

# COL1A1 Systematic Numbering of cDNA and Protein Sequences

The reference cDNA used here is based on GenBank RefSeq NM\_000088.3. The corresponding reference genomic DNA sequence is GenBank RefSeqGene NG\_007400.1.

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 “legacy” fashion with the first glycine of the triple helix numbered as “1”. Legacy amino acid numbers can be calculated by subtracting 178 from the systematic number.

Version 3: July 2009

## 5' Untranslated Region

```

                                TCGTCG TAGCAGACGG GAGTTTCTCC
                                -126  ----- +----- +----- -101

TCGGGGTCGG AGCAGGAGGC ACGCGGAGTG TGAGGCCACG CATGAGCGGA
-100 +----- +----- +----- +----- +----- -51

CGCTAACCCC CTCCCCAGCC ACAAAGAGTC TACATGTCTA GGGTCTAGAC
-50  +----- +----- +----- +----- +----- -1

```

## Signal Peptide

```

1  ATGTTTCAGCTTTGTGGACCTCCGGCTC CTGCTCCTCTTAGCGGCCACCGCCCTC 54
   -----+-----+----- -+-----+-----+-----
1  MetPheSerPheValAspLeuArgLeu LeuLeuLeuLeuAlaAlaThrAlaLeu 18
   M F S F V D L R L L L L L A A T A L

55  CTGACGCACGGC 66
   -----+-----
19  LeuThrHisGly 22
   L T H G

```

## N-terminal Propeptide

```

                                1_103 exon1 | exon2 104_298
67  CAAGAGGAAGGCCAAGTCGAGGGCCAA GACGAAGACATCCCACCAATCACCTGC 120
   ---+-----+-----+----- -+-----+-----+-----+
23  GlnGluGluGlyGlnValGluGlyGln AspGluAspIleProProIleThrCys 40
   Q E E G Q V E G Q D E D I P P I T C

121 GTACAGAACGGCCTCAGGTACCATGAC CGAGACGTGTGGAAACCCGAGCCCTGC 174
   -----+-----+----- -+-----+-----+-----
41  ValGlnAsnGlyLeuArgTyrHisAsp ArgAspValTrpLysProGluProCys 58
   V Q N G L R Y H D R D V W K P E P C

```

175 CGGATCTGCGTCTGCGACAACGGCAAG GTGTTGTGCGATGACGTGATCTGTGAC 228  
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
59 ArgIleCysValCysAspAsnGlyLys ValLeuCysAspAspValIleCysAsp 76  
R I C V C D N G K V L C D D V I C D

229 GAGACCAAGAAGACTGCCCCGGCGCCGAA GTCCCCGAGGGCGAGTGCTGTCCCGTC 282  
-+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
77 GluThrLysAsnCysProGlyAlaGlu ValProGluGlyGluCysCysProVal 94  
E T K N C P G A E V P E G E C C P V

104\_298 exon2\_1 | exon3 299\_333 299\_333 exon3\_1 | exon4 334\_369  
283 TGCCCCGACGGCTCAGAGTCACCCACC GACCAAGAAACCACCGGCGTCGAGGGA 336  
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
95 CysProAspGlySerGluSerProThr AspGlnGluThrThrGlyValGluGly 112  
C P D G S E S P T D Q E T T G V E G

334\_369 exon4\_1 | exon5 370\_471  
337 CCCAAGGGGAGACTGGCCCCGAGGC CCAAGGGGACCCGAGGCCCCCTGGC 390  
----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
113 ProLysGlyAspThrGlyProArgGly ProArgGlyProAlaGlyProProGly 130  
P K G D T G P R G P R G P A G P P G

391 CGAGATGGCATCCCTGGACAGCTTGA CTTCCCGGACCCCCGGACCCCCGGA 444  
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
131 ArgAspGlyIleProGlyGlnProGly LeuProGlyProProGlyProProGly 148  
R D G I P G Q P G L P G P P G P P G

370\_471 exon5\_1 | exon6 472\_543  
445 CCTCCCGGACCCCTGGCCTCGGAGGA AACTTTGCTCCC 483  
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
149 ProProGlyProProGlyLeuGlyGly AsnPheAlaPro 161  
P P G P P G L G G N F A P

### **N-terminal Telopeptide**

484 CAGCTGTCTTATGGCTATGATGAGAAA TCAACCGGAGGAATTTCCGTGCCT 534  
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
162 GlnLeuSerTyrGlyTyrAspGluLys SerThrGlyGlyIleSerValPro 178  
Q L S Y G Y D E K S T G G I S V P

### **Triple Helix**

472\_543 exon6\_1 | exon7 544\_588 544\_588 exon7\_1  
535 GGCCCCATGGGTCCCTCTGGTCCCTCGT GGTCTCCCTGGCCCCCTGGTGCACCT 588  
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
179 GlyProMetGlyProSerGlyProArg GlyLeuProGlyProProGlyAlaPro 196  
1 G P M G P S G P R G L P G P P G A P 18

| exon8 589\_642 589\_642 exon8\_1  
589 GGTCCCCAAGGCTTCCAAGGTCCCCCT GGTGAGCCTGGCGAGCTTGAGCTTCA 642  
-+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
197 GlyProGlnGlyPheGlnGlyProPro GlyGluProGlyGluProGlyAlaSer 214  
19 G P Q G F Q G P P G E P G E P G A S 36

