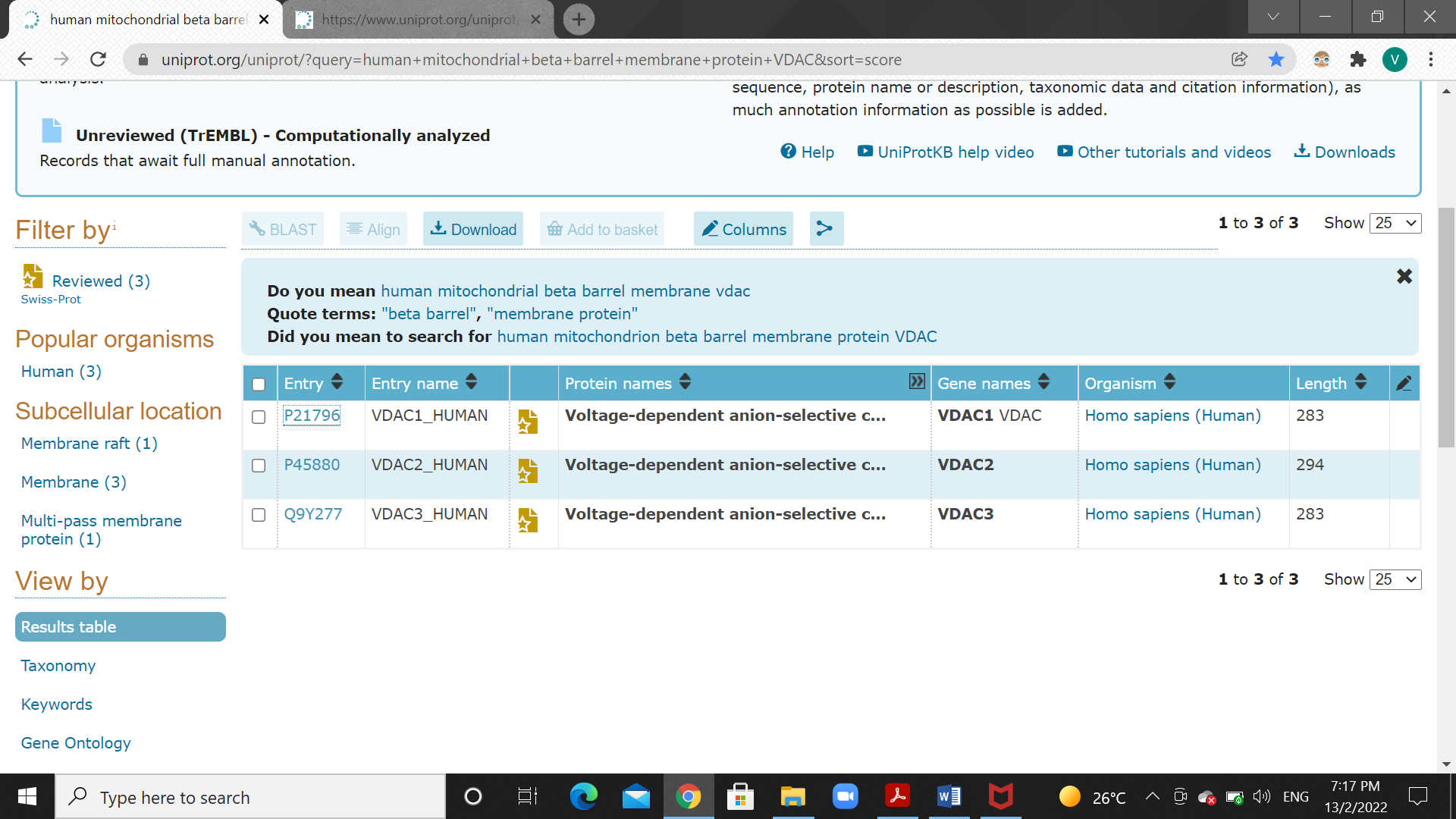
***Practicals 3***

*-BS19B032*

*-R. Vasantha Kumar*

*1) There are three human mitochondrial β barrel membrane protein VDAC: VDAC 1, VDAC 2, VDAC 3.*

