

An analysis of base frequencies in the anti-sense strands corresponding to the 180 human protein coding sequences

C.-T. Zhang¹ and K.-C. Chou²

¹Department of Physics, Tianjin University, Tianjin, China ²Computer-Aided Drug Discovery, Upjohn Laboratories, Kalamazoo, Michigan, U.S.A.

Accepted August 22, 1995

Summary. Can the anti-sense chain of DNA encode for a protein? Such a problem has been explored by means of the codon-analyzing graph developed recently.

Keywords: Amino acids – Sense chain – Anti-sense chain – Graphic expression – (G-non-G-N)_n pattern

Introduction

As is well-known, the DNA double helix consists of two strands being anti-parallel with each other. One strand forms the sense chain, and the complementary strand anti-parallel with it forms the anti-sense chain. Usually, it is thought that the main function of the anti-sense chain is to form a structure of double helix. Consider a coding sequence in the sense strand and the anti-sense sequence complementary to it. The latter usually can not encode for a protein. The first reason is that the anti-sense strand does not contain a promoter and other necessary sequence segment. The second reason is that once any one of the three codons – TTA, CTA, and TCA – appears in the interior of the sense-chain sequence, then one of the complementary codons - TAA, TAG, and TGA - must occur in the interior of anti-sense chain sequence, and these codons are none but a terminator. Usually the codons TTA, CTA and TCA may occur many times in a coding sequence. This implies that, there are many terminator codons existing in the corresponding anti-sense chain. A sequence with many terminators can of course not be used to encode for a protein. The third reason is that the last codon in the coding sequence of a sense chain is generally not CAT, therefore, the first codon in the anti-sense chain will not be the codon ATG, the initial codon usually required for a coding sequence. All of these have implied that it seems impossible for the anti-sense sequence complementary to the coding sequence

to encoding for a protein. Nevertheless, it is still worthwhile to study this problem at a deeper level.

Let us consider this problem from a pure theoretical point of view. Our question is: If there are no stop codons within the interior of the anti-sense sequences, can they encode for proteins? We shall explore this problem by observing the distribution of the frequencies of bases at each of the three codon positions. Of 2681 human protein coding sequences (Wata et al., 1991), we have found 180 coding sequences which do not contain any of the codons TTA, CTA and TCA. The names of such 180 coding sequences in the GenBank (Wada et al., 1991) are given in Table 1. This means the corresponding 180 anti-sense chains will not be divided by stop codons. Below, we shall use the symmetrical property of the codon-analyzing graph developed recently (Zhang and Zhang, 1991; Chou and Zhang, 1992; Zhang and Chou, 1994) to calculate the distribution of the base frequencies at each of the three codon positions for these anti-sense sequences.

Table 1. The 180 human protein coding sequences whose anti-sense sequences are not compartmentalized by stop codons

