

# COL1A2 Systematic Numbering of DNA and Protein Sequences

The reference cDNA used here is based on GenBank RefSeq NM\_000089.3. The corresponding reference genomic DNA sequence is GenBank RefSeqGene NG\_007405.1. The corresponding protein sequence is GenBank RefSeq NP\_000080.2.

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 “legacy” fashion with the first glycine of the triple helix numbered as “1”. Legacy amino acid numbers can be calculated by subtracting 90 from the systematic number.

Version 3: April 2010

## 5' Untranslated Region

```

                CACCA CGGCAGCAGG AGGTTTCGGC TAAGTTGGAG
                -135  ----- +-----+-----+----- -101

    GTACTGGCCA CGACTGCATG CCCGCGCCCG CCAGGTGATA CCTCCGCCGG
-100 +-----+-----+-----+-----+----- -51

    TGACCCAGGG GCTCTGCGAC ACAAGGAGTC TGCATGTCTA AGTGCTAGAC
-50  +-----+-----+-----+-----+----- -1

```

## Signal Peptide

```

1  ATGCTCAGCTTTGTGGATACGCGGACT TGTGCTGCTTGCGAGTAACCTTATGC 54
   -----+-----+-----+-----+-----+-----+-----
1  MetLeuSerPheValAspThrArgThr LeuLeuLeuLeuAlaValThrLeuCys 18
   M L S F V D T R T L L L L A V T L C

55  CTAGCAACATGC 66
   -----+-----
19  LeuAlaThrCys 22
   L A T C

```

## N-terminal Propeptide

```

1_70 Exon1  Exon2  Exon3  Exon4 97_133
67  CAATCTTTACAAGAGGAAACTGTAAGA AAGGGCCCGAGATAGAGGACCA 120
   -----+-----+-----+-----+-----+-----+
23  GlnSerLeuGlnGluGluThrValArg LysGlyProAlaGlyAspArgGlyPro 40
   Q S L Q E E T V R K G P A G D R G P

97_132 Exon4  Exon5 133_225
121 CGTGGAGAAAGGGGTCCACCAGGCCCC CCAGGCAGAGATGGTGAAGATGGTCCC 174
   -----+-----+-----+-----+-----+
41  ArgGlyGluArgGlyProProGlyPro ProGlyArgAspGlyGluAspGlyPro 58
   R G E R G P P G P P G R D G E D G P

```

133\_225 Exon5┘ Exon6 226\_

175 ACAGGCCCTCCTGGTCCACCTGGTCCT CCTGGCCCCCTGGTCTCGGTGGGAAC 228  
 -----+-----+-----+-----  
 59 ThrGlyProProGlyProProGlyPro ProGlyProProGlyLeuGlyGlyAsn 76  
 T G P P G P P G P P G P P G L G G N

229 TTTGCTGCT 237  
 -+-----  
 77 PheAlaAla 79  
 F A A

### ***N-terminal Telopeptide***

238 CAGTATGATGGAAAAGGAGTTGGACTT GGCCCT 270  
 ---+-----+-----+-----  
 80 GlnTyrAspGlyLysGlyValGlyLeu GlyPro 90  
 Q Y D G K G V G L G P

### ***Triple Helix***

226\_279 Exon6┘ Exon7 280\_324 280\_324 Exon7┘

271 GGACCAATGGGCTTAATGGGACCTAGA GGCCACCTGGTGCAGCTGGAGCCCCA 324  
 -----+-----+-----+-----  
 91 GlyProMetGlyLeuMetGlyProArg GlyProProGlyAlaAlaGlyAlaPro 108  
 1 G P M G L M G P R G P P G A A G A P 18

┘Exon8 325\_378 325\_378 Exon8┘

325 GGCCCTCAAGGTTTCCAAGGACCTGCT GGTGAGCCTGGTGAACCTGGTCAAACCT 378  
 -----+-----+-----+-----  
 109 GlyProGlnGlyPheGlnGlyProAla GlyGluProGlyGluProGlyGlnThr 126  
 19 G P Q G F Q G P A G E P G E P G Q T 36

┘Exon9 379\_432 379\_432 Exon9┘

379 GGTCTGCAGGTGCTCGTGGTCCAGCT GGCCCTCCTGGCAAGGCTGGTGAAGAT 432  
 -+-----+-----+-----+-----  
 127 GlyProAlaGlyAlaArgGlyProAla GlyProProGlyLysAlaGlyGluAsp 144  
 37 G P A G A R G P A G P P G K A G E D 54

┘Exon 10 433\_486 433\_486 Exon10┘

433 GGTCACCCTGGAAAACCCGGACGACCT GGTGAGAGAGGAGTTGTTGGACCACAG 486  
 -----+-----+-----+-----  
 145 GlyHisProGlyLysProGlyArgPro GlyGluArgGlyValValGlyProGln 162  
 55 G H P G K P G R P G E R G V V G P Q 72