	<b>┌exon9 643_696</b>		<b>643_696 exon9┐</b>	
643	GGTCCCATGGGTCCCCGAGGTCCCCCA	GGTCCCCCTGGAAAAGAAATGGAGATGAT	696	
	-----+-----+-----	+-----+-----+-----		
215	GlyProMetGlyProArgGlyProPro	GlyProProGlyLysAsnGlyAspAsp	232	
37	G P M G P R G P P	G P P G K N G D D	54	
	<b>┌exon10 697_750</b>		<b>697_750 exon10┐</b>	
697	GGGGAAGCTGGAAAACCTGGTCGTCCT	GGTGAGCGTGGGCCTCCTGGGCCCTCAG	750	
	---+-----+-----+---	-----+-----+-----+---		
233	GlyGluAlaGlyLysProGlyArgPro	GlyGluArgGlyProProGlyProGln	250	
55	G E A G K P G R P	G E R G P P G P Q	72	
	<b>┌exon11 751_804</b>		<b>751_804 exon11┐</b>	
751	GGTGCTCGAGGATTGCCCGAACAGCT	GGCCTCCCTGGAATGAAGGGACACAGA	804	
	-----+-----+-----	---+-----+-----+---		
251	GlyAlaArgGlyLeuProGlyThrAla	GlyLeuProGlyMetLysGlyHisArg	268	
73	G A R G L P G T A	G L P G M K G H R	90	
	<b>┌exon12 805_858</b>		<b>805_858 exon12┐</b>	
805	GGTTTCAGTGGTTTGGATGGTGCCAAG	GGAGATGCTGGTCCTGCTGGTCCTAAG	858	
	-----+-----+-----+---	-----+-----+-----+---		
269	GlyPheSerGlyLeuAspGlyAlaLys	GlyAspAlaGlyProAlaGlyProLys	286	
91	G F S G L D G A K	G D A G P A G P K	108	
	<b>┌exon13 859_903</b>	<b>859_903 exon13┐</b>	<b>┌exon14 904_957</b>	
859	GGTGAGCCTGGCAGCCCTGGTGAAAAT	GGAGCTCCTGGTCAGATGGGCCCCCGT	912	
	-+-----+-----+-----	---+-----+-----+-----		
287	GlyGluProGlySerProGlyGluAsn	GlyAlaProGlyGlnMetGlyProArg	304	
109	G E P G S P G E N	G A P G Q M G P R	126	
		<b>904_957 exon14┐</b>	<b>┌exon15 958_1002</b>	
913	GGCCTGCCTGGTGAGAGAGGTTCGCCCT	GGAGCCCCTGGCCCTGCTGGTGCTCGT	966	
	-----+-----+-----+---	+-----+-----+-----+---		
305	GlyLeuProGlyGluArgGlyArgPro	GlyAlaProGlyProAlaGlyAlaArg	322	
127	G L P G E R G R P	G A P G P A G A R	144	
		<b>958_1002 exon15┐</b>	<b>┌exon16 1003_1056</b>	
967	GGAAATGATGGTGCTACTGGTGCTGCC	GGGCCCCCTGGTCCCACCGCCCCGCT	1020	
	---+-----+-----+---	-----+-----+-----+---		
323	GlyAsnAspGlyAlaThrGlyAlaAla	GlyProProGlyProThrGlyProAla	340	
145	G N D G A T G A A	G P P G P T G P A	162	
		<b>1003_1056 exon16┐</b>	<b>┌exon17 1057_1155</b>	
1021	GGTCCTCCTGGCTTCCCTGGTGCTGTT	GGTGCTAAGGGTGAAGCTGGTCCCCAA	1074	
	-----+-----+-----	---+-----+-----+---		
341	GlyProProGlyPheProGlyAlaVal	GlyAlaLysGlyGluAlaGlyProGln	358	
163	G P P G F P G A V	G A K G E A G P Q	180	
1075	GGGCCCCGAGGCTCTGAAGGTCCCCAG	GGTGTGCGTGGTGAGCCTGGCCCCCCT	1128	
	-----+-----+-----+---	-----+-----+-----+---		
359	GlyProArgGlySerGluGlyProGln	GlyValArgGlyGluProGlyProPro	376	
181	G P R G S E G P Q	G V R G E P G P P	198	

