P G E P 306

1348_1455 exon21 | **exon22 1456_1509**

1420 GGTGCAAATGGGCTTCCAGGAGCTGCA GGAGAAAAGGGTGCCCTGGGTTCGGA 1473
 +-----+-----+----- -+-----+-----+-----
 474 GlyAlaAsnGlyLeuProGlyAlaAla GlyGluArgGlyAlaProGlyPheArg 491
 307 G A N G L P G A A G E R G A P G F R 324

1456_1509 exon22 | **exon23 1510_1608**

1474 GGACCTGCTGGACCAAATGGCATCCCA GGAGAAAAGGGTCTGCTGGAGAGCGT 1527
 -----+-----+-----+----- -+-----+-----+-----
 492 GlyProAlaGlyProAsnGlyIlePro GlyGluLysGlyProAlaGlyGluArg 509
 325 G P A G P N G I P G E K G P A G E R 342

1528	GGTGCTCCAGGCCCTGCAGGGCCCAGA	GGAGCTGCTGGAGAACCTGGCAGAGAT	1581
	--+-----+-----+-----	-----+-----+-----+---	
510	GlyAlaProGlyProAlaGlyProArg	GlyAlaAlaGlyGluProGlyArgAsp	527
343	G A P G P A G P R	G A A G E P G R D	360
1510_1608 exon23┘ rexon24 1609_1662			
1582	GGCGTCCCTGGAGGTCCAGGAATGAGG	GGCATGCCCGGAAGTCCAGGAGACCA	1635
	-----+-----+-----	+-----+-----+-----	
528	GlyValProGlyGlyProGlyMetArg	GlyMetProGlySerProGlyGlyPro	545
361	G V P G G P G M R	G M P G S P G G P	378
1609_1662 exon24┘ rexon25 1663_1761			
1636	GGAAGTGATGGGAAACCAGGGCCTCCC	GGAAGTCAAGGAGAAAGTGGTCGACCA	1689
	----+-----+-----+---	-----+-----+-----	
546	GlySerAspGlyLysProGlyProPro	GlySerGlnGlyGluSerGlyArgPro	563
379	G S D G K P G P P	G S Q G E S G R P	396
1690	GGTCCTCCTGGGCCATCTGGTCCCCGA	GGTCAGCCTGGTGTCATGGGCTTCCCC	1743
	+-----+-----+-----	---+-----+-----+---	
564	GlyProProGlyProSerGlyProArg	GlyGlnProGlyValMetGlyPhePro	581
397	G P P G P S G P R	G Q P G V M G F P	414
1663_1761 exon25┘ rexon26 1762_1815			
1744	GGTCCTAAAGGAAATGATGGTGCTCCT	GGTAAGAATGGAGAACGAGGTGGCCCT	1797
	-----+-----+-----+---	-----+-----+-----	
582	GlyProLysGlyAsnAspGlyAlaPro	GlyLysAsnGlyGluArgGlyGlyPro	599
415	G P K G N D G A P	G K N G E R G G P	432
1762_1815 exon26┘ rexon27 1816_1869			
1798	GGAGGACCTGGCCCTCAGGGTCTCCT	GGAAAGAATGGTGAAACTGGACCTCAG	1851
	--+-----+-----+-----	-----+-----+-----+---	
600	GlyGlyProGlyProGlnGlyProPro	GlyLysAsnGlyGluThrGlyProGln	617
433	G G P G P Q G P P	G K N G E T G P Q	450
1816_1869 exon27┘ rexon28 1870_1923			
1852	GGACCCCCAGGGCCTACTGGGCCTGGT	GGTGACAAAGGAGACACAGGACCCCT	1905
	-----+-----+-----	+-----+-----+-----	
618	GlyProProGlyProThrGlyProGly	GlyAspLysGlyAspThrGlyProPro	635
451	G P P G P T G P G	G D K G D T G P P	468
1870_1923 exon28┘ rexon29 1924_1977			
1906	GGTCCACAAGGATTACAAGGCTTGCCCT	GGTACAGGTGGTCCTCCAGGAGAAAAT	1959
	----+-----+-----+---	-----+-----+-----	
636	GlyProGlnGlyLeuGlnGlyLeuPro	GlyThrGlyGlyProProGlyGluAsn	653
469	G P Q G L Q G L P	G T G G P P G E N	486
1924_1977 exon29┘ rexon30 1978_2022			
1960	GGAAAACCTGGGGAACCAGGTCCAAAG	GGTGATGCCGGTGCACCTGGAGCTCCA	2013
	+-----+-----+-----	---+-----+-----+---	
654	GlyLysProGlyGluProGlyProLys	GlyAspAlaGlyAlaProGlyAlaPro	671
487	G K P G E P G P K	G D A G A P G A P	504