HUM18D HUM18U HUMACALX HUMACTASK HUMADRA HUMADRA2C HUMADRA2R HUMALPR HUMBCL2A HUMBCL2B HUMBCL2C HUMBETGLA HUMBM40 HUMCACY HUMCACYA HUMCACYA HUMCACYA HUMCACYA HUMCACYA HUMCACYA HUMCACYA HUMCCK3* HUMCNP HUMCNTFR HUMCOA2IT HUMCSH2** HUMCRYGBC**2 HUMCRYGQ2** HUMCRYGQ4** HUMCCYCPS3 HUMCSFIM3** HUMCSIST HUMCTRP HUMCVIB HUMCYCPS3 HUMCSFIM3** HUMFABPL HUMFABPLA HUMFKMKA HUMFSH3** HUMFSHB03** HUMGOS2A HUMGOS2P HUMGRYG2P2** HUMGRNAS6#1** HUMGNAS6#2** HUMGROB HUMGROB5** HUMGRSA2R HUMGROB HUMGROB5** HUMGROB5** HUMGSA2R HUMGROB HUMGROB5** HUMHBB#3 HUMHBB#5* HUMHBBAAZ HUMHBNFI HUMFEDOB HUMHISAB** HUMHISABPM HUMHISAB HUMHISAC HUMHISH2B HUMHISH3A HUMHISH4* HUMHMG17G HUMHMG1 HUMHMG1A HUMHMG1B HUMHMG174 HUMHMG7P HUMHMG7P3** HUMGHAAB HUMHMG7P HUMHMG1A HUMHMG1B HUMHMG1A HUMHMG1B HUMHMG1A HUMHMG1					
HUMBCL2C HUMBETGLA HUMBM40 HUMCACY HUMCACY HUMCACY HUMCACY HUMCACS* HUMCNP HUMCNTFR HUMCOA2IT HUMCSFIM3* HUMCSST HUMCRYGQ2* HUMCRYGQ4* HUMCRYGX4* HUMCYCPS3 HUMCYS3A3* HUMFABPL HUMFABPL HUMFABPLA HUMFKMKA HUMFSHBQ3* HUMGNAS6#1* HUMGNAS6#2* HUMGNAS6#1* HUMGNAS6#2* HUMGRPS#2* HUMGRAS6#1* HUMGNAS6#1* HUMGNAS6#1* HUMGRBFBA HUMGRBBH HUMGROB HUMGROB HUMGROB HUMGROB HUMGROB HUMHBB#3 HUMHBB#3 HUMHBB#5 HUMHBBAAZ HUMHBB#1 HUMHISAB HUMHISAC HUMHISAB HUMHISAC HUMHISHA HUMHISAB HUMHISAC HUMHISHA HUMHISAB HUMHISAC HUMHMGIA HUMHMGIA HUMHMGIA HUMHMGIA HUMHMGIA HUMHMGIA HUMHMGYA HUMHMOTOB HUMHMGYA HUMHMGYA HUMHMGYA HUMHMGYA HUMHMGYA HUMHMGYA HUMHMCACY HUMHMGYA HUMHMGYA HUMHMGYA HUMHBB#1 HUMHISAB HUMHISAC HUMHMGYA HUMHMGYA HUMHMGYA HUMHMGYA HUMHMGYA HUMHMGYA HUMHMGYA HUMHMOYA HUMHMGYA HUMHMOYA HUMHMOYA HUMHMOYA HUMHMOYA HUMHMOYA HUMHOXA HUMMICA HUMHOXA HUMMICA HUMICA HUMMICA HUMHICA HUMHICA HUMHICA HUMHICA HUMHICA HUMHICA HUMHI	HUM18D	HUM18U	HUMACALX	HUMACTASK	HUMADRA
HUMCAL HUMCAL HUMCAL HUMCAL HUMCAL HUMCCK3* HUMCNP HUMCNTFR HUMCOA2IT HUMCRPBP2* HUMCRYGO4* HUMCRYGCX4* HUMCRYGAX4* HUMCSIST HUMCTRP HUMCVIB HUMCYCPS3 HUMCSSST HUMFABPL HUMFABPLA HUMGOS2PE HUMGBR#1 HUMGSAS6#1* HUMGNAS6#1* HUMGNAS6#2* HUMGNAS6#1* HUMGROB HUMGROB5 HUMGROG5 HUMGROB5 HUMGROG5 HUMGROB5 HUMHBB#1 HUMHMGI7G HUMHMGI HUMHMGIA HUMHMGIB HUMHMGYA HUMHMGT2 HUMMGT2 HUMMGT2 HUMMGT2 HUMMMGT2 HUMMMGT2 HUMMMGT2 HUMMMGT2 HUMMMCT2 HUMMMTA HUMMHDC3B HUMMHDC3B HUMMHDC3B HUMMHDC3B HUMMHDCC HUMMHDCA#2 HUMMHDCBH HUMHBBH HUM	HUMADRA2C	HUMADRA2R	HUMALPR	HUMBCL2A	HUMBCL2B
HUMCCK3* HUMCNP HUMCNTFR HUMCOA2IT HUMCRBP2* HUMCRYGBC#2 HUMCRYGQ2* HUMCRYGQ4* HUMCRYGX4* HUMCSFIM3* HUMCSIST HUMCTRP HUMCVIB HUMCYCPS3 HUMCYSA3A3* HUMFABPL HUMFABPLA HUMFKMKA HUMFSH3* HUMFSHB2* HUMFSHBQ3* HUMG0S2A HUMG0S2PE HUMGBR#1 HUMGFI2A9 HUMGNAS6#1* HUMGNAS6#2* HUMGNAS6#3* HUMGROB5 HUMGROB5 HUMGROB5 HUMGROB5 HUMGROB6AIR HUMGSAIR HUMHISAB HUMHISAC HUMHISHB HUMHISAB HUMHISAC HUMHISHB HUMHISAB HUMHISAC HUMHISHB HUMHISAB HUMHISAC HUMHISHB HUMHISHAA HUMHISHA HUMHISAB HUMHISAC HUMHISHB HUMHISHAA HUMHISHA HUMHISAB HUMHISAC HUMHMGIB HUMHMGI HUMHMGIA HUMHMGIB HUMHMGI HUMHMGIA HUMHISAB HUMHISAB HUMHISAB HUMHISAB HUMHISAB HUMHISAB HUMHISHA HUMHISHA HUMHISHA HUMHISHA HUMHISHA HUMHISAB HUMHISAB HUMHISHA HUMMISHA	HUMBCL2C	HUMBETGLA	HUMBM40	HUMCACY	HUMCACYA
HUMCRYGBC#2 HUMCRYGQ2* HUMCYB HUMCYGS3 HUMCSST HUMCTRP HUMCVIB HUMCYGS3 HUMCYS3A3* HUMCSSB1* HUMFABPL HUMFABPLA HUMFKMKA HUMFSHB3* HUMGS2A HUMG0S2A HUMG0S2P HUMGRB#1 HUMGREP5#1* HUMGRF5#2* HUMGRO HUMGROB HUMGROB HUMGROB HUMGROB HUMGROB HUMGROB HUMGROB HUMGROB HUMGROB HUMHBB#3 HUMHBB#5 HUMHBB#3 HUMHBB#5 HUMHBBAAZ HUMHBB#1 HUMHBB#3 HUMHISAB HUMHISAB HUMHISAC HUMHISHB HUMHISHB HUMHISAB HUMHISAC HUMHISHB HUMHMGIB HUMHMGIB HUMHMGYD HUMHMGYD HUMHOX14 HUMHP2AA HUMHST HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGHBP1 HUMIGFB HUMIGBB HUMHIST HUMIGFB HUMICC HUMMET2 HUMMET2 HUMMET2B HUMMETE HUMMET2 HUMMET2B HUMMETE HUMMET HUMMET HUMMHDC3B HUMMHDR3A HUMHIDRB HUMMHDRB HUMMPCA#2 HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLCB HUMMYLCB HUMMPCOT1B HUMPPCOT2 HUMPPP HUMPROPAB#1 HUMPROPAB#1 HUMPROT1 HUMPROT1B HUMPROD4 HUMRSAA HUMSSAA HUMSSBAC HUMRSSA HUMRAAA HUMRSAA HUMRABPAR HUMPROD4 HUMPRO	HUMCAL	HUMCALCBE	HUMCALCR4*	HUMCAM3X1*	HUMCAMA
HUMCSIST HUMCABPL HUMFABPLA HUMFABPLA HUMFABPLA HUMFSHBQ3* HUMG0S2A HUMG0S2PE HUMGBR#1 HUMGF12A9 HUMGROS5PE HUMGROB5 HUMHBB#1 HUMHBB#2 HUMHBB#3 HUMHBB#3 HUMHBB#5 HUMHBBAAZ HUMHBNF1 HUMHISAB HUMHISAC HUMHISH2B HUMHISAB HUMHISAC HUMHISH2B HUMHMG17 HUMHMG17 HUMHMG17 HUMHMG17 HUMHMG18 HUMHMG18 HUMHMG18 HUMHMG19 HUMHMG19 HUMHMG19 HUMHMG19 HUMHMG10 HUMMG10 HUMHMG10 HUMMG10 H	HUMCCK3*	HUMCNP	HUMCNTFR	HUMCOA2IT	HUMCRBP2*
