**To test my speculation that the disulfide (C-C) domains of this protein are each encoded by one of the 22 exons of the gene, I have looked up the translation products of the CFH exons in ENSEMBL. A diagram of the 22 CFH exons on chromosome 1 appears below: **

**The translated products of each of the CFH exons are listed in the next section. Alternate exons are in blue. The red amino acids at the end of some sequences are amino acids encoded by a split codon -- one that starts in one exon and ends in the next. In all cases, these amino acids have been included with translation of the starting exon. The Cysteine units (C) that form the disulfide bonds in each domain are in larger font. The first line of sequence is the signal sequence used to mark the protein for secretion from the cell. The C-C domains begin with the second line and are numbered. Note that the second C-C domain is split between two smaller exons. In all other cases the beginning and ending C of the C-C domain are encoded in the same exon.**

**Signal: MRLLAKIICLMLWAICVAED**

**1. CNELPPRRNTEILTGSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLRKCQK**

**2. RPCGHPGDTPFGTFTLTGGNVFEYGVKAVYTCNEG**

**2. YQLLGEINYRECDTDGWTNDIPICEV**

**3. VKCLPVTAPENGKIVSSAMEPDREYHFGQAVRFVCNSGYKIEGDEEMHCSDDGFWSKEKPKCVE**

**4. ISCKSPDVINGSPISQKIIYKENERFQYKCNMGYEYSERGDAVCTESGWRPLPSCEE**

**5. KSCDNPYIPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGNTAKCTSTGWIPAPRCTL**

**6. KPCDYPDIKHGGLYHENMRRPYFPVAVGKYYSYYCDEHFETPSGSYWDHIHCTQDGWSPAVPCLR**

**7. KCYFPYLENGYNQNHGRKFVQGKSIDVACHPGYALPKAQTTVTCMENGWSPTPRCIRVK**

**8. TCSKSSIDIENGFISESQYTYALKEKAKYQCKLGYVTADGETSGSITCGKDGWSAQPTCIK**

**9. SCDIPVFMNARTKNDFTWFKLNDTLDYECHDGYESNTGSTTGSIVCGYNGWSDLPICYE**

**10. RECELPKIDVHLVPDRKKDQYKVGEVLKFSCKPGFTIVGPNSVQCYHFGLSPDLPICKE**

**11. QVQSCGPPPELLNGNVKEKTKEEYGHSEVVEYYCNPRFLMKGPNKIQCVDGEWTTLPVCIV**

**12. EESTCGDIPELEHGWAQLSSPPYYYGDSVEFNCSESFTMIGHRSITCIHGVWTQLPQCVA**

**13. IDKLKKCKSSNLIILEEHLKNKKEFDHNSNIRYRCRGKEGWIHTVCINGRWDPEVNCSM**

**14. AQIQLCPPPPQIPNSHNMTTTLNYRDGEKVSVLCQENYLIQEGEEITCKDGRWQSIPLCVE**

**15. KIPCSQPPQIEHGTINSSRSSQESYAHGTKLSYTCEGGFRISEENETTCYMGKWSSPPQCEG**

**16. LPCKSPPEISHGVVAHMSDSYQYGEEVTYKCFEGFGIDGPAIAKCLGEKWSHPPSCIK**

**17. TDCLSLPSFENAIPMGEKKDVYKAGEQVTYTCATYYKMDGASNVTCINSRWTGRPTCRD**

**18. TSCVNPPTVQNAYIVSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCLNGNWTEPPQCKD**

**19. STGKCGPPPPIDNGDITSFPLSVYAPASSVEYQCQNLYQLEGNKRITCRNGQWSEPPKCLH**

**20. PCVISREIMENYNIALRWTAKQKLYSRTGESVEFVCKRGYRLSSRSHTLRTTCWDGKLEYPTCAKR**

[**http://useast.ensembl.org/Homo\_sapiens/Transcript/Sequence\_Protein?db=core;g=ENSG00000000971;r=1:196651878-196747504;t=ENST00000367429**](http://useast.ensembl.org/Homo_sapiens/Transcript/Sequence_Protein?db=core;g=ENSG00000000971;r=1:196651878-196747504;t=ENST00000367429%20%20%20%20%20)