┘Exon11 487\_540 487\_540 Exon11┘

487 GGTGCTCGTGGTTTCCCTGGAACCTCCT GGACTTCCTGGCTTCAAAGGCATTAGG 540  
 ---+-----+-----+-----+-----  
 163 GlyAlaArgGlyPheProGlyThrPro GlyLeuProGlyPheLysGlyIleArg 180  
 73 G A R G F P G T P G L P G F K G I R 90

┘Exon12 541\_594 541\_594 Exon12┘

541 GGACACAATGGTCTGGATGGATTGAAG GGACAGCCCCGGTCTCTGGTGTGAAG 594  
 -----+-----+-----+-----  
 181 GlyHisAsnGlyLeuAspGlyLeuLys GlyGlnProGlyAlaProGlyValLys 198  
 91 G H N G L D G L K G Q P G A P G V K 108

	<b>┌Exon13 595_648</b>	<b>595_638 Exon13┐</b>	<b>┌Exon14 640_693</b>	
595	GGTGAACCTGGTGGCCCTGGTAAAAAT	GGAACCTCCAGGTCAAACAGGAGCCCGT		648
	-----+-----+-----+-----	-----+-----+-----+-----		
199	GlyGluProGlyAlaProGlyGluAsn	GlyThrProGlyGlnThrGlyAlaArg		216
109	G E P G A P G E N	G T P G Q T G A R		126
		<b>640_693 Exon14┐</b>	<b>┌Exon15 694_738</b>	
649	GGGCTTCCTGGTGAGAGAGGACGTGTT	GGTGCCCTGGCCAGCTGGTGGCCCGT		702
	-+-----+-----+-----+-----	-----+-----+-----+-----		
217	GlyLeuProGlyGluArgGlyArgVal	GlyAlaProGlyProAlaGlyAlaArg		234
127	G L P G E R G R V	G A P G P A G A R		144
		<b>694_738 Exon15┐</b>	<b>┌Exon16 739_792</b>	
703	GGCAGTGATGGAAGTGTGGGTCCCGTG	GGTCCCTGCTGGTCCCATTGGGTCTGCT		756
	-----+-----+-----+-----	+-----+-----+-----+-----		
235	GlySerAspGlySerValGlyProVal	GlyProAlaGlyProIleGlySerAla		252
145	G S D G S V G P V	G P A G P I G S A		162
		<b>694_792 Exon16┐</b>	<b>┌Exon17 793_891</b>	
757	GGCCCTCCAGGCTTCCCAGGTGCCCT	GGCCCCAAGGGTGAAATTGGAGCTGTT		810
	---+-----+-----+-----+-----	-----+-----+-----+-----		
253	GlyProProGlyPheProGlyAlaPro	GlyProLysGlyGluIleGlyAlaVal		270
163	G P P G F P G A P	G P K G E I G A V		180
811	GGTAACGCTGGTCCTGCTGGTCCCGCC	GGTCCCCGTGGTGAAGTGGGTCTTCCA		864
	-----+-----+-----+-----	---+-----+-----+-----		
271	GlyAsnAlaGlyProAlaGlyProAla	GlyProArgGlyGluValGlyLeuPro		288
181	G N A G P A G P A	G P R G E V G L P		198
		<b>793_891 Exon17┐</b>	<b>┌Exon18 892_936</b>	
865	GGCCTCTCCGGCCCCGTGGACCTCCT	GGTAATCCTGGAGCAAACGGCCTTACT		918
	-----+-----+-----+-----	-----+-----+-----+-----		
289	GlyLeuSerGlyProValGlyProPro	GlyAsnProGlyAlaAsnGlyLeuThr		306
199	G L S G P V G P P	G N P G A N G L T		216
		<b>892_936 Exon18┐</b>	<b>┌Exon19 937_1035</b>	
919	GGTGCCAAGGGTGCTGCTGGCCTTCCC	GGCGTTGCTGGGGCTCCCGGCCTCCCT		972
	-+-----+-----+-----+-----	-----+-----+-----+-----		
307	GlyAlaLysGlyAlaAlaGlyLeuPro	GlyValAlaGlyAlaProGlyLeuPro		324
217	G A K G A A G L P	G V A G A P G L P		234
973	GGACCCCGCGGTATTCTGGCCCTGTT	GGTGCTGCCGGTGCTACTGGTGCCAGA		1026
	-----+-----+-----+-----	+-----+-----+-----+-----		
325	GlyProArgGlyIleProGlyProVal	GlyAlaAlaGlyAlaThrGlyAlaArg		342
235	G P R G I P G P V	G A A G A T G A R		252
		<b>937_1035 Exon19┐</b>	<b>┌Exon20 1036_1089</b>	
1027	GGACTTGTTGGTGAGCCTGGTCCAGCT	GGCTCCAAAGGAGAGAGCGGTAACAAG		1080
	---+-----+-----+-----+-----	-----+-----+-----+-----		
343	GlyLeuValGlyGluProGlyProAla	GlySerLysGlyGluSerGlyAsnLys		360
253	G L V G E P G P A	G S K G E S G N K		270