HUMFABPL HUMFABPLA HUMFABPLA HUMGNASG#1* HUMGOS2A HUMGOS2PE HUMGNASG#1* HUMGNASG#1* HUMGNASG#2* HUMGNASG#1* HUMGROPSHUMGROB HUMGROB HUMGROB HUMGROB HUMGROB HUMGROB HUMGROB HUMHBB#1 HUMHBB#2 HUMHBB#3 HUMHBB#5 HUMHBBAAZ HUMHBNF1 HUMHEMOB HUMHISAB HUMHISAB HUMHISAC HUMHISH2B HUMHISH3A HUMHISH3 HUMHISAB HUMHISAC HUMHISH3B HUMHISH3B HUMHISAB HUMHISAC HUMHISHB HUMMILCA HUMMILCA HUMMET HUMMET HUMMET HUMMET HUMMET HUMMISH HUMMYLCA HUMMYLCB HUMMYLCB HUMMYLCB HUMMYLCB HUMMYLCB HUMMYLCB HUMMPODA HUMMPODA HUMPODA HUMPODA HUMPODT HUMPOD HUMPOD HUMPOD HUMPOD HUMPOD HUMPOD HUMPOD HUMPOD HUMPOD HUM	HUMCRYGBC#2	HUMCRYGQ2*	HUMCRYGQ4*	HUMCRYGX4*	HUMCSF1M3*
HUMFSHBQ3* HUMG0S2A HUMG0S2PE HUMGBR#1 HUMGF12A9 HUMGNAS6#1* HUMGNAS6#2* HUMGROB HUMGROB5 HUMGRPF5#2* HUMGRO HUMGROB HUMGROB5 HUMGSA1R HUMGSA2R HUMGTPBPA HUMHBB#1 HUMHBB#2 HUMHBB#3 HUMHISAB HUMHISAC HUMHISH2B HUMHISH3A HUMHISH4 HUMHMG1 HUMHMGAC HUMHMGP5 HUMHMGYB HUMHMGYB HUMHMGYB HUMHMGYB HUMHMGYB HUMHMGYB HUMHMGYB HUMHGP5AA HUMHMGYB HUMHGP5AA HUMHGP5AA HUMHGP5AA HUMHGP5AA HUMHGF1BP1 HUMIGF1BP5A HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMGF1BP1 HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMGF1BP1 HUMIFI56 HUML2AB HUMINIFI HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMET2PS HUMMET1E HUMMGSA HUMMGSAG HUMMHCACA#1 HUMMLC2 HUMMLC3NM HUMMLNA5* HUMMLC2 HUMMLC3NM HUMMLN HUMMBDQ HUMMIFA HUMMSF1 HUMMSF1 HUMMERB HUMMLNA5* HUMMYLCC HUMMYLC1 HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLC HUMMYLC1 HUMMYCB HUMNOPRR HUMOSF1 HUMOTCB HUMOTNPI HUMPGFBA HUMPPC9* HUMPCPAB#1 HUMPGP1X#1 HUMPROT1 HUMPOGFBA HUMPROD4* HUMPROP2AB#1 HUMPROP3AB#2 HUMPROT1 HUMPROT1B HUMPROT2 HUMRAP2 HUMRASH HUMRSSA HUMRDSAC HUMRPS11 HUMSAACT HUMSCL HUMSSAA HUMSSAC HUMSPARC10* HUMSLIPG HUMSTPA HUMTTYP HUMTTOS HUMTTOC HUMTTYP HUMTTOS HUMTTOC HUMTTYP	HUMCSIST	HUMCTRP	HUMCVIB	HUMCYCPS3	HUMCYS3A3*
HUMGNAS6#1* HUMGNAS6#2* HUMGNAS6#3* HUMGNAS6#4* HUMGRFP5#1* HUMGRFP5#2* HUMGRO HUMGROB HUMGROB5 HUMGROG5 HUMGSA1R HUMGSA2R HUMGTPBPA HUMHBB#1 HUMHBB#2 HUMHBB#3 HUMHBB#5 HUMHBAAZ HUMHBNF1 HUMHEMOB HUMHIS3PRM HUMHISAB HUMHLADZA HUMHISH2B HUMHISH3A HUMHISH4 HUMHLAB HUMHLADZA HUMHIL4G HUMHMG17 HUMHMG17G HUMHMG1 HUMHMG1A HUMHMG1B HUMHMGY HUMHMG4A HUMHMG7B HUMHMG7C HUMHMG7D HUMHMG7Y HUMHMG4A HUMHST HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMIGHBP1 HUMIF156 HUML2AB HUMINIFI HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMET2PS HUMMETIE HUMMGSA HUMMGSAG HUMMHCACA#1 HUMMHDC3B HUMMHDR3* HUMMHDRBC HUMMHDRDQ HUMMIFA HUMMC2 HUMMYLC2#1 HUMMLC2 HUMMYLCB HUMMYLCC HUMMYLC1 HUMMYCA#1 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYOL1 HUMNGFR HUMMYLCB HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPOFBA HUMPPC9* HUMPPCA9* HUMPPCPAB#1 HUMPROP1B HUMPROT1 HUMPROP1B HUMPROT2 HUMPROP2AB#1 HUMPROP1 HUMPROT1 HUMPROT1 HUMPROT1 HUMPROP1 HUMPROP1 HUMPROT1 HUMPROP1 HUMPROP1 HUMPROT1 HUMPROP1 HUMPROP1 HUMPROP1 HUMPROP1 HUMSSACT HUMSSA6* HUMSSACH HUMTHPP HUMTTOP1 HUMTTOP1 HUMTTOP1 HUMTTOP1	HUMFABPL	HUMFABPLA	HUMFKMKA	HUMFSH3*	HUMFSHB2*
HUMGRFP5#2* HUMGRO HUMGROB HUMGROB5 HUMGSA1R HUMGSA2R HUMGTPBPA HUMHBB#1 HUMHBB#2 HUMHBB#3 HUMHBB#5 HUMHBBAAZ HUMHBNF1 HUMHEMOB HUMHISAPRM HUMHISAB HUMHISAC HUMHISH2B HUMHISH3A HUMHISH4 HUMHLAB HUMHLADZA HUMHLL4G HUMHMG17 HUMHMG17G HUMHMGI HUMHMGIA HUMHMGIB HUMHMGY HUMHMGYA HUMHMGYB HUMHMGYC HUMHMGYD HUMHOX14 HUMHP2AA HUMHST HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMISGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMHCACA#1 HUMMHDC3B HUMMHDR3* HUMMHDRDC HUMMIFA HUMMLA3B HUMMLC2 HUMMHDRDC HUMMHDRDQ HUMMIFA HUMMT2A HUMMC3NM HUMMLN HUMMLN5* HUMMLNA5* HUMMYLCA HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYOL1 HUMMYCB HUMNOPMR HUMOSF1 HUMOTCB HUMOTPI HUMPGPBA HUMPPCO9* HUMPEPCA9* HUMPGPIX#1 HUMPNOT1 HUMPARP1 HUMPROD4* HUMPROP1AB#1 HUMPGPIX#1 HUMPROT1 HUMPROT1 HUMPROT2 HUMPRO HUMPROP1 HUMPROP1 HUMPROT2 HUMPROP1 HUMPROP1 HUMPROP1 HUMPROT1 HUMPROP1 HUMPROP1 HUMPROT2 HUMPROP1 HUMPROP1 HUMPROT4 HUMPROP1 HUMPROP1 HUMPROT5 HUMPROP1 HUMPROP1 HUMPROT6 HUMPROP1 HUMPROP1 HUMPROT7 HUMPROP1 HUMPROP1 HUMPROT6 HUMSLIPR HUMSISA6* HUMSISM HUMSISPDG HUMSLIPG HUMSLIPR HUMST19#1 HUMSPARC HUMSPARC10* HUMSLIPG HUMTTOS HUMTROPIA	HUMFSHBQ3*	HUMG0S2A	HUMG0S2PE	HUMGBR#1	HUMGFI2A9