**

*VDAC 1:* MAVPPTYADLGKSARDVFTKGYGFGLIKLDLKTKSENGLEFTSSGSANTETTKVTGSLET

KYRWTEYGLTFTEKWNTDNTLGTEITVEDQLARGLKLTFDSSFSPNTGKKNAKIKTGYKR

EHINLGCDMDFDIAGPSIRGALVLGYEGWLAGYQMNFETAKSRVTQSNFAVGYKTDEFQL

HTNVNDGTEFGGSIYQKVNKKLETAVNLAWTAGNSNTRFGIAAKYQIDPDACFSAKVNNS

SLIGLGYTQTLKPGIKLTLSALLDGKNVNAGGHKLGLGLEFQA

*Its function is to form channels in mitochondrial outer membrane and plasma membrane for diffusion of hydrophilic molecules.*

*VDAC 2:* MATHGQTCARPMCIPPSYADLGKAARDIFNKGFGFGLVKLDVKTKSCSGVEFSTSGSSNT

DTGKVTGTLETKYKWCEYGLTFTEKWNTDNTLGTEIAIEDQICQGLKLTFDTTFSPNTGK

KSGKIKSSYKRECINLGCDVDFDFAGPAIHGSAVFGYEGWLAGYQMTFDSAKSKLTRNNF