	<b>1057_1155 exon17</b>	<b>exon18 1156_1200</b>	
1129	GGCCCTGCTGGTGCTGCTGGCCCTGCT	GGAAACCCTGGTGCTGATGGACAGCCT	1182
	-+-----+-----+-----	-----+-----+-----+---	
377	GlyProAlaGlyAlaAlaGlyProAla	GlyAsnProGlyAlaAspGlyGlnPro	394
199	G P A G A A G P A	G N P G A D G Q P	216
	<b>1156_1200 exon18</b>	<b>exon19 1201_1299</b>	
1183	GGTGCTAAAGGTGCCAATGGTGCTCCT	GGTATTGCTGGTGCTCCTGGCTTCCCT	1236
	-----+-----+-----	+-----+-----+-----	
395	GlyAlaLysGlyAlaAsnGlyAlaPro	GlyIleAlaGlyAlaProGlyPhePro	412
217	G A K G A N G A P	G I A G A P G F P	234
1237	GGTGCCCGAGGCCCTCTGGACCCAG	GGCCCCGGCGCCCTCCTGGTCCCAAG	1290
	---+-----+-----+---	-----+-----+-----+---	
413	GlyAlaArgGlyProSerGlyProGln	GlyProGlyGlyProProGlyProLys	430
235	G A R G P S G P Q	G P G G P P G P K	252
<b>1201_1299</b>	<b>exon19</b>	<b>exon20 1300_1353</b>	
1291	GGTAACAGCGGTGAACCTGGTGCTCCT	GGCAGCAAAGGAGACACTGGTGCTAAG	1344
	-----+-----+-----	-----+-----+-----	
431	GlyAsnSerGlyGluProGlyAlaPro	GlySerLysGlyAspThrGlyAlaLys	448
253	G N S G E P G A P	G S K G D T G A K	270
<b>1300_1353</b>	<b>exon20</b>	<b>exon21 1354_1461</b>	
1345	GGAGAGCCTGGCCCTGTTGGTGTCAA	GGACCCCTGGCCCTGCTGGAGAGGAA	1398
	-----+-----+-----+---	-----+-----+-----	
449	GlyGluProGlyProValGlyValGln	GlyProProGlyProAlaGlyGluGlu	466
271	G E P G P V G V Q	G P P G P A G E E	288
1399	GGAAAGCGAGGAGCTCGAGGTGAACCC	GGACCCACTGGCCTGCCCGACCCCT	1452
	-+-----+-----+-----	-----+-----+-----+---	
467	GlyLysArgGlyAlaArgGlyGluPro	GlyProThrGlyLeuProGlyProPro	484
289	G K R G A R G E P	G P T G L P G P P	306
<b>1354_1461</b>	<b>exon21</b>	<b>exon22 1462_1515</b>	
1453	GGCGAGCGTGGTGGACCTGGTAGCCGT	GGTTTCCTGGCGCAGATGGTGTGCT	1506
	-----+-----+-----+---	+-----+-----+-----	
485	GlyGluArgGlyGlyProGlySerArg	GlyPheProGlyAlaAspGlyValAla	502
307	G E R G G P G S R	G F P G A D G V A	324
<b>1462_1515</b>	<b>exon22</b>	<b>exon23 1516_1614</b>	
1507	GGTCCCAAGGGTCCCCTGGTGAACGT	GGTTCTCCTGGCCCTGCTGGCCCCAAA	1560
	---+-----+-----+---	-----+-----+-----+---	
503	GlyProLysGlyProAlaGlyGluArg	GlySerProGlyProAlaGlyProLys	520
325	G P K G P A G E R	G S P G P A G P K	342
1561	GGATCTCCTGGTGAAGCTGGTCTGCC	GGTGAAGCTGGTCTGCCTGGTGCCAAG	1614
	-----+-----+-----+---	---+-----+-----+---	
521	GlySerProGlyGluAlaGlyArgPro	GlyGluAlaGlyLeuProGlyAlaLys	538
343	G S P G E A G R P	G E A G L P G A K	360
	<b>exon24 1615_1668</b>	<b>1615_1668 exon24</b>	
1615	GGTCTGACTGGAAGCCCTGGCAGCCCT	GGTCTGATGGCAAACTGGCCCCCT	1668
	-----+-----+-----+---	-----+-----+-----	
539	GlyLeuThrGlySerProGlySerPro	GlyProAspGlyLysThrGlyProPro	556
361	G L T G S P G S P	G P D G K T G P P	378

**exon25 1669\_1767**

1669 GGTCCCGCCGGTCAAGATGGTCGCCCC GGACCCCGAGGCCACCTGGTGCCCGT 1722  
 -+-----+-----+----- -+-----+-----+-----  
 557 GlyProAlaGlyGlnAspGlyArgPro GlyProProGlyProProGlyAlaArg 574  
 379 G P A G Q D G R P G P P G P P G A R 396

**1669\_1767 exon25 | exon26 1768\_1821**

1723 GGTCAAGGCTGGTGTGATGGGATTCCCT GGACCTAAAGGTGCTGCTGGAGAGCCC 1776  
 -----+-----+----- +-----+-----+-----  
 575 GlyGlnAlaGlyValMetGlyPhePro GlyProLysGlyAlaAlaGlyGluPro 592  
 397 G Q A G V M G F P G P K G A A G E P 414

**1768\_1821 exon26 | exon27 1822\_1875**

1777 GGCAAGGCTGGAGAGCGAGGTGTTCCC GGACCCCTGGCGCTGTCGGTCCCTGCT 1830  
 ---+-----+-----+----- -+-----+-----+-----  
 593 GlyLysAlaGlyGluArgGlyValPro GlyProProGlyAlaValGlyProAla 610  
 415 G K A G E R G V P G P P G A V G P A 432