HUMGSA1R HUMGSA2R HUMGTPBPA HUMHBB#1 HUMHBB#2 HUMHBB#3 HUMHBB#5 HUMHBBAAZ HUMHBNF1 HUMHEMOB HUMHIS3PRM HUMHISAB HUMHISAC HUMHISH2B HUMHISH3A HUMHISH4 HUMHLAB HUMHLADZA HUMHLL4G HUMHMG17 HUMHMG17G HUMHMG1 HUMHMG1A HUMHMG1B HUMHMGY HUMHMGYA HUMHMGYB HUMHMGYC HUMHMGYD HUMHOX14 HUMHP2AA HUMHST HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMIGHBP1 HUMIIF156 HUMIL2AB HUMINIF1 HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMET2PS HUMMET1E HUMMGSA HUMMGSAG HUMMHCACA#1 HUMMHDC3B HUMMHDR3* HUMMHDRBC HUMMHDRDQ HUMMIFA HUMMLC2 HUMMLC3NM HUMMLN HUMMLN5* HUMMLNA5* HUMMT2A HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYOL1 HUMNGFR HUMNYLCB HUMNOPMR HUMOSF1 HUMOTCB HUMNIFI HUMPGFBA HUMPEPC9* HUMPEPCA9* HUMPGP1X#1 HUMPHIDYIN HUMPDGFBA HUMPROD4* HUMPPCA9* HUMPGP1X#1 HUMPHIDYIN HUMPARP1 HUMPROD4* HUMPROP2AB#1 HUMPROP2AB#2 HUMPROT1 HUMPROT1B HUMPROT2 HUMPRP HUMPROPA HUMPROP3ARC HUMRSACT HUMSCL HUMSSAC* HUMSSM HUMRSACT HUMSCL HUMSSAC* HUMSSM HUMSSPARC10* HUMSPR2A HUMSYB2A5* HUMST39#1 HUMTC2 HUMTTOPIA	HUMGNAS6#1*	HUMGNAS6#2*	HUMGNAS6#3*	HUMGNAS6#4*	HUMGRFP5#1*
HUMHBB#3 HUMHBB#5 HUMHBBAAZ HUMHBNF1 HUMHEMOB HUMHIS3PRM HUMHISAB HUMHISAC HUMHISH2B HUMHISH3A HUMHISH4 HUMHLAB HUMHLADZA HUMHLLAG HUMHMG17 HUMHMG17G HUMHMG1B HUMHMG1B HUMHMGYB HUMHMGYA HUMHST HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMIGHBP1 HUMIET56 HUMIL2AB HUMINIF1 HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMET2PS HUMMETIE HUMMGSA HUMMGSAG HUMMHCACA#1 HUMMHDC3B HUMMHDR3* HUMMHDRBC HUMMHDRDQ HUMMIFA HUMMLC2 HUMMLC3NM HUMMLN HUMMLN5* HUMMYLCA#1 HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYLCA#1 HUMNGFR HUMNC2B HUMMPPCO9* HUMOSF1 HUMOTCB HUMOTNP1 HUMPDGFBA HUMPPCO9* HUMPPCA9* HUMPGPIX#1 HUMPHIDYIN HUMPARP1 HUMPROD4* HUMPROP2AB#1 HUMPROP2AB#2 HUMPROT1 HUMPROP1BA HUMPROT2 HUMPRP HUMPROPAB HUMPROT1 HUMPROT1B HUMPSG3* HUMRAP2 HUMRASH HUMPROT1 HUMPROT1B HUMPSG3* HUMRAP2 HUMRASH HUMRSY79 HUMRODG HUMRSISPDG HUMSLIPG HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC10* HUMSPR2A HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC10* HUMSPR2A HUMSYB2A5* HUMTYPB4 HUMTTMP2 HUMTIMP2 HUMTNC2* HUMTNCS HUMTROC HUMTTOPIA	HUMGRFP5#2*	HUMGRO	HUMGROB	HUMGROB5	HUMGROG5
HUMHISAPRM HUMHISAB HUMHISAC HUMHISH2B HUMHISH3A HUMHISH4 HUMHLAB HUMHLADZA HUMHLL4G HUMHMG17 HUMHMG17G HUMHMG1 HUMHMG1A HUMHMG1B HUMHMGY HUMHMGYA HUMHMGYB HUMHMGYC HUMHMGYD HUMHOX14 HUMHP2AA HUMHST HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMIGHBP1 HUMIFI56 HUMIL2AB HUMINIFI HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMET2PS HUMMETIE HUMMGSA HUMMGSAG HUMMHCACA#1 HUMMHDC3B HUMMHDR3* HUMMHDRBC HUMMHDRDQ HUMMIFA HUMMLC2 HUMMLC3NM HUMMLN HUMMLN5* HUMMYLCB HUMMYLC1 HUMMYLC4#1 HUMMYLCA#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYOL1 HUMNGFR HUMNK21 HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPDGFBA HUMPEPC9* HUMPEPCA9* HUMPGPIX#1 HUMPHIDYIN HUMPDGFBA HUMPEPC9* HUMPRP HUMPGPIX#1 HUMPROT1 HUMPROT1B HUMPROT2 HUMPRP HUMPROP2AB#2 HUMPROT1 HUMPROT1B HUMPS2G3* HUMRAP2 HUMRASH HUMPROT1 HUMPS2 HUMPS11 HUMSAACT HUMSCL HUMSISPAC HUMSISPDG HUMSLIPG HUMSLIPR HUMSSPARC10* HUMSPR2A HUMSYB2A5* HUMTYPA HUMTTMP2 HUMTIMP2 HUMTNC2* HUMTNCS HUMTROC HUMTTOPIA	HUMGSA1R	HUMGSA2R	HUMGTPBPA	HUMHBB#1	HUMHBB#2
HUMHISH4 HUMHLAB HUMHLADZA HUMHLLAG HUMHMG17 HUMHMG17G HUMHMGI HUMHMGIA HUMHMGIB HUMHMGY HUMHMGYA HUMHMGYB HUMHMGYC HUMHMGYD HUMHOX14 HUMHP2AA HUMHST HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMIGHBP1 HUMIFI56 HUMIL2AB HUMINIFI HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMET2PS HUMMETIE HUMMGSA HUMMGSAG HUMMHCACA#1 HUMMHDC3B HUMMHDR3* HUMMHDRBC HUMMHDRDQ HUMMIFA HUMMLC2 HUMMC23M HUMMLN HUMMLN5* HUMMLNA5* HUMMT2A HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYOL1 HUMNGFR HUMNK21 HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPDGFBA HUMPEPC9* HUMPEPCA9* HUMPGPIX#1 HUMPHIDYIN HUMPPARP1 HUMPROD4* HUMPROP2AB#1 HUMPROP2AB#2 HUMPROT1 HUMPROT1B HUMPSC3* HUMRAP2 HUMRASH HUMPROT1 HUMPSC HUMRPS11 HUMSAACT HUMSCL HUMSISA6* HUMSISM HUMSISPDG HUMSLIPG HUMSLIPR HUMSNRAA HUMSSM HUMSSPARC10* HUMSPARA HUMTHBP HUMTNCS HUMTROC HUMTTOPIA	HUMHBB#3	HUMHBB#5	HUMHBBAAZ	HUMHBNF1	HUMHEMOB
HUMHMG17G HUMHMGI HUMHMGIA HUMHMGIB HUMHMGY HUMHMGYA HUMHMGYB HUMHMGYC HUMHMGYD HUMHOX14 HUMHP2AA HUMHST HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMIGHBP1 HUMIFI56 HUML2AB HUMINIFI HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMHDC3B HUMMHDR3* HUMMHDRBC HUMMHDRDQ HUMMIFA HUMMLC2 HUMMLC3NM HUMMLN HUMMLN5* HUMMLNA5* HUMMT2A HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPDGFBA HUMPEPC9* HUMPEPCA9* HUMPGPIX#1 HUMPHIDYIN HUMPDGFBA HUMPROD4* HUMPROP2AB#1 HUMPGPIX#1 HUMPROT1 HUMPROT1B HUMPROT2 HUMPRP HUMPROPA HUMPROT1 HUMPS2 HUMPS2G3* HUMRAP2 HUMRASH HUMRSY79 HUMRODG HUMRSSI1 HUMSAACT HUMSLIPR HUMSISA6* HUMSISM HUMSISPDG HUMSLIPG HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC10* HUMTHBP HUMTNCS HUMTROCL HUMTHYP HUMTIMP2 HUMTROPIA	HUMHIS3PRM	HUMHISAB	HUMHISAC	HUMHISH2B	HUMHISH3A