**1036\_1089 Exon20** | **Exon21 1090\_1197**

1081 GGTGAGCCCGGCTCTGCTGGGCCCCAA GGTCTCCTGGTCCCAGTGGTGAAGAA 1134  
 -----+-----+----- -+-----+-----+-----  
 361 GlyGluProGlySerAlaGlyProGln GlyProProGlyProSerGlyGluGlu 378  
 271 G E P G S A G P Q G P P G P S G E E 288

1135 GGAAAGAGAGGCCCTAATGGGGAAGCT GGATCTGCCGCCCTCCAGGACCTCCT 1188  
 -----+-----+-----+----- -+-----+-----+-----  
 379 GlyLysArgGlyProAsnGlyGluAla GlySerAlaGlyProProGlyProPro 396  
 289 G K R G P N G E A G S A G P P G P P 306

**1090\_1197 Exon21** | **Exon22 1198\_1251**

1189 GGGCTGAGAGGTAGTCTGCTGGTTCTCGT GGTCTTCCTGGAGCTGATGGCAGAGCT 1242  
 -+-----+-----+----- -+-----+-----+-----  
 397 GlyLeuArgGlySerProGlySerArg GlyLeuProGlyAlaAspGlyArgAla 414  
 307 G L R G S P G S R G L P G A D G R A 324

**1198\_1251 Exon22** | **Exon23 1252\_1350**

1243 GGCCTCATGGGCCCTCCTGGTAGTCGT GGTGCAAGTGGCCCTGCTGGAGTCCGA 1296  
 -----+-----+-----+----- +-----+-----+-----+-----  
 415 GlyValMetGlyProProGlySerArg GlyAlaSerGlyProAlaGlyValArg 432  
 325 G V M G P P G S R G A S G P A G V R 342

**1252\_1350 Exon23**

1297 GGACCTAATGGAGATGCTGGTCCGCCCT GGGGAGCCTGGTCTCATGGGACCCAGA 1350  
 ---+-----+-----+----- -+-----+-----+-----+-----  
 433 GlyProAsnGlyAspAlaGlyArgPro GlyGluProGlyLeuMetGlyProArg 450  
 343 G P N G D A G R P G E P G L M G P R 360

**Exon24 1351\_1404** | **1351\_1404 Exon24**

1351 GGTCTTCCTGGTTCGCCCTGGAAATATC GGCCCCGCTGGAAAAGAAGGTCCTGTC 1404  
 -----+-----+-----+----- -+-----+-----+-----+-----  
 451 GlyLeuProGlySerProGlyAsnIle GlyProAlaGlyLysGluGlyProVal 468  
 361 G L P G S P G N I G P A G K E G P V 378

**Exon25 1405\_1503**

1405 GGCCTCCCTGGCATTGACGGCAGGCCT GGCCCAATTGGCCCAGCTGGAGCAAGA 1458  
 -----+-----+-----+----- -+-----+-----+-----+-----  
 469 GlyLeuProGlyIleAspGlyArgPro GlyProIleGlyProAlaGlyAlaArg 486  
 379 G L P G I D G R P G P I G P A G A R 396

**1405\_1503 Exon25** | **Exon26 1504\_1557**

1459 GGAGAGCCTGGCAACATTGGATTCCCT GGACCCAAAGGCCCCACTGGTGATCCT 1512  
 -+-----+-----+-----+----- -+-----+-----+-----+-----  
 487 GlyGluProGlyAsnIleGlyPhePro GlyProLysGlyProThrGlyAspPro 504  
 397 G E P G N I G F P G P K G P T G D P 414

**1504\_1557 Exon26** | **Exon27 1558\_1611**

1513 GGCAAAAACGGTGATAAAGGTCATGCT GGTCTTGCTGGTGCTCGGGGTGCTCCA 1566  
 -----+-----+-----+----- +-----+-----+-----+-----  
 505 GlyLysAsnGlyAspLysGlyHisAla GlyLeuAlaGlyAlaArgGlyAlaPro 522  
 415 G K N G D K G H A G L A G A R G A P 432