**1822\_1875 exon27 | exon28 1876\_1929**

1831 GGCAAAGATGGAGAGGCTGGAGCTCAG GGACCCCTGGCCCTGCTGGTCCCGCT 1884  
 -----+-----+----- -+-----+-----+-----  
 611 GlyLysAspGlyGluAlaGlyAlaGln GlyProProGlyProAlaGlyProAla 628  
 433 G K D G E A G A Q G P P G P A G P A 450

**1876\_1929 exon28 | exon29 1930\_1983**

1885 GGCGAGAGAGGTGAACAAGGCCCTGCT GGCTCCCCGGATTCCAGGGTCTCCCT 1938  
 -----+-----+----- -+-----+-----+-----  
 629 GlyGluArgGlyGluGlnGlyProAla GlySerProGlyPheGlnGlyLeuPro 646  
 451 G E R G E Q G P A G S P G F Q G L P 468

**1930\_1983 exon29 | exon30 1984\_2028**

1939 GGTCTGCTGGTCTCCAGGTGAAGCA GGCAAACCTGGTGAACAGGGTGTTCCT 1992  
 -+-----+-----+----- -+-----+-----+-----  
 647 GlyProAlaGlyProProGlyGluAla GlyLysProGlyGluGlnGlyValPro 664  
 469 G P A G P P G E A G K P G E Q G V P 486

**1984\_2028 exon30 | exon31 2029\_2127**

1993 GGAGACCTTGGCGCCCTGGCCCCCTCT GGAGCAAGAGGCGAGAGAGGTTCCCT 2046  
 -----+-----+----- +-----+-----+-----  
 665 GlyAspLeuGlyAlaProGlyProSer GlyAlaArgGlyGluArgGlyPhePro 682  
 487 G D L G A P G P S G A R G E R G F P 504

2047 GGCGAGCGTGGTGTGCAAGGTCCCCCT GGTCTGCTGGTCCCCGAGGGGCCAAC 2100  
 ---+-----+-----+----- -+-----+-----+-----  
 683 GlyGluArgGlyValGlnGlyProPro GlyProAlaGlyProArgGlyAlaAsn 700  
 505 G E R G V Q G P P G P A G P R G A N 522

**2029\_2127 exon31 | exon32 2128\_2235**

2101 GGTGCTCCCGCAACGATGGTGCTAAG GGTGATGCTGGTGCCCTGGAGCTCCC 2154  
 -----+-----+----- -+-----+-----+-----  
 701 GlyAlaProGlyAsnAspGlyAlaLys GlyAspAlaGlyAlaProGlyAlaPro 718  
 523 G A P G N D G A K G D A G A P G A P 540

2155	GGTAGCCAGGGCGCCCCTGGCCTTCAG	GGAATGCCTGGTGAACGTGGTGCAGCT	2208
	-----+-----+-----+-----	-----+-----+-----+-----	
719	GlySerGlnGlyAlaProGlyLeuGln	GlyMetProGlyGluArgGlyAlaAla	736
541	G S Q G A P G L Q	G M P G E R G A A	558
	<b>2128_2235 exon32</b>	<b>rexon33/34 2236_2343</b>	
2209	GGTCTTCCAGGGCCTAAGGGTGACAGA	GGTGATGCTGGTCCCAAAGGTGCTGAT	2262
	-+-----+-----+-----+-----	-----+-----+-----+-----	
737	GlyLeuProGlyProLysGlyAspArg	GlyAspAlaGlyProLysGlyAlaAsp	754
559	G L P G P K G D R	G D A G P K G A D	576
2263	GGCTCTCCTGGCAAAGATGGCGTCCGT	GGTCTGACTGGCCCCATTGGTCTCCT	2316
	-----+-----+-----+-----	+-----+-----+-----+-----	
755	GlySerProGlyLysAspGlyValArg	GlyLeuThrGlyProIleGlyProPro	772
577	G S P G K D G V R	G L T G P I G P P	594
	<b>2236_2343 exon33/34</b>	<b>rexon35 2344_2397</b>	
2317	GGCCCTGCTGGTGCCCTGGTGACAAG	GGTGAAAGTGGTCCCAGCGGCCCTGCT	2370
	-----+-----+-----+-----	-----+-----+-----+-----	
773	GlyProAlaGlyAlaProGlyAspLys	GlyGluSerGlyProSerGlyProAla	790
595	G P A G A P G D K	G E S G P S G P A	612
	<b>2344_2397 exon35</b>	<b>rexon36 2398_2451</b>	
2371	GGTCCCACTGGAGCTCGTGGTGCCTCC	GGAGACCGTGGTGAGCCTGGTCCCTCC	2424
	-----+-----+-----+-----	-+-----+-----+-----+-----	
791	GlyProThrGlyAlaArgGlyAlaPro	GlyAspArgGlyGluProGlyProPro	808
613	G P T G A R G A P	G D R G E P G P P	630
	<b>2398_2451 exon36</b>	<b>rexon37 2452_2559</b>	
2425	GGCCCTGCTGGCTTTGCTGGCCCCCCT	GGTGCTGACGGCCAACCTGGTGCTAAA	2478
	-----+-----+-----+-----	-----+-----+-----+-----	
809	GlyProAlaGlyPheAlaGlyProPro	GlyAlaAspGlyGlnProGlyAlaLys	826
631	G P A G F A G P P	G A D G Q P G A K	648
2479	GGCGAACCTGGTGATGCTGGTGCTAAA	GGCGATGCTGGTCCCCCTGGCCCTGCC	2532
	-+-----+-----+-----+-----	-----+-----+-----+-----	
827	GlyGluProGlyAspAlaGlyAlaLys	GlyAspAlaGlyProProGlyProAla	844
649	G E P G D A G A K	G D A G P P G P A	666
	<b>2452_2559 exon37</b>	<b>rexon38 2560_2613</b>	
2533	GGACCCGCTGGACCCCTGGCCCCATT	GGTAATGTTGGTGCTCCTGGAGCCAAA	2586
	-----+-----+-----+-----	+-----+-----+-----+-----	
845	GlyProAlaGlyProProGlyProIle	GlyAsnValGlyAlaProGlyAlaLys	862
667	G P A G P P G P I	G N V G A P G A K	684
	<b>2560_2613 exon38</b>	<b>rexon39 2614_2667</b>	
2587	GGTGCTCGCGGCAGCGCTGGTCCCCCT	GGTGCTACTGGTTCCCTGGTGCTGCT	2640
	---+-----+-----+-----+-----	-----+-----+-----+-----+-----	
863	GlyAlaArgGlySerAlaGlyProPro	GlyAlaThrGlyPheProGlyAlaAla	880
685	G A R G S A G P P	G A T G F P G A A	702
	<b>2614_2667 exon39</b>	<b>rexon40 2668_2829</b>	
2641	GGCCGAGTCGGTCTCCTGGCCCCCTCT	GGAAATGCTGGACCCCTGGCCCCCTCT	2694
	-----+-----+-----+-----	-+-----+-----+-----+-----	
881	GlyArgValGlyProProGlyProSer	GlyAsnAlaGlyProProGlyProPro	898
703	G R V G P P G P S	G N A G P P G P P	720