HUMHMGYA HUMHMGYB HUMHMGYC HUMHMGYD HUMHOX14 HUMHP2AA HUMHST HUMIGF27 HUMIGFBP5A HUMIGFBP6 HUMIGH3A HUMIGHBP1 HUMIIFI56 HUMIL2AB HUMINIFI HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMET2PS HUMMETIE HUMMGSA HUMMGSAG HUMMHCACA#1 HUMMHDC3B HUMMLC3NM HUMMLNA5* HUMMT2A HUMMLC3NM HUMMLNA5* HUMMYLC4#1 HUMMYLC4#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYC11 HUMNGFR HUMNK21 HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPDGFBA HUMPEPC9* HUMPEPCA9* HUMPGPIX#1 HUMPHIDYIN HUMPDGFBA HUMPROD4* HUMPROP2AB#1 HUMPGPIX#1 HUMPROT1 HUMPARP1 HUMPROT2 HUMPRP HUMPROP2AB#2 HUMPROT1 HUMPS2 HUMPS2G3* HUMRAP2 HUMRASH HUMPROT1 HUMPS2 HUMRPS11 HUMSAACT HUMSAACT HUMSISA6* HUMSISM HUMSISPDG HUMSLIPG HUMSLIPR HUMSISAACC HUMSPARC10* HUMSPACA HUMTYB4 HUMTYPP HUMTCSM HUMTNC2* HUMTNCS HUMTROCL	HUMHISH4	HUMHLAB	HUMHLADZA	HUMHLL4G	HUMHMG17
HUMHP2AAHUMHSTHUMIGF27HUMIGFBP5AHUMIGFBP6HUMIGH3AHUMIGHBP1HUMIFI56HUMIL2ABHUMINIFIHUMISKHUMKSGFAHUMLECHUMLEC14KHUMMETHUMMET2HUMMET2PSHUMMETIEHUMMGSAHUMMGSAGHUMMHCACA#1HUMMHDC3BHUMMHDR3*HUMMHDRBCHUMMHDRDQHUMMIFAHUMMLC2HUMMLC3NMHUMMLNHUMMLN5*HUMMLNA5*HUMMYLCAHUMMYLCA#1HUMNGFRHUMNK21HUMNOPMRHUMOSF1HUMOTCBHUMOTNPIHUMPDGFBAHUMPEPC9*HUMPEPCA9*HUMPGPIX#1HUMPHIDYINHUMPPARP1HUMPROD4*HUMPROP2AB#1HUMPROP2AB#2HUMPROT1HUMPROT1BHUMPROT2HUMPRPHUMPRP0AHUMPRT1AHUMPS2HUMPS2G3*HUMRAP2HUMRASHHUMRBSY79HUMRODGHUMRPS11HUMSAACTHUMSCLHUMSISA6*HUMSISMHUMSISPDGHUMSLIPGHUMSCLHUMSISA6*HUMSISMHUMSPARC10*HUMSPR2AHUMSYB2A5*HUMTS19#1HUMTC2HUMTCSMHUMTHBPHUMTHYB10HUMTHYB4HUMTHYPHUMTIMP2HUMTNC2*HUMTNCSHUMTROCHUMTROPIA	HUMHMG17G	HUMHMGI	HUMHMGIA	HUMHMGIB	HUMHMGY
HUMIGH3A HUMIGHBP1 HUMIFI56 HUMIL2AB HUMINIFI HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMET2PS HUMMETIE HUMMGSA HUMMGSAG HUMMHCACA#1 HUMMHDC3B HUMMHDR3* HUMMHDRBC HUMMHDRDQ HUMMIFA HUMMLC2 HUMMLC3NM HUMMLN HUMMLN5* HUMMLNA5* HUMMT2A HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYOL1 HUMNGFR HUMNK21 HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPDGFBA HUMPEPC9* HUMPEPCA9* HUMPGPIX#1 HUMPHIDYIN HUMPPARP1 HUMPROD4* HUMPROP2AB#1 HUMPROP2AB#2 HUMPROT1 HUMPROT1B HUMPROT2 HUMPRP HUMPROPA HUMPRT1A HUMPS2G3* HUMRAP2 HUMRASH HUMPST1B HUMRPS11 HUMSAACT HUMSCL HUMSISA6* HUMSISM HUMSISPDG HUMSLIPG HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC10* HUMSPR2A HUMSYB2A5* HUMTS19#1 HUMTC2 HUMTCSM HUMTHBP HUMTNCS HUMTROCL HUMTROPIA	HUMHMGYA	HUMHMGYB	HUMHMGYC	HUMHMGYD	HUMHOX14
HUMISK HUMKSGFA HUMLEC HUMLEC14K HUMMET HUMMET2 HUMMET2PS HUMMETIE HUMMGSA HUMMGSAG HUMMHCACA#1 HUMMHDC3B HUMMHDR3* HUMMHDRBC HUMMHDRDQ HUMMIFA HUMMLC2 HUMMLC3NM HUMMLN HUMMLN5* HUMMLNA5* HUMMT2A HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYOL1 HUMNGFR HUMNK21 HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPDGFBA HUMPEPC9* HUMPEPCA9* HUMPGPIX#1 HUMPHIDYIN HUMPPARP1 HUMPROD4* HUMPROP2AB#1 HUMPROP2AB#2 HUMPROT1 HUMPROT1B HUMPROT2 HUMPRP HUMPROPA HUMPRT1A HUMPS2 HUMPS2G3* HUMRAP2 HUMRASH HUMRSY79 HUMRODG HUMRPS11 HUMSAACT HUMSCL HUMSISA6* HUMSISM HUMSISPDG HUMSLIPG HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC10* HUMSPR2A HUMSYB2A5* HUMT519#1 HUMTC2 HUMTCSM HUMTHBP HUMTNCS HUMTROC HUMTROPIA	HUMHP2AA	HUMHST	HUMIGF27	HUMIGFBP5A	HUMIGFBP6
HUMMET2HUMMET2PSHUMMETIEHUMMGSAHUMMGSAGHUMMHCACA#1HUMMHDC3BHUMMHDR3*HUMMHDRBCHUMMHDRDQHUMMIFAHUMMLC2HUMMLC3NMHUMMLNHUMMLN5*HUMMLNA5*HUMMT2AHUMMYLCA#1HUMMYLCA#2HUMMYLCBHUMMYLCCHUMMYLV1HUMMYOL1HUMNGFRHUMNK21HUMNOPMRHUMOSF1HUMOTCBHUMOTNPIHUMPDGFBAHUMPEPC9*HUMPEPCA9*HUMPGPIX#1HUMPRIDYINHUMPPARP1HUMPROD4*HUMPROP2AB#1HUMPROP2AB#2HUMPROT1HUMPROT1BHUMPROT2HUMPRPHUMPRP0AHUMPRT1AHUMPS2HUMPS2G3*HUMRASHHUMRSY99HUMRODGHUMRPS11HUMSAACTHUMSCLHUMSISA6*HUMSISMHUMSISPDGHUMSLIPGHUMSLIPRHUMSNRAAHUMSPARCHUMSPARC10*HUMSPR2AHUMSYB2A5*HUMT519#1HUMTC2HUMTCSMHUMTHBPHUMTHYB10HUMTHYB4HUMTHYPHUMTIMP2HUMTNC2*HUMTNCSHUMTROCHUMTROPIA	HUMIGH3A	HUMIGHBP1	HUMIIF I 56	HUMIL2AB	HUMINIFI
HUMMHCACA#1HUMMHDC3BHUMMHDR3*HUMMHDRBCHUMMHDRDQHUMMIFAHUMMLC2HUMMLC3NMHUMMLNHUMMLN5*HUMMLNA5*HUMMT2AHUMMYLCA#1HUMMYLCA#2HUMMYLCBHUMMYLCCHUMMYLV1HUMMYOL1HUMNGFRHUMNK21HUMNOPMRHUMOSF1HUMOTCBHUMOTNPIHUMPDGFBAHUMPEPC9*HUMPEPCA9*HUMPGPIX#1HUMPHIDYINHUMPPARP1HUMPROD4*HUMPROP2AB#1HUMPROP2AB#2HUMPROT1HUMPROT1BHUMPROT2HUMPRPHUMPR0AHUMPRT1AHUMPS2HUMPS2G3*HUMRAP2HUMRASHHUMRSY79HUMRODGHUMRPS11HUMSAACTHUMSCLHUMSISA6*HUMSISMHUMSISPDGHUMSLIPGHUMSLIPRHUMSNRAAHUMSPARCHUMSPARC10*HUMSPR2AHUMSYB2A5*HUMTS19#1HUMTC2HUMTCSMHUMTHBPHUMTHYB10HUMTHYB4HUMTHYPHUMTIMP2HUMTNC2*HUMTNCSHUMTROCHUMTROPIA	HUMISK	HUMKSGFA	HUMLEC	HUMLEC14K	HUMMET