		<b>1588_1611 Exon27</b>	<b>Exon28</b>	<b>1612_1665</b>	
1567	GGTCCTGATGGAAACAATGGTGCTCAG	GGACCTCCTGGACCACAGGGTGTTC	CAA	1620	
	----+-----+-----+-----	-----+-----+-----+-----			
523	GlyProAspGlyAsnAsnGlyAlaGln	GlyProProGlyProGlnGlyValGln		540	
433	G P D G N N G A Q	G P P G P Q G V Q		450	
		<b>1612_1665 Exon28</b>	<b>Exon29</b>	<b>1666_1719</b>	
1621	GGTGGAAAAGGTGAACAGGGTCCCCCT	GGTCCTCCAGGCTTCCAGGGTCTGCCT		1674	
	-----+-----+-----+-----	-+-----+-----+-----			
541	GlyGlyLysGlyGluGlnGlyProPro	GlyProProGlyPheGlnGlyLeuPro		558	
451	G G K G E Q G P P	G P P G F Q G L P		468	
		<b>1666_1719 Exon29</b>	<b>Exon30</b>	<b>1720_1764</b>	
1675	GGCCCCCTCAGGTCCCCTGGTGAAGTT	GGCAAACCAGGAGAAAGGGGTCTCCAT		1728	
	-----+-----+-----+-----	-----+-----+-----+-----			
559	GlyProSerGlyProAlaGlyGluVal	GlyLysProGlyGluArgGlyLeuHis		576	
469	G P S G P A G E V	G K P G E R G L H		486	
		<b>1720_1764 Exon30</b>	<b>Exon31</b>	<b>1765_1863</b>	
1729	GGTGAGTTTGGTCTCCCTGGTCCTGCT	GGTCCAAGAGGGGAACGCGGTCCCCCA		1782	
	-+-----+-----+-----+-----	-----+-----+-----+-----			
577	GlyGluPheGlyLeuProGlyProAla	GlyProArgGlyGluArgGlyProPro		594	
487	G E F G L P G P A	G P R G E R G P P		504	
		<b>1783_1836 Exon31</b>	<b>Exon32</b>	<b>1864_1971</b>	
1783	GGTGAGAGTGGTGCTGCCGGTCTACT	GGTCCTATTGGAAGCCGAGGTCTTCT		1836	
	-----+-----+-----+-----	+-----+-----+-----+-----			
595	GlyGluSerGlyAlaAlaGlyProThr	GlyProIleGlySerArgGlyProSer		612	
505	G E S G A A G P T	G P I G S R G P S		522	
		<b>1765_1863 Exon31</b>	<b>Exon32</b>	<b>1864_1971</b>	
1837	GGACCCCCAGGGCCTGATGGAAACAAG	GGTGAACCTGGTGTGGTTGGTGCTGTG		1890	
	----+-----+-----+-----	-----+-----+-----+-----			
613	GlyProProGlyProAspGlyAsnLys	GlyGluProGlyValValGlyAlaVal		630	
523	G P P G P D G N K	G E P G V V G A V		540	
		<b>1891_1944 Exon32</b>	<b>Exon33</b>	<b>1972_2025</b>	
1891	GGCACTGCTGGTCCATCTGGTCCCTAGT	GGACTCCCAGGAGAGAGGGGTGCTGCT		1944	
	-----+-----+-----+-----	-+-----+-----+-----			
631	GlyThrAlaGlyProSerGlyProSer	GlyLeuProGlyGluArgGlyAlaAla		648	
541	G T A G P S G P S	G L P G E R G A A		558	
		<b>1864_1971 Exon32</b>	<b>Exon33</b>	<b>1972_2025</b>	
1945	GGCATACTGGAGGCAAGGGAGAAAAG	GGTGAACCTGGTCTCAGAGGTGAAATT		1998	
	-----+-----+-----+-----	-----+-----+-----+-----			
649	GlyIleProGlyGlyLysGlyGluLys	GlyGluProGlyLeuArgGlyGluIle		666	
559	G I P G G K G E K	G E P G L R G E I		576	
		<b>1972_2025 Exon33</b>	<b>Exon34</b>	<b>2026_2079</b>	
1999	GGTAACCCTGGCAGAGATGGTGCTCGT	GGTGCTCCTGGTGCTGTAGGTGCCCT		2052	
	-+-----+-----+-----+-----	-----+-----+-----+-----			
667	GlyAsnProGlyArgAspGlyAlaArg	GlyAlaProGlyAlaValGlyAlaPro		684	
577	G N P G R D G A R	G A P G A V G A P		594	