2695	GGTCCTGCTGGCAAAGAAGGCGGCAAA	GGTCCCCGTGGTGAGACTGGCCCTGCT	2748
	-----+-----+-----+-----	-----+-----+-----+-----	
899	GlyProAlaGlyLysGluGlyGlyLys	GlyProArgGlyGluThrGlyProAla	916
721	G P A G K E G G K	G P R G E T G P A	738
2749	GGACGTCCTGGTGAAGTTGGTCCCCCT	GGTCCCCCTGGCCCTGCTGGCGAGAAA	2802
	-+-----+-----+-----+-----	-+-----+-----+-----+-----	
917	GlyArgProGlyGluValGlyProPro	GlyProProGlyProAlaGlyGluLys	934
739	G R P G E V G P P	G P P G P A G E K	756
	<b>2668_2829 exon40</b>   <b>exon41 2830_2937</b>		
2803	GGATCCCCTGGTGCTGATGGTCCTGCT	GGTGCTCCTGGTACTCCCGGGCCTCAA	2856
	-----+-----+-----+-----	+-----+-----+-----+-----	
935	GlySerProGlyAlaAspGlyProAla	GlyAlaProGlyThrProGlyProGln	952
757	G S P G A D G P A	G A P G T P G P Q	774
2857	GGTATTGCTGGACAGCGTGGTGTGGTC	GGCCTGCCTGGTCAGAGAGGAGAGAGA	2910
	---+-----+-----+-----+-----	---+-----+-----+-----+-----	
953	GlyIleAlaGlyGlnArgGlyValVal	GlyLeuProGlyGlnArgGlyGluArg	970
775	G I A G Q R G V V	G L P G Q R G E R	792
	<b>2830_2937 exon41</b>   <b>exon42 2938_3045</b>		
2911	GGCTTCCCTGGTCTTCCTGGCCCCCTCT	GGTGAACCTGGCAAACAAGGTCCCTCT	2964
	-----+-----+-----+-----	-+-----+-----+-----+-----	
971	GlyPheProGlyLeuProGlyProSer	GlyGluProGlyLysGlnGlyProSer	988
793	G F P G L P G P S	G E P G K Q G P S	810
2965	GGAGCAAGTGGTGAACGTGGTCCCCCT	GGTCCCATGGGCCCCCTGGATTGGCT	3018
	-----+-----+-----+-----	-----+-----+-----+-----	
989	GlyAlaSerGlyGluArgGlyProPro	GlyProMetGlyProProGlyLeuAla	1006
811	G A S G E R G P P	G P M G P P G L A	828
	<b>2938_3045 exon42</b>   <b>exon43 3046_3099</b>		
3019	GGACCCCCTGGTGAATCTGGACGTGAG	GGGGCTCCTGGTGCCGAAGGTTCCCCT	3072
	-+-----+-----+-----+-----	-+-----+-----+-----+-----	
1007	GlyProProGlyGluSerGlyArgGlu	GlyAlaProGlyAlaGluGlySerPro	1024
829	G P P G E S G R E	G A P G A E G S P	846
	<b>3046_3099 exon43</b>   <b>exon44 3100_3207</b>		
3073	GGACGAGACGGTTCTCCTGGCGCCAAG	GGTGACCGTGGTGAGACCGCCCCGCT	3126
	-----+-----+-----+-----	+-----+-----+-----+-----	
1025	GlyArgAspGlySerProGlyAlaLys	GlyAspArgGlyGluThrGlyProAla	1042
847	G R D G S P G A K	G D R G E T G P A	864
3127	GGACCCCCTGGTGCTCCTGGTGCTCCT	GGTGCCCCTGGCCCCGTTGGCCCTGCT	3180
	---+-----+-----+-----+-----	---+-----+-----+-----+-----	
1043	GlyProProGlyAlaProGlyAlaPro	GlyAlaProGlyProValGlyProAla	1060
865	G P P G A P G A P	G A P G P V G P A	882
	<b>3100_3207 exon44</b>   <b>exon45 3208_3261</b>		
3181	GGCAAGAGTGGTGATCGTGGTGAGACT	GGTCTGCTGGTCCC GCCGTCCTGTCT	3234
	-----+-----+-----+-----	-+-----+-----+-----+-----	
1061	GlyLysSerGlyAspArgGlyGluThr	GlyProAlaGlyProAlaGlyProVal	1078
883	G K S G D R G E T	G P A G P A G P V	900