HUMMIFA HUMMLO2 HUMMLO3NM HUMMLN HUMMLN5* HUMMYLCA#1 HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYOL1 HUMNGFR HUMNK21 HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPDGFBA HUMPEPC9* HUMPEPCA9* HUMPGPIX#1 HUMPROD4* HUMPROP2AB#1 HUMPROP2AB#2 HUMPROT1 HUMPROT1 HUMPROT2 HUMPRP HUMPROA HUMPRT1A HUMPS2 HUMPS2G3* HUMRAP2 HUMRASH HUMRBSY79 HUMRODG HUMRPS11 HUMSISPDG HUMSLIPG HUMSLIPG HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC HUMSPARC HUMSPARC HUMSPARC HUMTCSM HUMTHYB10 HUMTROPIA	HUMMET2	HUMMET2PS	HUMMETIE	HUMMGSA	HUMMGSAG
HUMMLNA5* HUMMT2A HUMMYLCA#1 HUMMYLCA#2 HUMMYLCB HUMMYLCC HUMMYLV1 HUMMYOL1 HUMNGFR HUMNK21 HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPDGFBA HUMPEPC9* HUMPEPCA9* HUMPGPIX#1 HUMPHIDYIN HUMPPARP1 HUMPROD4* HUMPROP2AB#1 HUMPROP2AB#2 HUMPROT1 HUMPROT1B HUMPROT2 HUMPRP HUMPROA HUMPRT1A HUMPS2 HUMPS2G3* HUMRAP2 HUMRASH HUMRBSY79 HUMRODG HUMRPS11 HUMSAACT HUMSCL HUMSISA6* HUMSISM HUMSISPDG HUMSLIPG HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC10* HUMSPR2A HUMSYB2A5* HUMT519#1 HUMTC2 HUMTCSM HUMTHPP HUMTNCS HUMTROPIA	HUMMHCACA#1	HUMMHDC3B	HUMMHDR3*	HUMMHDRBC	
HUMMYLCCHUMMYLV1HUMMYOL1HUMNGFRHUMNK21HUMNOPMRHUMOSF1HUMOTCBHUMOTNPIHUMPDGFBAHUMPEPC9*HUMPEPCA9*HUMPGPIX#1HUMPHIDYINHUMPPARP1HUMPROD4*HUMPROP2AB#1HUMPROP2AB#2HUMPROT1HUMPROT1BHUMPROT2HUMPRPHUMPRP0AHUMPRT1AHUMPS2HUMPS2G3*HUMRAP2HUMRASHHUMRBSY79HUMRODGHUMRPS11HUMSAACTHUMSLIPRHUMSISA6*HUMSISMHUMSISPDGHUMSLIPGHUMSLIPRHUMSNRAAHUMSPARCHUMSPARC10*HUMSPR2AHUMSYB2A5*HUMT519#1HUMTC2HUMTCSMHUMTHBPHUMTHYB10HUMTHYB4HUMTHYPHUMTIMP2HUMTNC2*HUMTNCSHUMTROCHUMTROPIA	HUMMIFA	HUMMLC2	HUMMLC3NM	HUMMLN	HUMMLN5*
HUMNOPMR HUMOSF1 HUMOTCB HUMOTNPI HUMPDGFBA HUMPEPC9* HUMPEPCA9* HUMPGPIX#1 HUMPHIDYIN HUMPPARP1 HUMPROD4* HUMPROP2AB#1 HUMPROP2AB#2 HUMPROT1 HUMPROT1B HUMPROT2 HUMPRP HUMPRP0A HUMPRT1A HUMPS2 HUMPS2G3* HUMRAP2 HUMRASH HUMRBSY79 HUMRODG HUMRPS11 HUMSAACT HUMSCL HUMSISA6* HUMSISM HUMSISPDG HUMSLIPG HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC10* HUMSPACA HUMSYB2A5* HUMT519#1 HUMTC2 HUMTCSM HUMTHBP HUMTHYB10 HUMTHYB4 HUMTHYP HUMTIMP2 HUMTNC2* HUMTNCS HUMTROCI HUMTROPIA	HUMMLNA5*	HUMMT2A	HUMMYLCA#1	HUMMYLCA#2	
HUMPEPC9*HUMPEPCA9*HUMPGPIX#1HUMPHIDYINHUMPPARP1HUMPROD4*HUMPROP2AB#1HUMPROP2AB#2HUMPROT1HUMPROT1BHUMPROT2HUMPRPHUMPRP0AHUMPRT1AHUMPS2HUMPS2G3*HUMRAP2HUMRASHHUMRBSY79HUMRODGHUMRPS11HUMSAACTHUMSCLHUMSISA6*HUMSISMHUMSISPDGHUMSLIPGHUMSLIPRHUMSNRAAHUMSPARCHUMSPARC10*HUMSPR2AHUMSYB2A5*HUMT519#1HUMTC2HUMTCSMHUMTHBPHUMTHYB10HUMTHYB4HUMTHYPHUMTIMP2HUMTNC2*HUMTNCSHUMTROCHUMTROPIA	HUMMYLCC	HUMMYLV1	HUMMYOL1	HUMNGFR	
HUMPROD4*HUMPROP2AB#1HUMPROP2AB#2HUMPROT1HUMPROT1BHUMPROT2HUMPRPHUMPRP0AHUMPRT1AHUMPS2HUMPS2G3*HUMRAP2HUMRASHHUMRBSY79HUMRODGHUMRPS11HUMSAACTHUMSCLHUMSISA6*HUMSISMHUMSISPDGHUMSLIPGHUMSLIPRHUMSNRAAHUMSPARCHUMSPARC10*HUMSPR2AHUMSYB2A5*HUMT519#1HUMTC2HUMTCSMHUMTHPPHUMTHYB10HUMTHYB4HUMTHYPHUMTIMP2HUMTNC2*HUMTNCSHUMTROCHUMTROPIA	HUMNOPMR	HUMOSF1	HUMOTCB	HUMOTNPI	
HUMPROT2HUMPRPHUMPRP0AHUMPRT1AHUMPS2HUMPS2G3*HUMRAP2HUMRASHHUMRBSY79HUMRODGHUMRPS11HUMSAACTHUMSCLHUMSISA6*HUMSISMHUMSISPDGHUMSLIPGHUMSLIPRHUMSNRAAHUMSPARCHUMSPARC10*HUMSPR2AHUMSYB2A5*HUMT519#1HUMTC2HUMTCSMHUMTHPPHUMTHYB10HUMTHYB4HUMTHYPHUMTIMP2HUMTNC2*HUMTNCSHUMTROCHUMTROPIA	HUMPEPC9*	HUMPEPCA9*	HUMPGPIX#1		
HUMPS2G3* HUMRAP2 HUMRASH HUMRBSY79 HUMRODG HUMRPS11 HUMSAACT HUMSCL HUMSISA6* HUMSISM HUMSISPDG HUMSLIPG HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC10* HUMSPR2A HUMSYB2A5* HUMT519#1 HUMTC2 HUMTCSM HUMTHBP HUMTHYB10 HUMTHYB4 HUMTHYP HUMTIMP2 HUMTNC2* HUMTNCS HUMTROC HUMTROPIA	HUMPROD4*	HUMPROP2AB#1	HUMPROP2AB#2	HUMPROT1	
HUMRPS11HUMSAACTHUMSCLHUMSISA6*HUMSISMHUMSISPDGHUMSLIPGHUMSLIPRHUMSNRAAHUMSPARCHUMSPARC10*HUMSPR2AHUMSYB2A5*HUMT519#1HUMTC2HUMTCSMHUMTHBPHUMTHYB10HUMTHYB4HUMTHYPHUMTIMP2HUMTNC2*HUMTNCSHUMTROCHUMTROPIA	HUMPROT2	HUMPRP			
HUMSISPDG HUMSLIPG HUMSLIPR HUMSNRAA HUMSPARC HUMSPARC10* HUMSPR2A HUMSYB2A5* HUMT519#1 HUMTC2 HUMTCSM HUMTHBP HUMTHYB10 HUMTHYB4 HUMTHYP HUMTIMP2 HUMTNC2* HUMTNCS HUMTROC HUMTROPIA	HUMPS2G3*	HUMRAP2			
HUMSPARC10* HUMSPR2A HUMSYB2A5* HUMTS19#1 HUMTC2 HUMTCSM HUMTHBP HUMTHYB10 HUMTHYB4 HUMTHYP HUMTIMP2 HUMTNC2* HUMTNCS HUMTROC HUMTROPIA	HUMRPS11				
HUMTCSM HUMTHBP HUMTHYB10 HUMTHYB4 HUMTHYP HUMTIMP2 HUMTNC2* HUMTNCS HUMTROPIA	HUMSISPDG				
HUMTIMP2 HUMTNC2* HUMTNCS HUMTROC HUMTROPIA	HUMSPARC10*				
TIONTINE TIONTINE					
HUMU1C HUMUBI13 HUMYB1A HUMYUBG1 HUMZNFBPAA					
	HUMU1C	HUMUBI13	HUMYB1A	HUMYUBG1	HUMZNFBPAA