1978_2022 exon30 | **exon31 2023_2121**
 2014 GGAGGCAAGGGTGATGCTGGTGCCCT GGTGAACGTGGACCTCCTGGATTGGCA 2067
 -----+-----+-----+ -----+-----+-----
 672 GlyGlyLysGlyAspAlaGlyAlaPro GlyGluArgGlyProProGlyLeuAla 689
 505 G G K G D A G A P G E R G P P G L A 522

2023_2121 exon31

2068 GGGCCCCAGGACTTAGAGGTGGAGCT GGTCCCCCTGGTCCCGAAGGAGGAAAG 2121
 --+-----+-----+-----+ -----+-----+-----+-----
 690 GlyAlaProGlyLeuArgGlyGlyAla GlyProProGlyProGluGlyGlyLys 707
 523 G A P G L R G G A G P P G P E G G K 540

exon32 2122_2229
 2122 GGTGCTGCTGGTCCTCCTGGGCCACCT GGTGCTGCTGGTACTCCTGGTCTGCAA 2175
 -----+-----+-----+ -+-----+-----+-----
 708 GlyAlaAlaGlyProProGlyProPro GlyAlaAlaGlyThrProGlyLeuGln 725
 541 G A A G P P G P P G A A G T P G L Q 558

2122_2229 exon32

2176 GGAATGCCTGGAGAAAGAGGAGGTCTT GGAAGTCCTGGTCCAAAGGGTGACAAG 2229
 ----+-----+-----+-----+ -----+-----+-----+-----
 726 GlyMetProGlyGluArgGlyGlyLeu GlySerProGlyProLysGlyAspLys 743
 559 G M P G E R G G L G S P G P K G D K 576

exon33 2230_2283 **2230_2283 exon33**
 2230 GGTGAACCAGGCGGTCCAGGTGCTGAT GGTGTCCAGGGAAAGATGGCCCAAGG 2283
 +-----+-----+-----+ -----+-----+-----+-----
 744 GlyGluProGlyGlyProGlyAlaAsp GlyValProGlyLysAspGlyProArg 761
 577 G E P G G P G A D G V P G K D G P R 594

exon34 2284_2337 **2284_2337 exon34**
 2284 GGTCTACTGGTCCTATTGGTCTCCTT GGCCCAGCTGGCCAGCCTGGAGATAAG 2337
 -----+-----+-----+ -----+-----+-----+-----
 762 GlyProThrGlyProIleGlyProPro GlyProAlaGlyGlnProGlyAspLys 779
 595 G P T G P I G P P G P A G Q P G D K 612

exon35 2338_2391 **2338_2391 exon35**
 2338 GGTGAAGGTGGTGCCCCCGGACTTCCA GGTATAGCTGGACCTCGTGGTAGCCCT 2391
 --+-----+-----+-----+ -----+-----+-----+-----
 780 GlyGluGlyGlyAlaProGlyLeuPro GlyIleAlaGlyProArgGlySerPro 797
 613 G E G G A P G L P G I A G P R G S P 630

exon36 2392_2445 **2392_2445 exon36**
 2392 GGTGAGAGAGGTGAAACTGGCCCTCCA GGACCTGCTGGTTTCCCTGGTGTCTCCT 2445
 -----+-----+-----+ -+-----+-----+-----
 798 GlyGluArgGlyGluThrGlyProPro GlyProAlaGlyPheProGlyAlaPro 815
 631 G E R G E T G P P G P A G F P G A P 648

exon37 2446_2553
 2446 GGACAGAATGGTGAACCTGGTGGTAAA GGAGAAAGAGGGGCTCCGGGTGAGAAA 2499
 ----+-----+-----+-----+ -----+-----+-----+-----
 816 GlyGlnAsnGlyGluProGlyGlyLys GlyGluArgGlyAlaProGlyGluLys 833
 649 G Q N G E P G G K G E R G A P G E K 666