	<b>3208_3261 exon45</b>	<b>exon46 3262_3369</b>	
3235	GGCCCTGTTGGCGCCCGTGGCCCCGCC	GGACCCCAAGGCCCCCGTGGTGACAAG	3288
	-----+-----+-----+-----	-----+-----+-----+-----	
1079	GlyProValGlyAlaArgGlyProAla	GlyProGlnGlyProArgGlyAspLys	1096
901	G P V G A R G P A	G P Q G P R G D K	918
3289	GGTGAGACAGGCGAACAGGGCGACAGA	GGCATAAAGGGTCACCGTGGCTTCTCT	3342
	-+-----+-----+-----	-----+-----+-----+-----	
1097	GlyGluThrGlyGluGlnGlyAspArg	GlyIleLysGlyHisArgGlyPheSer	1114
919	G E T G E Q G D R	G I K G H R G F S	936
	<b>3262_3369 exon46</b>	<b>exon47 3370_3423</b>	
3343	GGCCTCCAGGGTCCCCCTGGCCCTCCT	GGCTCTCCTGGTGAACAAGGTCCCTCT	3396
	-----+-----+-----	+-----+-----+-----	
1115	GlyLeuGlnGlyProProGlyProPro	GlySerProGlyGluGlnGlyProSer	1132
937	G L Q G P P G P P	G S P G E Q G P S	954
	<b>3370_3423 exon47</b>	<b>exon48 3424_3531</b>	
3397	GGAGCCTCTGGTCTGCTGGTCCCCGA	GGTCCCCCTGGCTCTGCTGGTGTCTCT	3450
	---+-----+-----+-----	-----+-----+-----+-----	
1133	GlyAlaSerGlyProAlaGlyProArg	GlyProProGlySerAlaGlyAlaPro	1150
955	G A S G P A G P R	G P P G S A G A P	972
3451	GGCAAAGATGGACTCAACGGTCTCCCT	GGCCCCATTGGGCCCCCTGGTCTCGC	3504
	-----+-----+-----	---+-----+-----+-----	
1151	GlyLysAspGlyLeuAsnGlyLeuPro	GlyProIleGlyProProGlyProArg	1168
973	G K D G L N G L P	G P I G P P G P R	990
	<b>3424_3531 exon48</b>	<b>exon49 3532_3814</b>	
3505	GGTCGCACTGGTGATGCTGGTCTGTT	GGTCCCCCGGCCCTCCTGGACCTCCT	3558
	-----+-----+-----	-----+-----+-----	
1169	GlyArgThrGlyAspAlaGlyProVal	GlyProProGlyProProGlyProPro	1186
991	G R T G D A G P V	G P P G P P G P P	1008
3559	GGTCCCCCTGGTCTCCCTCC		3576
	-+-----+-----		
1187	GlyProProGlyProPro		1192
1009	G P P G P P		1014

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

3577	AGCGCTGGTTTTCGACTTCAGCTTCTTG	CCCCAGCCACCTCAAGAGAAGGCTCAC	3630
	---+-----+-----+-----	-----+-----+-----+-----	
1193	SerAlaGlyPheAspPheSerPheLeu	ProGlnProProGlnGluLysAlaHis	1210
	S A G F D F S F L	P Q P P Q E K A H	
3631	GATGGTGGCCGCTACTACCGGGCT		3654
	-----+-----+-----		
1211	AspGlyGlyArgTyrTyrArgAla		1218
	D G G R Y Y R A		