	<b>2026_2079 Exon34</b>	<b>Exon35 2080_2133</b>	
2053	GGTCCTGCTGGAGCCACAGGTGACCCGG	GGCGAAGCTGGGGCTGCTGGTCCTGCT	2106
	-----+-----+-----	+-----+-----+-----	
685	GlyProAlaGlyAlaThrGlyAspArg	GlyGluAlaGlyAlaAlaGlyProAla	702
595	G P A G A T G D R	G E A G A A G P A	612
	<b>2080_2133 Exon35</b>	<b>Exon36 2134_2187</b>	
2107	GGTCCTGCTGGTCCTCGGGGAAGCCCT	GGTGAACGTGGTGAGGTCGGTCCTGCT	2160
	---+-----+-----+---	-----+-----+-----+	
703	GlyProAlaGlyProArgGlySerPro	GlyGluArgGlyGluValGlyProAla	720
613	G P A G P R G S P	G E R G E V G P A	630
	<b>2134_2187 Exon36</b>	<b>Exon37 2188_2295</b>	
2161	GGCCCCAATGGATTTGCTGGTCCTGCT	GGTGTCTGCTGGTCAACCTGGTGCTAAA	2214
	-----+-----+-----	-+-----+-----+---	
721	GlyProAsnGlyPheAlaGlyProAla	GlyAlaAlaGlyGlnProGlyAlaLys	738
631	G P N G F A G P A	G A A G Q P G A K	648
2215	GGAGAAAGAGGAGCCAAAGGGCCTAAG	GGTGAACCGGTGTTGTTGGTCCCACA	2268
	-----+-----+-----+	-----+-----+-----+	
739	GlyGluArgGlyAlaLysGlyProLys	GlyGluAsnGlyValValGlyProThr	756
649	G E R G A K G P K	G E N G V V G P T	666
	<b>2188_2295 Exon37</b>	<b>Exon38 2296_2349</b>	
2269	GGCCCCGTTGGAGCTGCTGGCCCAGCT	GGTCCAAATGGTCCCCCGGTCTCTGCT	2322
	-+-----+-----+-----	----+-----+-----+---	
757	GlyProValGlyAlaAlaGlyProAla	GlyProAsnGlyProProGlyProAla	774
667	G P V G A A G P A	G P N G P P G P A	684
	<b>2296_2349 Exon38</b>	<b>Exon39 2350_2403</b>	
2323	GGAAGTCGTGGTGATGGAGGCCCCCT	GGTATGACTGGTTTCCCTGGTGCTGCT	2376
	-----+-----+-----	+-----+-----+-----	
775	GlySerArgGlyAspGlyGlyProPro	GlyMetThrGlyPheProGlyAlaAla	792
685	G S R G D G G P P	G M T G F P G A A	702
	<b>2350_2403 Exon39</b>	<b>Exon40 2404_2565</b>	
2377	GGACGACTGGTCCCCAGGACCCTCT	GGTATTTCTGGCCCTCCTGGTCCCCCT	2430
	---+-----+-----+---	-----+-----+-----+	
793	GlyArgThrGlyProProGlyProSer	GlyIleSerGlyProProGlyProPro	810
703	G R T G P P G P S	G I S G P P G P P	720
2431	GGTCCTGCTGGGAAAGAAGGGCTTCGT	GGTCCTCGTGGTGACCAAGGTCCAGTT	2484
	-----+-----+-----	-+-----+-----+---	
811	GlyProAlaGlyLysGluGlyLeuArg	GlyProArgGlyAspGlnGlyProVal	828
721	G P A G K E G L R	G P R G D Q G P V	738
2485	GGCCGAACTGGAGAAGTAGGTGCAGTT	GGTCCCCCTGGCTTCGCTGGTGAGAAG	2538
	-----+-----+-----+	-----+-----+-----+	
829	GlyArgThrGlyGluValGlyAlaVal	GlyProProGlyPheAlaGlyGluLys	846
739	G R T G E V G A V	G P P G F A G E K	756