		2446_2553 exon37	
2500	GGTGAAGGAGGCCCTCCTGGAGTTGCA	GGACCCCTGGAGGTCTGGACCTGCT	2553
	+-----+-----+-----	---+-----+-----+---	
834	GlyGluGlyGlyProProGlyValAla	GlyProProGlyGlySerGlyProAla	851
667	G E G G P P G V A	G P P G G S G P A	684
		2554_2607 exon38	
2554	GGTCCTCCTGGTCCCCAAGGTGTCAAA	GGTGAACGTGGCAGTCTGGTGGACCT	2607
	-----+-----+-----+-----	-----+-----+-----+-----	
852	GlyProProGlyProGlnGlyValLys	GlyGluArgGlySerProGlyGlyPro	869
685	G P P G P Q G V K	G E R G S P G G P	702
		2608_2661 exon39	
2608	GGTGCTGCTGGCTTCCCTGGTGCTCGT	GGTCTTCCTGGTCCCTCCTGGTAGTAAT	2661
	--+-----+-----+-----	-----+-----+-----+---	
870	GlyAlaAlaGlyPheProGlyAlaArg	GlyLeuProGlyProProGlySerAsn	887
703	G A A G F P G A R	G L P G P P G S N	720
		2662_2823 exon40	
2662	GGTAACCCAGGACCCCCAGGTCCCAGC	GGTCTCCAGGCAAGGATGGGCCCCCA	2715
	-----+-----+-----+-----	-+-----+-----+-----	
888	GlyAsnProGlyProProGlyProSer	GlySerProGlyLysAspGlyProPro	905
721	G N P G P P G P S	G S P G K D G P P	738
2716	GGTCCTGCGGGTAACACTGGTGCTCCT	GGCAGCCCTGGAGTGCTGGACCAAAA	2769
	----+-----+-----+---	-----+-----+-----	
906	GlyProAlaGlyAsnThrGlyAlaPro	GlySerProGlyValSerGlyProLys	923
739	G P A G N T G A P	G S P G V S G P K	756
		2662_2823 exon40	
2770	GGTGATGCTGGCCAACCAGGAGAGAAG	GGATCGCCTGGTGCCAGGGCCCACCA	2823
	+-----+-----+-----+---	---+-----+-----+---	
924	GlyAspAlaGlyGlnProGlyGluLys	GlySerProGlyAlaGlnGlyProPro	941
757	G D A G Q P G E K	G S P G A Q G P P	774
		2824_2931 exon41	
2824	GGAGCTCCAGGCCCACTTGGGATTGCT	GGGATCACTGGAGCACGGGGTCTTGCA	2877
	-----+-----+-----+-----	-----+-----+-----+-----	
942	GlyAlaProGlyProLeuGlyIleAla	GlyIleThrGlyAlaArgGlyLeuAla	959
775	G A P G P L G I A	G I T G A R G L A	792
		2824_2931 exon41	
2878	GGACCACCAGGCATGCCAGGTCTTAGG	GGAAGCCCTGGCCCTCAGGGTGTCAGG	2931
	--+-----+-----+-----	-----+-----+-----+---	
960	GlyProProGlyMetProGlyProArg	GlySerProGlyProGlnGlyValLys	977
793	G P P G M P G P R	G S P G P Q G V K	810
		2932_3039 exon42	
2932	GGTGAAAGTGGGAAACCAGGAGCTAAC	GGTCTCAGTGGAGAACGTGGTCCCCCT	2985
	-----+-----+-----+-----	-+-----+-----+-----	
978	GlyGluSerGlyLysProGlyAlaAsn	GlyLeuSerGlyGluArgGlyProPro	995
811	G E S G K P G A N	G L S G E R G P P	828