AVGYRTGDFQLHTNVNDGTEFGGSIYQKVCEDLDTSVNLAWTSGTNCTRFGIAAKYQLDP

TASISAKVNNSSLIGVGYTQTLRPGVKLTLSALVDGKSINAGGHKVGLALELEA

*Its function is similar to VDAC 1 to form channels in plasma membrane.*

*VDAC 3:* MCNTPTYCDLGKAAKDVFNKGYGFGMVKIDLKTKSCSGVEFSTSGHAYTDTGKASGNLET

KYKVCNYGLTFTQKWNTDNTLGTEISWENKLAEGLKLTLDTIFVPNTGKKSGKLKASYKR

DCFSVGSNVDIDFSGPTIYGWAVLAFEGWLAGYQMSFDTAKSKLSQNNFALGYKAADFQL

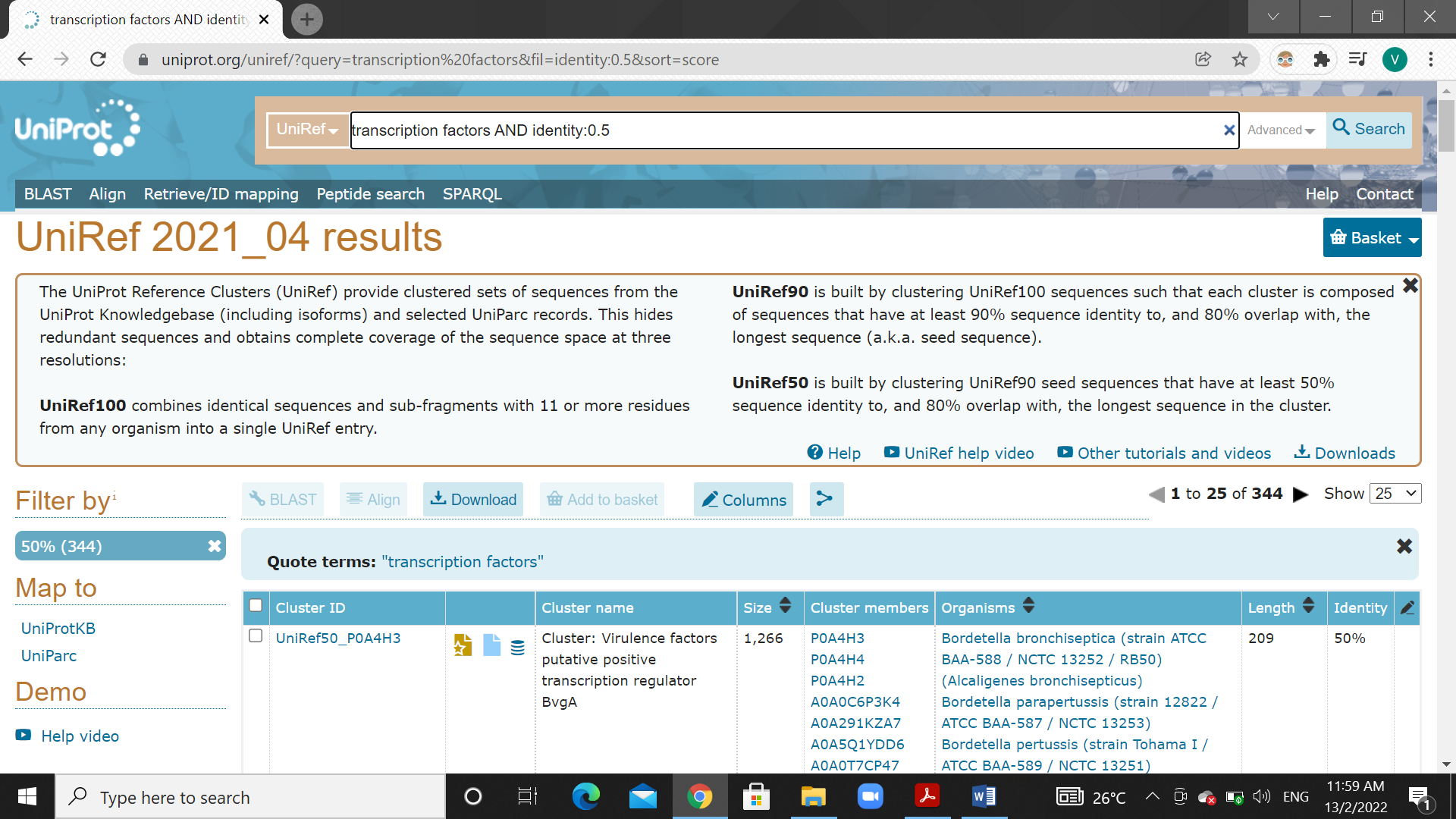
HTHVNDGTEFGGSIYQKVNEKIETSINLAWTAGSNNTRFGIAAKYMLDCRTSLSAKVNNA

SLIGLGYTQTLRPGVKLTLSALIDGKNFSAGGHKVGLGFELEA

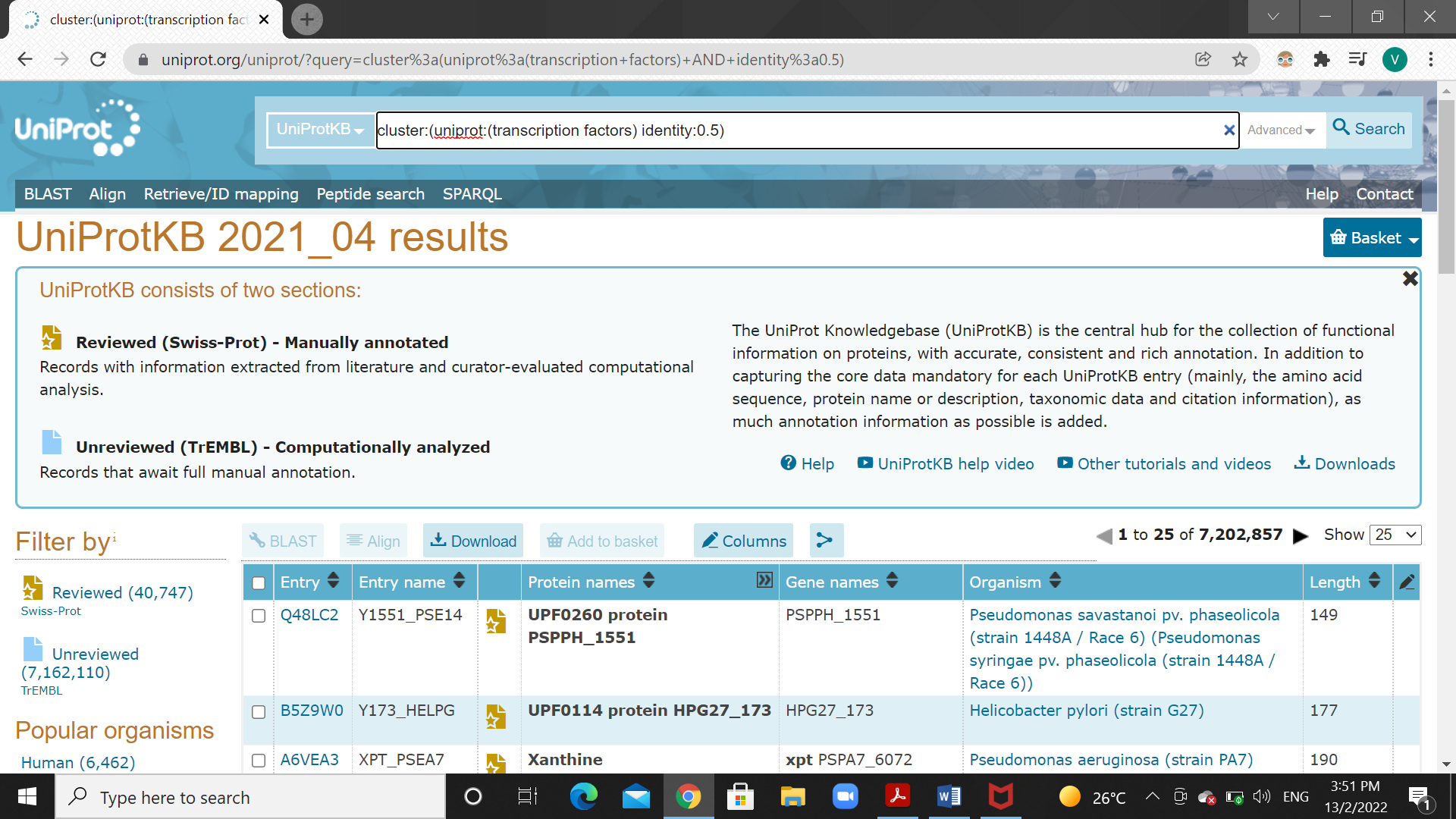
*Function similar to VDAC 1 and VDAC 2.*