	<b>2404_2565 Exon40</b>	<b>Exon41 2566_2673</b>	
2539	GGTCCCTCTGGAGAGGCTGGTACTGCT	GGACCTCCTGGCACTCCAGGTCCTCAG	2592
	-+-----+-----+-----	-----+-----+-----+--	
847	GlyProSerGlyGluAlaGlyThrAla	GlyProProGlyThrProGlyProGln	864
757	G P S G E A G T A	G P P G T P G P Q	774
2593	GGTCTTCTTGGTGCTCCTGGTATTCTG	GGTCTCCCTGGCTCGAGAGGTGAACGT	2646
	-----+-----+-----	+-----+-----+-----	
865	GlyLeuLeuGlyAlaProGlyIleLeu	GlyLeuProGlySerArgGlyGluArg	882
775	G L L G A P G I L	G L P G S R G E R	792
	<b>2566_2673 Exon41</b>	<b>Exon42 2674_2781</b>	
2647	GGTCTACCAGGTGTTGCTGGTGTGTG	GGTGAACCTGGTCCTCTTGGCATTGCC	2700
	---+-----+-----+---	-----+-----+-----+---	
883	GlyLeuProGlyValAlaGlyAlaVal	GlyGluProGlyProLeuGlyIleAla	900
793	G L P G V A G A V	G E P G P L G I A	810
2701	GGCCCTCCTGGGGCCCGTGGTCTCCT	GGTGCTGTGGGTAGTCCTGGAGTCAAC	2754
	-----+-----+-----	--+-----+-----+---	
901	GlyProProGlyAlaArgGlyProPro	GlyAlaValGlySerProGlyValAsn	918
811	G P P G A R G P P	G A V G S P G V N	828
	<b>2674_2781 Exon42</b>	<b>Exon43 2782_2835</b>	
2755	GGTGCTCCTGGTGAAGCTGGTCGTGAT	GGCAACCCTGGGAACGATGGTCCCCCA	2808
	-----+-----+-----+---	-----+-----+-----+---	
919	GlyAlaProGlyGluAlaGlyArgAsp	GlyAsnProGlyAsnAspGlyProPro	936
829	G A P G E A G R D	G N P G N D G P P	846
	<b>2782_2835 Exon43</b>	<b>Exon44 2836-2943</b>	
2809	GGTCGCGATGGTCAACCCGGACACAAG	GGAGAGCGCGGTTACCCTGGCAATATT	2862
	-+-----+-----+-----	---+-----+-----+---	
937	GlyArgAspGlyGlnProGlyHisLys	GlyGluArgGlyTyrProGlyAsnIle	954
847	G R D G Q P G H K	G E R G Y P G N I	864
2863	GGTCCCGTTGGTGCTGCAGGTGCACCT	GGTCCTCATGGCCCCGTGGGTCCTGCT	2916
	-----+-----+-----+---	+-----+-----+-----+---	
955	GlyProValGlyAlaAlaGlyAlaPro	GlyProHisGlyProValGlyProAla	972
865	G P V G A A G A P	G P H G P V G P A	882
	<b>2890_2943 Exon44</b>	<b>Exon45 2944_2997</b>	
2917	GGCAAACATGGAAACCGTGGTGAACCT	GGTCCTTCTGGTCCTGTTGGTCCTGCT	2970
	---+-----+-----+---	-----+-----+-----+---	
973	GlyLysHisGlyAsnArgGlyGluThr	GlyProSerGlyProValGlyProAla	990
883	G K H G N R G E T	G P S G P V G P A	900
	<b>2944_2997 Exon45</b>	<b>Exon46 2998_3105</b>	
2971	GGTGCTGTTGGCCCAAGAGGTCTTAGT	GGCCCACAAGGCATTCGTGGCGATAAG	3024
	-----+-----+-----+---	--+-----+-----+---	
991	GlyAlaValGlyProArgGlyProSer	GlyProGlnGlyIleArgGlyAspLys	1008
901	G A V G P R G P S	G P Q G I R G D K	918
3025	GGAGAGCCCAGGTGAAAAGGGGCCCAAG	GGTCTTCTGGCTTAAAGGGACACAAT	3078
	-----+-----+-----+---	-----+-----+-----+---	
1009	GlyGluProGlyGluLysGlyProArg	GlyLeuProGlyLeuLysGlyHisAsn	1026
919	G E P G E K G P R	G L P G L K G H N	936

	<b>2998_3105 Exon46</b>	<b>Exon47 3105_3159</b>	
3079	GGATTGCAAGGTCTGCCTGGTATCGCT	GGTCACCATGGTGATCAAGGTGCTCCT	3132
	-+-----+-----+-----	-----+-----+-----+--	
1027	GlyLeuGlnGlyLeuProGlyIleAla	GlyHisHisGlyAspGlnGlyAlaPro	1044
937	G L Q G L P G I A	G H H G D Q G A P	954
	<b>3105_3159 Exon47</b>	<b>Exon48 3160_3267</b>	
3133	GGCTCCGTGGGTCTGCTGGTCTTAGG	GGCCCTGCTGGTCTTCTGGCCCTGCT	3186
	-----+-----+-----	+-----+-----+-----	
1045	GlySerValGlyProAlaGlyProArg	GlyProAlaGlyProSerGlyProAla	1062
955	G S V G P A G P R	G P A G P S G P A	972
3187	GGAAAAGATGGTTCGCACTGGACATCCT	GGTACAGTTGGACCTGCTGGCATTCTGA	3240
	---+-----+-----+---	-----+-----+-----+---	
1063	GlyLysAspGlyArgThrGlyHisPro	GlyThrValGlyProAlaGlyIleArg	1080
973	G K D G R T G H P	G T V G P A G I R	990
	<b>3160_3267 Exon48</b>	<b>Exon49 3268_3526</b>	
3241	GGCCCTCAGGGTCACCAAGGCCCTGCT	GGCCCCCTGGTCCCCCTGGCCCTCCT	3294
	-----+-----+-----	--+-----+-----+---	
1081	GlyProGlnGlyHisGlnGlyProAla	GlyProProGlyProProGlyProPro	1098
991	G P Q G H Q G P A	G P P G P P G P P	1008
3295	GGACCTCCAGGTGTAAGC		3312
	-----+-----+---		
1099	GlyProProGlyValSer		1104
1009	G P P G V S		1014