		2932_3039 exon42	
2986	GGACCCCAGGGTCTTCTGGTCTGGCT	GGTACAGCTGGTGAACCTGGAAGAGAT	3039
	-----+-----+-----+--	-----+-----+-----	
996	GlyProGlnGlyLeuProGlyLeuAla	GlyThrAlaGlyGluProGlyArgAsp	1013
829	G P Q G L P G L A	G T A G E P G R D	846
	exon43 3040_3093	3040_3093 exon43	
3040	GGAAACCCTGGATCAGATGGTCTTCCA	GGCCGAGATGGATCTCCTGGTGGCAAG	3093
	+-----+-----+-----	---+-----+-----+---	
1014	GlyAsnProGlySerAspGlyLeuPro	GlyArgAspGlySerProGlyGlyLys	1031
847	G N P G S D G L P	G R D G S P G G K	864
	exon44 3094_3201		
3094	GGTGATCGTGGTGAAAATGGCTCTCCT	GGTGCCCCTGGCGCTCCTGGTCATCCA	3147
	-----+-----+-----+---	-----+-----+-----	
1032	GlyAspArgGlyGluAsnGlySerPro	GlyAlaProGlyAlaProGlyHisPro	1049
865	G D R G E N G S P	G A P G A P G H P	882
		3094_3201 exon44	
3148	GGCCCACCTGGTCCTGTCTGGTCCAGCT	GGAAAGAGTGGTGACAGAGGAGAAAGT	3201
	--+-----+-----+-----+---	-----+-----+-----+---	
1050	GlyProProGlyProValGlyProAla	GlyLysSerGlyAspArgGlyGluSer	1067
883	G P P G P V G P A	G K S G D R G E S	900
	exon45 3202_3255	3202_3255 exon45	
3202	GGCCCTGCTGGCCCTGCTGGTGTCTCCC	GGTCTGCTGGTTCCTGAGGTGCTCCT	3255
	-----+-----+-----	-+-----+-----+-----	
1068	GlyProAlaGlyProAlaGlyAlaPro	GlyProAlaGlySerArgGlyAlaPro	1085
901	G P A G P A G A P	G P A G S R G A P	918
	exon46 3256_3363		
3256	GGTCTCAAGGCCACGTGGTGACAAA	GGTGAAACAGGTGAACGTGGAGCTGCT	3309
	-----+-----+-----+---	-----+-----+-----	
1086	GlyProGlnGlyProArgGlyAspLys	GlyGluThrGlyGluArgGlyAlaAla	1103
919	G P Q G P R G D K	G E T G E R G A A	936
		3256_3363 exon46	
3310	GGCATCAAAGGACATCGAGGATTCCCT	GGTAATCCAGGTGCCCCAGGTTCTCCA	3363
	+-----+-----+-----	---+-----+-----+---	
1104	GlyIleLysGlyHisArgGlyPhePro	GlyAsnProGlyAlaProGlySerPro	1121
937	G I K G H R G F P	G N P G A P G S P	954
	exon47 3364_3417	3364_3417 exon47	
3364	GGCCCTGCTGGTCAGCAGGGTGCATC	GGCAGTCCAGGACCTGCAGGCCCCAGA	3417
	-----+-----+-----+---	-----+-----+-----	
1122	GlyProAlaGlyGlnGlnGlyAlaIle	GlySerProGlyProAlaGlyProArg	1139
955	G P A G Q Q G A I	G S P G P A G P R	972
	exon48 3418_3525		
3418	GGACCTGTTGGACCCAGTGGACCTCCT	GGCAAAGATGGAACCAGTGGACATCCA	3471
	--+-----+-----+-----+---	-----+-----+-----+---	
1140	GlyProValGlyProSerGlyProPro	GlyLysAspGlyThrSerGlyHisPro	1157
973	G P V G P S G P P	G K D G T S G H P	990

3418_3525 exon48

3472	GGTCCCATTGGACCACCAGGGCCTCGA	GGTAACAGAGGTGAAAAGAGGATCTGAG	3525
	-----+-----+-----	-+-----+-----+-----	
1158	GlyProIleGlyProProGlyProArg	GlyAsnArgGlyGluArgGlySerGlu	1175
991	G P I G P P G P R	G N R G E R G S E	1008

exon49 3526_3823

3526	GGCTCCCCAGGCCACCCAGGGCAACCA	GGCCCTCCTGGACCTCCTGGTGCCCT	3579
	----+-----+-----+--	-----+-----+-----	
1176	GlySerProGlyHisProGlyGlnPro	GlyProProGlyProProGlyAlaPro	1193
1009	G S P G H P G Q P	G P P G P P G A P	1026

3580	GGTCCTTGC	3588
	+-----	
1194	GlyProCys	1196
1027	G P C	1029

C-terminal Telopeptide

3589	TGTGGTGGTGTGGAGCCGCTGCCATT	GCTGGGATTGGAGGTGAAAAAGCTGGC	3642
	-+-----+-----+-----	----+-----+-----+--	
1197	CysGlyGlyValGlyAlaAlaAlaIle	AlaGlyIleGlyGlyGluLysAlaGly	1214
	C G G V G A A A I	A G I G G E K A G	