## C-terminal Propeptide

3655	GATGATGCCAATGTGGTTTCGTGACCGT	GACCTCGAGGTGGACACCACCCTCAAG	3708
	-----+-----+-----+-----	-----+-----+-----	
1219	AspAspAlaAsnValValArgAspArg	AspLeuGluValAspThrThrLeuLys	1236
	D D A N V V R D R	D L E V D T T L K	
3709	AGCCTGAGCCAGCAGATCGAGAACATC	CGGAGCCCAGAGGGCAGCCGCAAGAAC	3762
	-+-----+-----+-----	----+-----+-----+---	
1237	SerLeuSerGlnGlnIleGluAsnIle	ArgSerProGluGlySerArgLysAsn	1254
	S L S Q Q I E N I	R S P E G S R K N	
		<b>3532_3814 exon49   exon50 3815_4005</b>	
3763	CCCCCCCCCACCTGCCGTGACCTCAAG	ATGTGCCACTCTGACTGGAAGAGTGGA	3816
	-----+-----+-----+-----	+-----+-----+-----	
1255	ProAlaArgThrCysArgAspLeuLys	MetCysHisSerAspTrpLysSerGly	1272
	P A R T C R D L K	M C H S D W K S G	
3817	GAGTACTGGATTGACCCCAACCAAGGC	TGCAACCTGGATGCCATCAAAGTCTTC	3870
	---+-----+-----+-----	-----+-----+-----+---	
1273	GluTyrTrpIleAspProAsnGlnGly	CysAsnLeuAspAlaIleLysValPhe	1290
	E Y W I D P N Q G	C N L D A I K V F	
3871	TGCAACATGGAGACTGGTGAGACCTGC	GTGTACCCCACTCAGCCCAGTGTGGCC	3924
	-----+-----+-----+-----	--+-----+-----+-----	
1291	CysAsnMetGluThrGlyGluThrCys	ValTyrProThrGlnProSerValAla	1308
	C N M E T G E T C	V Y P T Q P S V A	
3925	CAGAAGAACTGGTACATCAGCAAGAAC	CCCAAGGACAAGAGGCATGTCTGGTTC	3978
	-----+-----+-----+-----	-----+-----+-----+-----	
1309	GlnLysAsnTrpTyrIleSerLysAsn	ProLysAspLysArgHisValTrpPhe	1326
	Q K N W Y I S K N	P K D K R H V W F	
		<b>3815_4005 exon50   exon51 4006_4248</b>	
3979	GGCGAGAGCATGACCGATGGATTCCAG	TTTCGAGTATGGCGGCCAGGGCTCCGAC	4032
	-+-----+-----+-----	----+-----+-----+---	
1327	GlyGluSerMetThrAspGlyPheGln	PheGluTyrGlyGlyGlnGlySerAsp	1344
	G E S M T D G F Q	F E Y G G Q G S D	
4033	CCTGCCGATGTGGCCATCCAGCTGACC	TTCCTGCGCCTGATGTCCACCGAGGCC	4086
	-----+-----+-----+-----	+-----+-----+-----+---	
1345	ProAlaAspValAlaIleGlnLeuThr	PheLeuArgLeuMetSerThrGluAla	1362
	P A D V A I Q L T	F L R L M S T E A	
4087	TCCAGAACATCACCTACCACTGCAAG	AACAGCGTGGCCTACATGGACCAGCAG	4140
	---+-----+-----+-----	-----+-----+-----+---	
1363	SerGlnAsnIleThrTyrHisCysLys	AsnSerValAlaTyrMetAspGlnGln	1380
	S Q N I T Y H C K	N S V A Y M D Q Q	
4141	ACTGGCAACCTCAAGAAGGCCCTGCTC	CTCCAGGGCTCCAACGAGATCGAGATC	4194
	-----+-----+-----+-----	--+-----+-----+-----	
1381	ThrGlyAsnLeuLysLysAlaLeuLeu	LeuGlnGlySerAsnGluIleGluIle	1398
	T G N L K K A L L	L Q G S N E I E I	

**4006\_4248 exon51**

```

4195  CGCGCCGAGGGCAACAGCCGCTTCACC TACAGCGTCACTGTCGATGGCTGCACG  4248
-----+-----+-----+-----+-----+-----+-----+-----+
1399  ArgAlaGluGlyAsnSerArgPheThr TyrSerValThrValAspGlyCysThr  1416
R  A  E  G  N  S  R  F  T  Y  S  V  T  V  D  G  C  T

┌exon52 4249_end of mRNA
4249  AGTCACACCCGGAGCCTGGGGCAAGACA GTGATTGAATACAAAACCACCAAGACC  4302
-+-----+-----+-----+-----+-----+-----+-----+
1417  SerHisThrGlyAlaTrpGlyLysThr ValIleGluTyrLysThrThrLysThr  1434
S  H  T  G  A  W  G  K  T  V  I  E  Y  K  T  T  K  T

4303  TCCCGCCTGCCCATCATCGATGTGGCC CCCTTGGACGTTGGTGCCCCAGACCAG  4356
-----+-----+-----+-----+-----+-----+-----+
1435  SerArgLeuProIleIleAspValAla ProLeuAspValGlyAlaProAspGln  1452
S  R  L  P  I  I  D  V  A  P  L  D  V  G  A  P  D  Q

4357  GAATTCGGCTTCGACGTTGGCCCTGTC TGCTTCCTGTAA  4395
----+-----+-----+-----+-----+-----+
1453  GluPheGlyPheAspValGlyProVal CysPheLeuEnd  1465
E  F  G  F  D  V  G  P  V  C  F  L