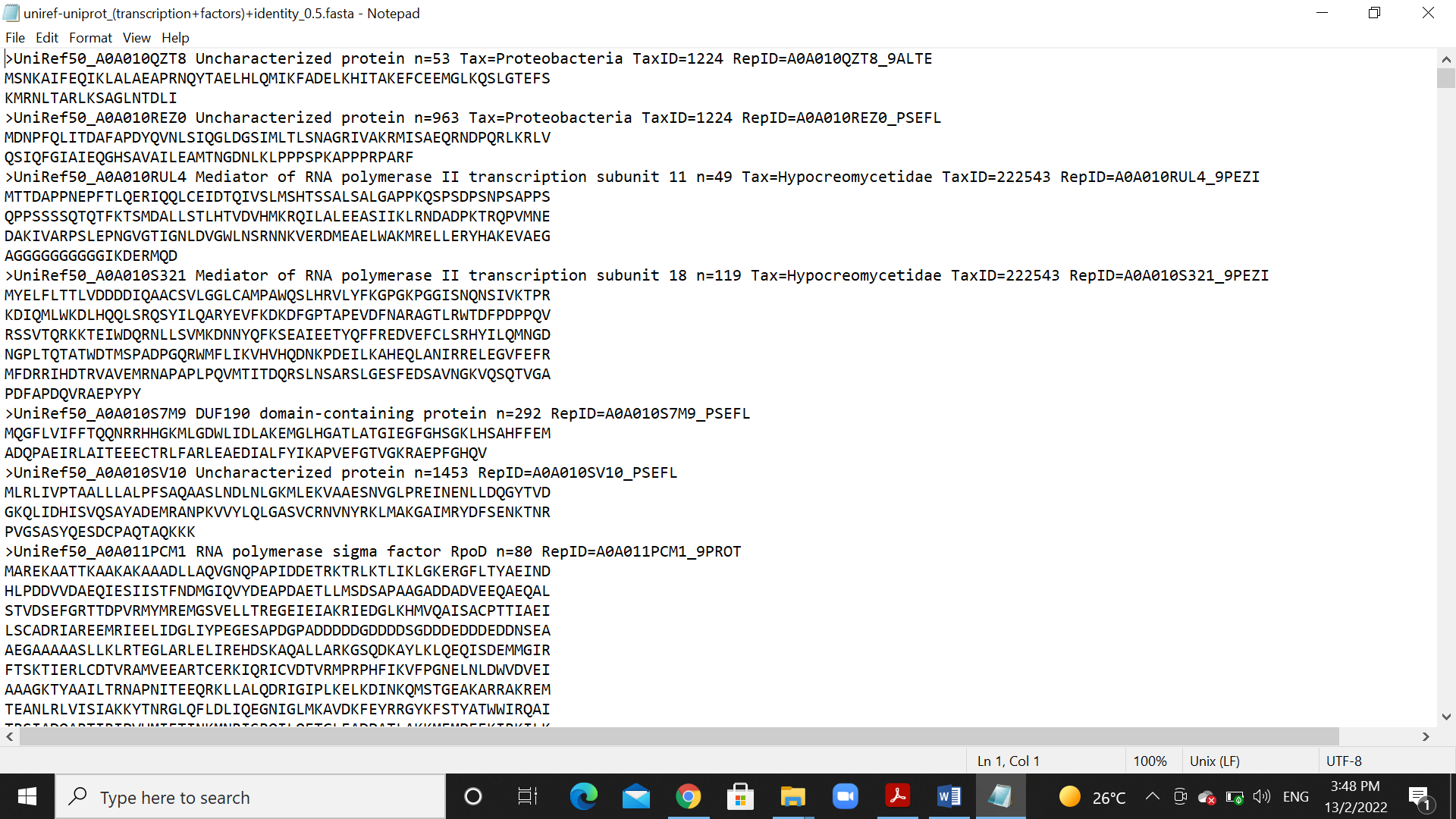
*Only one transmembrane segments.*

*2) The number of clusters for transcription factors with 50% sequence identity is 344.*

**

*The number of sequences for transcription factors with 50% sequence identity is 7,202,857.*