3643	GGTTTTGCCCGTATTATGGA	3663
	-----+-----+---	
1215	GlyPheAlaProTyrTyrGly	1221
	G F A P Y Y G	

C-terminal Propeptide

3664	GATGAACCAATGGATTTCAAAATCAAC	ACCGATGAGATTATGACTTCACTCAAG	3717
	-----+-----+-----+	-----+-----+-----	
1222	AspGluProMetAspPheLysIleAsn	ThrAspGluIleMetThrSerLeuLys	1239
	D E P M D F K I N	T D E I M T S L K	

3718	TCTGTTAATGGACAAATAGAAAGCCTC	ATTAGTCCTGATGGTTCTCGTAAAAAC	3771
	--+-----+-----+-----	----+-----+-----+--	
1240	SerValAsnGlyGlnIleGluSerLeu	IleSerProAspGlySerArgLysAsn	1257
	S V N G Q I E S L	I S P D G S R K N	

3526_3823 exon49 | exon50 3824_4011

3772	CCCCTAGAAACTGCAGAGACCTGAAA	TTCTGCCATCCTGAACTCAAGAGTGA	3825
	-----+-----+-----+---	-+-----+-----+-----	
1258	ProAlaArgAsnCysArgAspLeuLys	PheCysHisProGluLeuLysSerGly	1275
	P A R N C R D L K	F C H P E L K S G	

3826	GAATACTGGGTTGACCCTAACCAAGGA	TGCAAATTGGATGCTATCAAGGTATTC	3879
	----+-----+-----+---	-----+-----+-----	
1276	GluTyrTrpValAspProAsnGlnGly	CysLysLeuAspAlaIleLysValPhe	1293
	E Y W V D P N Q G	C K L D A I K V F	

3880 TGTAATATGGAAACTGGGGAAACATGC ATAAGTGCCAATCCTTTGAATGTTCCA 3933
+-----+-----+----- -+-----+-----+-----
1294 CysAsnMetGluThrGlyGluThrCys IleSerAlaAsnProLeuAsnValPro 1311
C N M E T G E T C I S A N P L N V P
3934 CGGAAACACTGGTGGACAGATTCTAGT GCTGAGAAGAAACACGTTTGGTTTGGGA 3987
-----+-----+-----+----- -+-----+-----+-----
1312 ArgLysHisTrpTrpThrAspSerSer AlaGluLysLysHisValTrpPheGly 1329
R K H W W T D S S A E K K H V W F G

3824_4011 exon50_1-exon51 4012_4254

3988 GAGTCCATGGATGGTGGTTTTTCAGTTT AGCTACGGCAATCCTGAACTTCCTGAA 4041
--+-----+-----+----- -+-----+-----+-----+-----
1330 GluSerMetAspGlyGlyPheGlnPhe SerTyrGlyAsnProGluLeuProGlu 1347
E S M D G G F Q F S Y G N P E L P E
4042 GATGTCCTTGATGTGCAGCTGGCATT CTTGACTTCTCTCCAGCCGAGCTTCC 4095
-----+-----+----- -+-----+-----+-----
1348 AspValLeuAspValGlnLeuAlaPhe LeuArgLeuLeuSerSerArgAlaSer 1365
D V L D V Q L A F L R L L S S R A S
4096 CAGAACATCACATATCACTGCAAAAAT AGCATTGCATACATGGATCAGGCCAGT 4149
----+-----+-----+----- -+-----+-----+-----
1366 GlnAsnIleThrTyrHisCysLysAsn SerIleAlaTyrMetAspGlnAlaSer 1383
Q N I T Y H C K N S I A Y M D Q A S
4150 GGAAATGTAAAGAAGGCCCTGAAGCTG ATGGGGTCAAATGAAGGTGAATTCAAG 4203
+-----+-----+----- -+-----+-----+-----
1384 GlyAsnValLysLysAlaLeuLysLeu MetGlySerAsnGluGlyGluPheLys 1401
G N V K K A L K L M G S N E G E F K