### ***C-terminal Telopeptide***

3313	GGTGGTGGTTATGACTTTGGTTACGAT	GGAGACTTCTACAGGGCT	3357
	-----+-----+-----	+-----+-----	
1105	GlyGlyGlyTyrAspPheGlyTyrAsp	GlyAspPheTyrArgAla	1119
	G G G Y D F G Y D	G D F Y R A	

### ***C-terminal Propeptide***

3358	GACCAGCCTCGCTCAGCACCTTCTCTC	AGACCCAAGGACTATGAAGTTGATGCT	3411
	-+-----+-----+---	-----+-----+-----+---	
1120	AspGlnProArgSerAlaProSerLeu	ArgProLysAspTyrGluValAspAla	1137
	D Q P R S A P S L	R P K D Y E V D A	
3412	ACTCTGAAGTCTCTCAACAACCAGATT	GAGACCCCTTCTTACTCCTGAAGGCTCT	3465
	-----+-----+-----	-+-----+-----+---	
1138	ThrLeuLysSerLeuAsnAsnGlnIle	GluThrLeuLeuThrProGluGlySer	1155
	T L K S L N N Q I	E T L L T P E G S	
3466	AGAAAGAACCCAGCTCGCACATGCCGT	GACTTGAGACTCAGCCACCCAGAGTGG	3519
	----+-----+-----+---	-----+-----+-----	
1156	ArgLysAsnProAlaArgThrCysArg	AspLeuArgLeuSerHisProGluTrp	1173
	R K N P A R T C R	D L R L S H P E W	



**3268\_3526**

**Exon 49 | Exon50 3527\_3711**

3520	AGCAGTGGTTACTACTGGATTGACCCT	AACCAAGGATGCACTATGGATGCTATC	3573
	+-----+-----+-----+	---+-----+-----+---	
1174	SerSerGlyTyrTyrTrpIleAspPro	AsnGlnGlyCysThrMetAspAlaIle	1191
	S S G Y Y W I D P	N Q G C T M D A I	
3574	AAAGTATACTGTGATTTCTCTACTGGC	GAAACCTGTATCCGGGCCCAACCTGAA	3627
	-----+-----+-----+	-----+-----+-----+	
1192	LysValTyrCysAspPheSerThrGly	GluThrCysIleArgAlaGlnProGlu	1209
	K V Y C D F S T G	E T C I R A Q P E	
3628	AACATCCCAGCCAAGAACTGGTATAGG	AGCTCCAAGGACAAGAAACACGTCTGG	3681
	--+-----+-----+-----+	-----+-----+-----+---	
1210	AsnIleProAlaLysAsnTrpTyrArg	SerSerLysAspLysLysHisValTrp	1227
	N I P A K N W Y R	S S K D K K H V W	

**3527\_3711 Exon50 | Exon51 3712\_3954**

3682	CTAGGAGAAACTATCAATGCTGGCAGC	CAGTTTGAATATAATGTAGAAGGAGTG	3735
	-----+-----+-----+	-----+-----+-----+	
1228	LeuGlyGluThrIleAsnAlaGlySer	GlnPheGluTyrAsnValGluGlyVal	1245
	L G E T I N A G S	Q F E Y N V E G V	
3736	ACTTCCAAGGAAATGGCTACCCAACCTT	GCCTTCATGCGCCTGCTGGCCAACCTAT	3789
	----+-----+-----+---	-----+-----+-----+---	
1246	ThrSerLysGluMetAlaThrGlnLeu	AlaPheMetArgLeuLeuAlaAsnTyr	1263
	T S K E M A T Q L	A F M R L L A N Y	
3790	GCCTCTCAGAACATCACCTACCACTGC	AAGAACAGCATTGCATACATGGATGAG	3843
	+-----+-----+-----+	---+-----+-----+---	
1264	AlaSerGlnAsnIleThrTyrHisCys	LysAsnSerIleAlaTyrMetAspGlu	1281
	A S Q N I T Y H C	K N S I A Y M D E	
3844	GAGACTGGCAACCTGAAAAAGGCTGTC	ATTCTACAGGGCTCTAATGATGTTGAA	3897
	-----+-----+-----+	-----+-----+-----+	
1282	GluThrGlyAsnLeuLysLysAlaVal	IleLeuGlnGlySerAsnAspValGlu	1299
	E T G N L K K A V	I L Q G S N D V E	
3898	CTTGTTGCTGAGGGCAACAGCAGGTTC	ACTTACACTGTTCTTGTAGATGGCTGC	3951
	--+-----+-----+-----+	-----+-----+-----+---	
1300	LeuValAlaGluGlyAsnSerArgPhe	ThrTyrThrValLeuValAspGlyCys	1317
	L V A E G N S R F	T Y T V L V D G C	