**

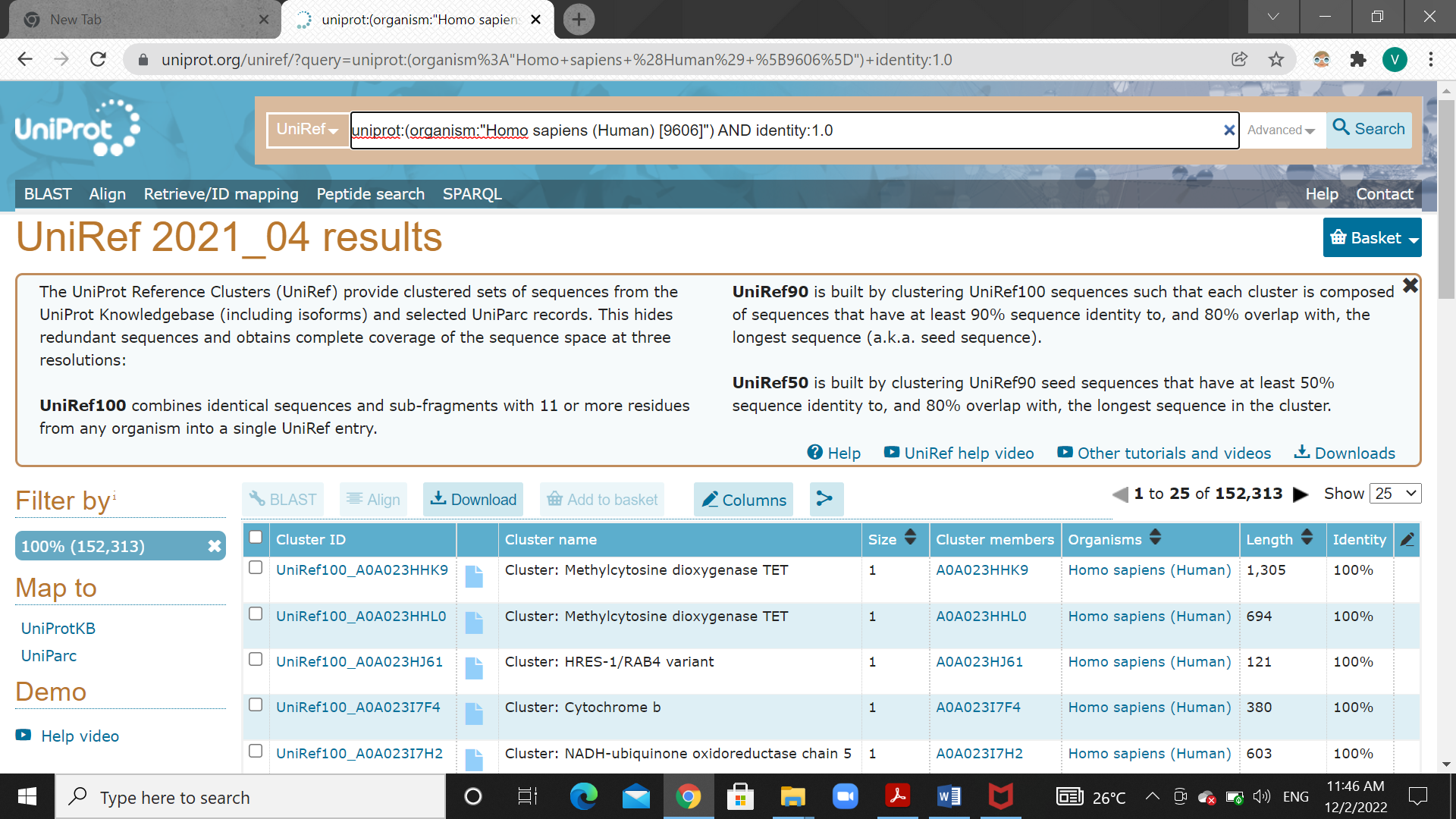
**

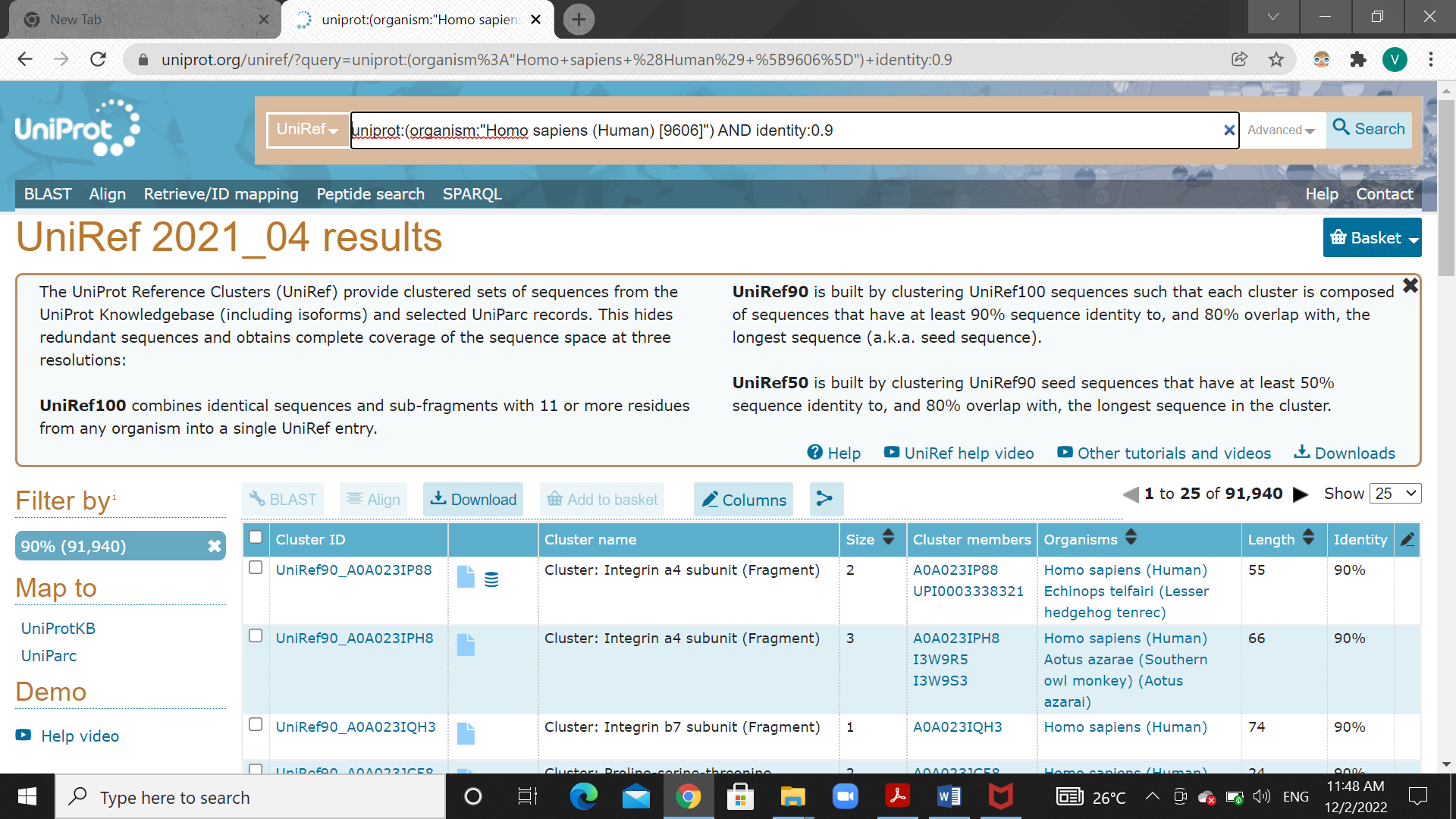
*3) The number of protein sequences from Homo sapiens that are obtained at identity cutoff of 100%, 90% and 50% sequence identity are:*