4012_4254 exon51_1-exon52 4255_end

4204 GCTGAAGGAAATAGCAAATTCACCTAC ACAGTTCTGGAGGATGGTTGCACGAAA 4257
-----+-----+-----+----- -+-----+-----+-----
1402 AlaGluGlyAsnSerLysPheThrTyr ThrValLeuGluAspGlyCysThrLys 1419
A E G N S K F T Y T V L E D G C T K
4258 CACACTGGGGAATGGAGCAAAACAGTC TTTGAATATCGAACACGCAAGGCTGTG 4311
--+-----+-----+----- -+-----+-----+-----
1420 HisThrGlyGluTrpSerLysThrVal PheGluTyrArgThrArgLysAlaVal 1437
H T G E W S K T V F E Y R T R K A V
4312 AGACTACCTATTGTAGATATTGCACCC TATGACATTGGTGGTCTTGATCAAGAA 4365
-----+-----+----- -+-----+-----+-----
1438 ArgLeuProIleValAspIleAlaPro TyrAspIleGlyGlyProAspGlnGlu 1455
R L P I V D I A P Y D I G G P D Q E
4366 TTTGGTGTGGACGTTGGCCCTGTTTGC TTTTATAA 4401
-----+-----+----- -+-----+-----
1456 PheGlyValAspValGlyProValCys PheLeuEnd 1467
F G V D V G P V C F L

3' Untranslated region

```

*1   ACCAAACTCT ATCTGAAATC CCAACAAAAA AAATTTAACT CCATATGTGT *50
-----+ -----+ -----+ -----+ -----+
*51   TCCTCTTGTT CTAATCTTGT CAACCAGTGC AAGTGACCGA CAAAATTCCA *100
-----+ -----+ -----+ -----+ -----+
*101  GTTATTTATT TCCAAAATGT TTGGAAACAG TATAATTTGA CAAAGAAAAA *150
-----+ -----+ -----+ -----+ -----+
*151  TGATACTTCT CTTTTTTTGC TGTTCACCA AATACAATTC AAATGCTTTT *200
-----+ -----+ -----+ -----+ -----+
*201  TGTTTTATTT TTTTACCAAT TCCAATTTCA AAATGTCTCA ATGGTGCTAT *250
-----+ -----+ -----+ -----+ -----+
*251  AATAAATAAA CTTCAACACT CTTTATGATA ACAACACTGT GTTATATTCT *300
-----+ -----+ -----+ -----+ -----+
*301  TTGAATCCTA GCCCATCTGC AGAGCAATGA CTGTGCTCAC CAGTAAAAGA *350
-----+ -----+ -----+ -----+ -----+
*351  TAACCTTTCT TTCTGAAATA GTCAAATACG AAATTAGAAA AGCCCTCCCT *400
-----+ -----+ -----+ -----+ -----+
*401  ATTTTAACTA CCTCAACTGG TCAGAAACAC AGATTGTATT CTATGAGTCC *450
-----+ -----+ -----+ -----+ -----+
*451  CAGAAGATGA AAAAAATTTT ATACGTTGAT AAAACTTATA AATTTCATTG *500
-----+ -----+ -----+ -----+ -----+
*501  ATTAATCTCC TGGAAGATTG GTTTAAAAAG AAAAGTGTA TGAAGAATT *550
-----+ -----+ -----+ -----+ -----+
*551  TAAAGAAATA TTTTAAAGC CACAATTATT TTAATATTGG ATATCAACTG *600
-----+ -----+ -----+ -----+ -----+
*601  CTTGTAAAGG TGCTCCTCTT TTTTCTTGTC ATTGCTGGTC AAGATTACTA *650
-----+ -----+ -----+ -----+ -----+
*651  ATATTTGGGA AGGCTTTAAA GACGCATGTT ATGGTGCTAA TGTACTIONTCA *700
-----+ -----+ -----+ -----+ -----+
*701  CTTTTAAACT CTAGATCAGA ATTGTTGACT TGCATTCAGA ACATAAATGC *750
-----+ -----+ -----+ -----+ -----+
*751  ACAAATCTG TACATGTCTC CCATCAGAAA GATTCATTGG CATGCCACAG *800
-----+ -----+ -----+ -----+ -----+
*801  GGGATTCTCC TCCTTCATCC TGTAAGGTC AACAATAAAA ACCAAATTAT *850
-----+ -----+ -----+ -----+ -----+
*851  GGGGCTGCTT TTGTCACACT AGCATAGAGA ATGTGTTGAA ATTTAACTTT *900
-----+ -----+ -----+ -----+ -----+
*901  GTAAGCTTGT ATGTGGTTGT TGATCTTTTT TTTCTTACA GACACCCATA *950
-----+ -----+ -----+ -----+ -----+
*951  ATAAAATATC ATATTAAAAT TC
-----+ -----+ -- *972

```