^a Names used in the GenBank (Wada et al., 1991), where the definitions for the symbols # and * (Aota et al., 1988; Maruyama et al., 1986) are explicitly given.

Method

Let the occurrence frequencies of bases A, C, G and T at the *i*th (i = 1, 2, 3) codon position in a coding sequence be denoted by a_i , c_i , g_i and t_i , respectively. Obviously, we have

$$a_i + c_i + g_i + t_i = 1, \quad 0 < a_i, c_i, g_i, t_i < 1.$$
 (1)

Because the following formulation is generally valid regardless of which of the three codon positions is referred to, for brevity the subscript i will be omitted below except for those cases where a special reference mark is needed for distinction. It is implied from eq.1 that, of the four real numbers a, c, g, and t, only three are independent. Therefore, it is instructive to make the following transformation:

$$\begin{cases} x = 2(a+g) - 1 \\ y = 2(a+c) - 1 \\ z = 2(a+t) - 1 \end{cases}$$
 (2)

In the anti-sense sequences, we should carry out a complementary transform; i.e., $A \Rightarrow T$, $C \Rightarrow G$, $G \Rightarrow C$, and $T \Rightarrow A$. It follows after such a transformation that

$$\begin{cases} \overline{x} = -x \\ \overline{y} = -y \\ \overline{z} = z \end{cases}$$
 (3)

where \overline{x} , \overline{y} , and \overline{z} represent the corresponding coordinates of an anti-sense chain. Considering the fact that the DNA double helix is an anti-parallel helix structure, we should also perform the following transformation: codon position $1 \Rightarrow$ codon position 3 in the antisense chain; codon position $2 \Rightarrow$ codon position 2 in the anti-sense chain; and codon position $3 \Rightarrow$ codon position 1 in the anti-sense chain. Combining these two transformations, we have

$$\begin{cases} \overline{x_1} = -x_3 \\ \overline{y_1} = -y_3 \\ \overline{z_1} = z_3 \end{cases}$$
 (4)

etc.

In summary, our method includes the following procedures:

- 1. For each of the 180 human protein coding sequences, the frequencies of bases in each codon position can be calculated from the codon usage table (Wata, 1991).
- 2. For the *i*th codon position, calculate the sense point coordinates x_i , y_i , and z_i (i = 1, 2, 3) by using eq. 2.
- 3. Display each of the 180 sense points thus obtained in the *i*th 3-dimensional (3-D) space for i = 1, 2, and 3.
- 4. Calculate the corresponding anti-sense point coordinates $\overline{x_i}$, $\overline{y_i}$, and $\overline{z_i}$ by using eq. 4 for i = 1, 2, and 3.
- 5. Display each of the 180 anti-sense points in the *i*th 3-D space for i = 1, 2, and 3.
- 6. Comparing the distribution of these two kinds of points in each of the three 3-D spaces.

The results are shown in Figs. 1-3, in which the point corresponding to a coding sequence is denoted by an open circle \circ , while the point corresponding to an anti-sense sequence denoted by a filled circle \bullet .

Analysis and discussion

Let us pay attention to Fig. 1 first, which represents the distribution of the occurrence frequencies of bases at the first codon position. The distribution of the points for the coding sequences in Fig. 1 shows that most of the bases at the first codon position are G and A, and G is the most dominant base, a result in a good agreement with that described previously (Chou and Zhang, 1992). However, the distribution of the points for the anti-sense sequences shows that most of the bases in the first codon position in this case are G and C. However, G is no longer the most dominant base. The number of G occurring at the first position is roughly equal to that of C. Note that the two distributions are much different, and yet they have a little overlap region.

Consider Fig. 2, which represents the distribution of the base frequencies at the second codon position. For the sense chain sequences, as can be seen from the graph that A is the dominant base, while G is the least dominant base. However, for the anti-sense sequences, according to the graph method we find that T is the dominant base and C is the least dominant base. This is a natural result due to the complementary principle.

Finally, let us consider Fig. 3, which shows the distribution of base frequencies at the third codon position. For the sense chain coding sequences, it can be seen from the graphic expression that G and C are the most dominant bases with $g_3 \approx c_3$. A and T are the least occurring bases. However, for the antisense sequences, it is observed from Fig. 3 that the pyrimidine bases C and T are the most dominant bases at the third position.

The average frequencies for each of the four bases at each of the three codon positions for both the sense chain and the anti-sense chain sequences are listed in Table 2.

A recent study (Zhang and Chou, 1994) shows that the distribution patterns of base frequencies for the first two codon positions reflect the origin for producing native folding structures of proteins. These patterns should be basically species-independent. While the distribution patterns of base frequencies at the third codon position are generally species-dependent. On the other hand, summarizing the results discussed above, we find that the codon usage and hence the distribution patterns of base frequencies at three codon positions are quite different between the coding sequences and the anticoding sequences. For the 180 anti-coding sequences complementary to 180 human protein coding sequences, although they contain no stop codons and hence are not compartmentalized by terminators, it seems unlikely for them to be able to encode for native human proteins. As pointed out previously (Zhang and Chou, 1994) that the requirement of forming stable structures of proteins exerts considerable severe constraints on the frequencies of bases at the first two codon positions. However, the distribution patterns of base frequencies at the first two codon positions for the 180 anti-coding sequences do not generally satisfy the constraint requirements. Merely from this point of view we come to the conclusion that, generally speaking, these 180 anti-coding sequences could not be used to encode for native human protein. Nevertheless, especially note that the point distributions in Figs. 1 through 3 have

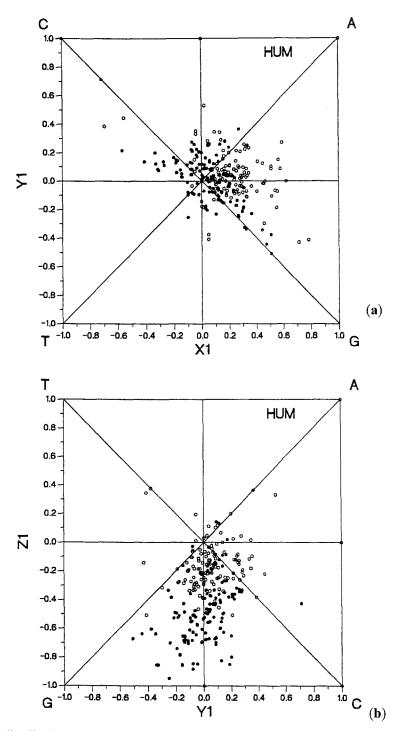
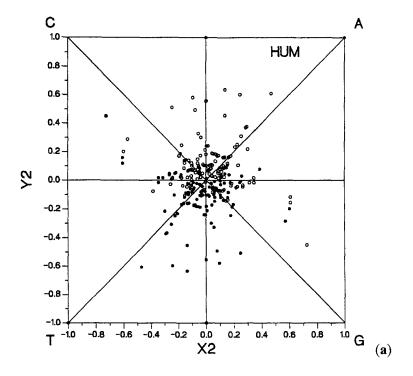


Fig. 1. The distribution of the base frequencies at the first codon positon for both the sense chain and the anti-sense chain sequences. The point corresponding to a sense chain is called a sense point, denoted by an open circle ○, and that to an anti-sense chain called an anti-sense point, denoted by a filled circle ●. The distribution graphs obtained by projecting all the points to (a) the X-Y plane, and (b) the Y-Z plane, respectively. For more detail about the graph, see, e.g. (Zhang and Zhang, 1991; Chou and Zhang, 1992; Zhang and Chou, 1994)