**3712\_3954**

**Exon51 | Exon52 3955\_end of mRNA**

3952	TCTAAAAAGACAAATGAATGGGGAAAG	ACAATCATTGAATACAAAACAAATAAG	4005
	-----+-----+-----+	---+-----+-----+---	
1318	SerLysLysThrAsnGluTrpGlyLys	ThrIleIleGluTyrLysThrAsnLys	1335
	S K K T N E W G K	T I I E Y K T N K	
4006	CCATCACGCCTGCCCTTCCTTGATATT	GCACCTTTGGACATCGGTGGTGCTGAC	4059
	----+-----+-----+---	-----+-----+-----+---	
1336	ProSerArgLeuProPheLeuAspIle	AlaProLeuAspIleGlyGlyAlaAsp	1353
	P S R L P F L D I	A P L D I G G A D	

4060 CAGGAATTCTTTGTGGACATTGGCCCA GTCTGTTTCAAATAA 4101  
 +-----+-----+----- -+-----+  
 1354 GlnGluPhePheValAspIleGlyPro ValCysPheLysEnd 1367  
 Q E F F V D I G P V C F K

### 3' Untranslated region

\*1 ATGAACTCAA TCTAAATTAA AAAAGAAAGA AATTTGAAAA AACTTTCTCT \*50  
 -----+ -----+ -----+ -----+ -----+  
 \*51 TTGCCATTTT TTCTTCTTCT TTTTAACTG AAAGCTGAAT CCTTCCATTT \*100  
 -----+ -----+ -----+ -----+ -----+  
 \*101 CTTCTGCACA TCTACTTGCT TAAATTGTGG GCAAAAGAGA AAAAGAAGGA \*150  
 -----+ -----+ -----+ -----+ -----+  
 \*151 TTGATCAGAG CATTGTGCAA TACAGTTTCA TTAACTCCTT CCCCCGCTCC \*200  
 -----+ -----+ -----+ -----+ -----+  
 \*201 CCCAAAAATT TGAATTTTTT TTTCAACACT CTTACACCTG TTATGGAAAA \*250  
 -----+ -----+ -----+ -----+ -----+  
 \*251 TGTCAACCTT TGTAAGAAAA CAAAAATAAA AATTGAAAAA TAAAAACCAT \*300  
 -----+ -----+ -----+ -----+ -----+  
 \*301 AAACATTTGC ACCACTTGTG GCTTTTGAAT ATCTTCCACA GAGGGAAGTT \*350  
 -----+ -----+ -----+ -----+ -----+  
 \*351 TAAAACCCAA ACTTCCAAAG GTTTAAACTA CCTCAAAAACA CTTTCCCATG \*400  
 -----+ -----+ -----+ -----+ -----+  
 \*401 AGTGTGATCC ACATTGTTAG GTGCTGACCT AGACAGAGAT GAACTGAGGT \*450  
 -----+ -----+ -----+ -----+ -----+  
 \*451 CCTTGTTTTG TTTTGTTCAT AATACAAAGG TGCTAATTAA TAGTATTTCA \*500  
 -----+ -----+ -----+ -----+ -----+  
 \*501 GATACTTGAA GAATGTTGAT GGTGCTAGAA GAATTTGAGA AGAAATACTC \*550  
 -----+ -----+ -----+ -----+ -----+  
 \*551 CTGTATTGAG TTGTATCGTG TGGTGTATTT TTTAAAAAAT TTGATTTAGC \*600  
 -----+ -----+ -----+ -----+ -----+  
 \*601 ATTCATATTT TCCATCTTAT TCCAATTAA AAGTATGCAG ATTATTTGCC \*650  
 -----+ -----+ -----+ -----+ -----+  
 \*651 CAAATCTTCT TCAGATTCAG CATTGTCTTCT TTGCCAGTCT CATTTTCATC \*700  
 -----+ -----+ -----+ -----+ -----+  
 \*701 TTCTTCCATG GTTCCACAGA AGCTTTGTTT CTTGGGCAAG CAGAAAAATT \*750  
 -----+ -----+ -----+ -----+ -----+  
 \*751 AAATTGTACC TATTTTGTAT ATGTGAGATG TTTAAATAAA TTGTGAAAAA \*800  
 -----+ -----+ -----+ -----+ -----+  
 \*801 AATGAAATAA AGCATGTTTG GTTTTCCAAA AGAACATAT \*839  
 -----+ -----+ -----+ -----+ -----+