```

### 3' Untranslated region

```

*1  ACTCCCTCCA TCCCAACCTG GCTCCCTCCC ACCCAACCAA CTTTCCCCCC *50
-----+-----+-----+-----+-----+
*51  AACCCGGAAA CAGACAAGCA ACCCAAAGTG AACCCCTCA AAAGCCAAAA *100
-----+-----+-----+-----+-----+
*101 AATGGGAGAC AATTTACAT GGACTTTGGA AAATATTTTT TTCCTTTGCA *150
-----+-----+-----+-----+-----+
*151 TTCATCTCTC AAACTTAGTT TTTATCTTTG ACCAACCGAA CATGACCAAA *200
-----+-----+-----+-----+-----+
*201 AACCAAAAGT GCATTCAACC TTACCAAAAA AAAAAAAAAA AAAAGAATAA *250
-----+-----+-----+-----+-----+
*251 ATAAATAACT TTTTAAAAAA GGAAGCTTGG TCCACTTGCT TGAAGACCCA *300
-----+-----+-----+-----+-----+
*301 TGCGGGGGTA AGTCCCTTTC TGCCCCTTGG GCTTATGAAA CCCCAATGCT *350
-----+-----+-----+-----+-----+
*351 GCCCTTTCTG CTCCTTTCTC CACACCCCCC TTGGGGCCTC CCCTCCACTC *400
-----+-----+-----+-----+-----+
*401 CTTCCCAAAT CTGTCTCCCC AGAAGACACA GGAAACAATG TATTGTCTGC *450
-----+-----+-----+-----+-----+
*451 CCAGCAATCA AAGGCAATGC TCAAACACCC AAGTGGCCCC CACCCCTCAGC *500
-----+-----+-----+-----+-----+
*501 CCGCTCCTGC CCGCCCAGCA CCCCCAGGCC CTGGGGGACC TGGGGTTCTC *550
-----+-----+-----+-----+-----+

```

```

*551  AGACTGCCAA AGAAGCCTTG CCATCTGGCG CTCCCATGGC TCTTGCAACA *600
-----+ -----+ -----+ -----+ -----+

*601  TCTCCCCTTC GTTTTTTGAGG GGGTCATGCC GGGGGAGCCA CCAGCCCCTC *650
-----+ -----+ -----+ -----+ -----+

*651  ACTGGGTTTCG GAGGAGAGTC AGGAAGGGCC ACGACAAAGC AGAAACATCG *700
-----+ -----+ -----+ -----+ -----+

*701  GATTTGGGGA ACGCGTGTCA ATCCCTTGTG CCGCAGGGCT GGGCGGGAGA *750
-----+ -----+ -----+ -----+ -----+

*751  GACTGTTCTG TTCCTTGTGT AACTGTGTTG CTGAAAGACT ACCTCGTTCT *800
-----+ -----+ -----+ -----+ -----+

*801  TGTCTTGATG TGTCACCGGG GCAACTGCCT GGGGGCGGGG ATGGGGGCAG *850
-----+ -----+ -----+ -----+ -----+

*851  GGTGGAAGCG GCTCCCCATT TTATACCAAA GGTGCTACAT CTATGTGATG *900
-----+ -----+ -----+ -----+ -----+

*901  GGTGGGGTGG GGAGGGAATC ACTGGTGCTA TAGAAATTGA GATGCCCCCC *950
-----+ -----+ -----+ -----+ -----+

*951  CAGGCCAGCA AATGTTCCCT TTTGTTCAAA GTCTATTTTT ATTCCCTGAT *1000
-----+ -----+ -----+ -----+ -----+

*1001 ATTTTTCTTT TTTTTTTTTT TTTTTTGTGG ATGGGGACTT GTGAATTTTT *1050
-----+ -----+ -----+ -----+ -----+

*1051 CTAAAGGTGC TATTTAACAT GGGAGGAGAG CGTGTGCGGC TCCAGCCCAG *1100
-----+ -----+ -----+ -----+ -----+

*1101 CCCGCTGCTC ACTTTCCACC CTCTCTCCAC CTGCCTCTGG CTTCTCAGGC *1150
-----+ -----+ -----+ -----+ -----+

*1151 CTCTGCTCTC CGACCTCTCT CCTCTGAAAC CCTCCTCCAC AGCTGCAGCC *1200
-----+ -----+ -----+ -----+ -----+

*1201 CATCCTCCCG GCTCCCTCCT AGTCTGTCC TCGTCCCTCTG TCCCCGGGTT *1250
-----+ -----+ -----+ -----+ -----+

*1251 TCAGAGACAA CTTCCCAAAG CACAAAGCAG TTTTTCCTCC TAGGGGTGGG *1300
-----+ -----+ -----+ -----+ -----+

*1301 AGGAAGCAAA AGACTCTGTA CCTATTTTGT ATGTGTATAA TAATTTGAGA *1350
-----+ -----+ -----+ -----+ -----+

*1351 TGTTTTTAAT TATTTTGATT GCTGGAATAA AGCATGTGGA AATGACCCAA *1400
-----+ -----+ -----+ -----+ -----+

*1401 ACATAA *1406
-----

```