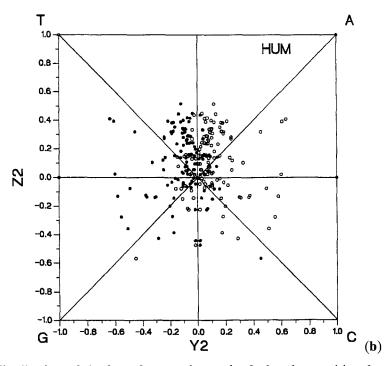
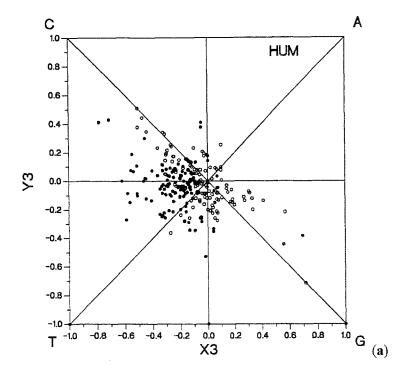


Fig. 2. The distribution of the base frequencies at the 2nd codon position for both the sense chain and the anti-sense chain sequences. The point corresponding to a sense chain is called a sense point, denoted by an open circle ○, and that to an anti-sense chain called an anti-sense point, denoted by a filled circle ●. The distribution graphs obtained by projecting all the points to (a) the X-Y plane, and (b) the Y-Z plane, respectively



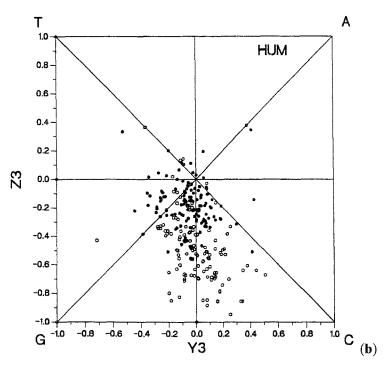


Fig. 3. The distribution of the base frequencies at the 3rd codon position for both the sense chain and the anti-sense chain sequences. The point corresponding to a sense chain is called a sense point, denoted by an open circle \odot , and that to an anti-sense chain called an anti-sense point, denoted by a filled circle \bullet . The distribution graphs obtained by projecting all the points to (a) the X-Y plane, (b) the Y-Z plane, respectively

Codon position	Sense chains			Anti-sense chains				
	$\langle A \rangle$	(C)	$\langle G \rangle$	$\langle T \rangle$	$\langle A \rangle$	(C)	$\langle G \rangle$	$\langle T \rangle$
1	0.288	0.244	0.326	0.142	0.141	0.367	0.376	0.116
2	0.308	0.231	0.217	0.244	0.244	0.217	0.231	0.308
3	0.116	0.376	0.367	0.141	0.142	0.326	0.244	0.288

Table 2. The average frequencies for each of the four bases at each of the three codon positions for both the 180 sense chains and the 180 anti-sense chains

remarkable overlap regions. This implies that at least for few of the 180 anticoding sequences, it is still possible to be used to encode for stable human proteins. If adding the codon ATG in front and adding stop codon later, as well as providing some necessary fragments such as promoter to each of these few anti-coding sequences, we could make these sequences be expressible, and their products might be stable proteins.

The purposes of this study are two-fold. First, we hope to explore the biological function of the anti-sense sequences in contrast to the coding sequences. Second, we hope to discuss the question as what kind of DNA sequences can encode for stable proteins, and what kind can not. Why? The tool used is to analyze the distribution of base frequencies for both strands. It was found that most of the anti-sense sequences corresponding to the 2681 human coding sequences can not function to encode for protein sequences owing to many stop codons dwelling within them. It seems that there are no direct biological functions for these sequences, except for forming a double helix structure. However, of the 2681 anti-sense sequences, we did find 180 sequences in which there are no stop codons. For such 180 anti-sense sequences, if expressed by means of some genetic engineering technique, they should be able to lead to the formation of amino acid sequences, at least from a theoretical point of view. Thus, another question is that if these amino acid sequences can be folded into stable proteins? It is well-known that a random DNA sequence can not encode for a stable protein. Actually, Trifonov (1987) has found that irrespective of the choice of the genes and species, the base G is always found to prefer to occur at the first positions of the triplets, while avoiding the second ones in any coding sequences. He described the pattern of these coding sequences as a universal three-base periodical pattern (G-non-G-N), where N means any bases. Such a pattern could maintain the correct reading frame during the translation (Trifonov, 1987). We have observed the same pattern for the coding sequences by using the graphic method (Zhang and Zhang, 1991; Chou and Zhang, 1992; Zhang and Chou, 1994). This pattern is considered as the necessary condition for any protein-coding sequences. The following question comes up immediately: Do the 180 anti-sense sequences possess such a pattern? A simple method to answer this question is to observe the point distribution in Figs. 1-3, and see if some overlap regions exist between the two counterparts. Interestingly, some little overlap regions do exist. Furthermore, it is seen from Table 1 that G is the most dominant base at the first position, and meanwhile G is the less dominant base at the second position for the anti-sense sequences. Therefore, it might be possible for a few of the anti-sense sequences to encode for stable human proteins based on the consideration of the Trifonov's pattern G-non G-N. The above discussion is only a theoretical exploration. Actually, these sequences have no promoters, no stop codons or no termination codons, it is difficult to express them and to produce the corresponding proteins. Nevertheless, the modern technique of genetic engineering provides the possibility to cut the sequences down and then insert the sequences reversely into the DNA sequence. Here "reversely" means the exchange of the sense strand and the anti-sense strand. Note that the mRNA-identical strand is referred to as the sense chain in this paper. If such an "operation" is successful and the anti-sense sequences are expressed, it would be exciting to see what kind of proteins will be produced. These proteins, if exist, may called the "anti-proteins". Thus, a series of questions will emerge. Do these anti-proteins have some "complementary" structural relationship with their counterparts? What kind of biological functions do they bear? And so forth.

What we offered in this short article is merely an introductory remark so that others may come up with valuable in-depth discussions, just like what is implied in an ancient Chinese proverb as saying "cast a brick to attract jade". We hope to see more studies, both theoretical and experimental, can be stimulated around this interesting but not yet sufficiently investigated topic.

Acknowledgements

We would like to thank Professor T. Ikemura for supplying the relevant data for codon usages. We are indebted to Dr. Wei-Zhu Zhong for help in drawing all the Figures in this report. Valuable discussions with Drs. Ming Xu and Roger L. Yu are also gratefully acknowledged. The present study was supported in part by the grant no. 19577104 and no. 39570187 from China Natural Science Foundation.

References

- Aota S, Gojobori T, Ishibashi F, Maruyama T, Ikemura T (1988) Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res 16 [Suppl]: r315–r402
- Chou KC, Zhang CT (1992) Diagrammatization of codon usage in 339 human immunodeficiency virus proteins and its biological implication. AIDS Res Hum Retroviruses 12: 1967–1976
- Maruyama T, Gojobori T, Aota S, Ikemura T (1986) Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res 14 [Suppl]: r151-r197
- Trifonov EN (1987) Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16 S rRNA nucleotide sequences. J Mol Biol 194: 643–652
- Wata K, Wata Y, Doi H, Ishibashi F, Gojobori T, Ikemura T (1991) Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res 19 [Suppl]: r1981-r1986

Zhang CT, Chou KC (1994) A graphic approach to analyzing codon usage in 1562 E. Coli protein coding sequences. J Mol Biol 238: 1–8

Zhang CT, Zhang R (1991) Analysis of distribution of bases in the coding sequences by a diagrammatic technique. Nucleic Acids Res 19: 6313–6317

Authors' address: Dr. C.-T. Zhang, Department of Physics, Tianjin University, Tianjin 300072, China, Dr. K.-C. Chou, Computer-Aided Drug Discovery, Upjohn Laboratories, Kalamazoo, MI 49001-4940, U.S.A.

Received July 1, 1995