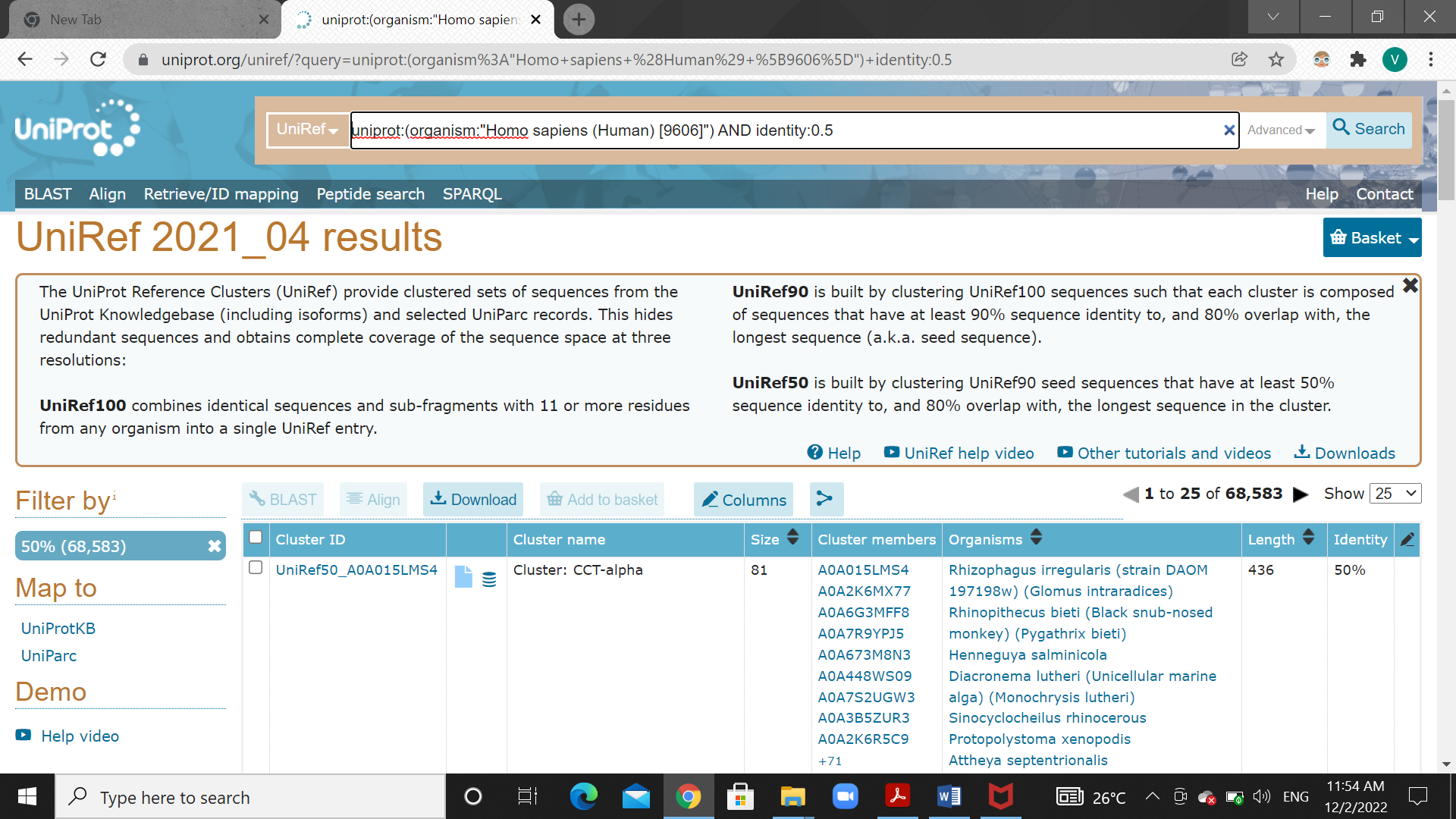
*100%: 152,313*

*90%: 91,940*

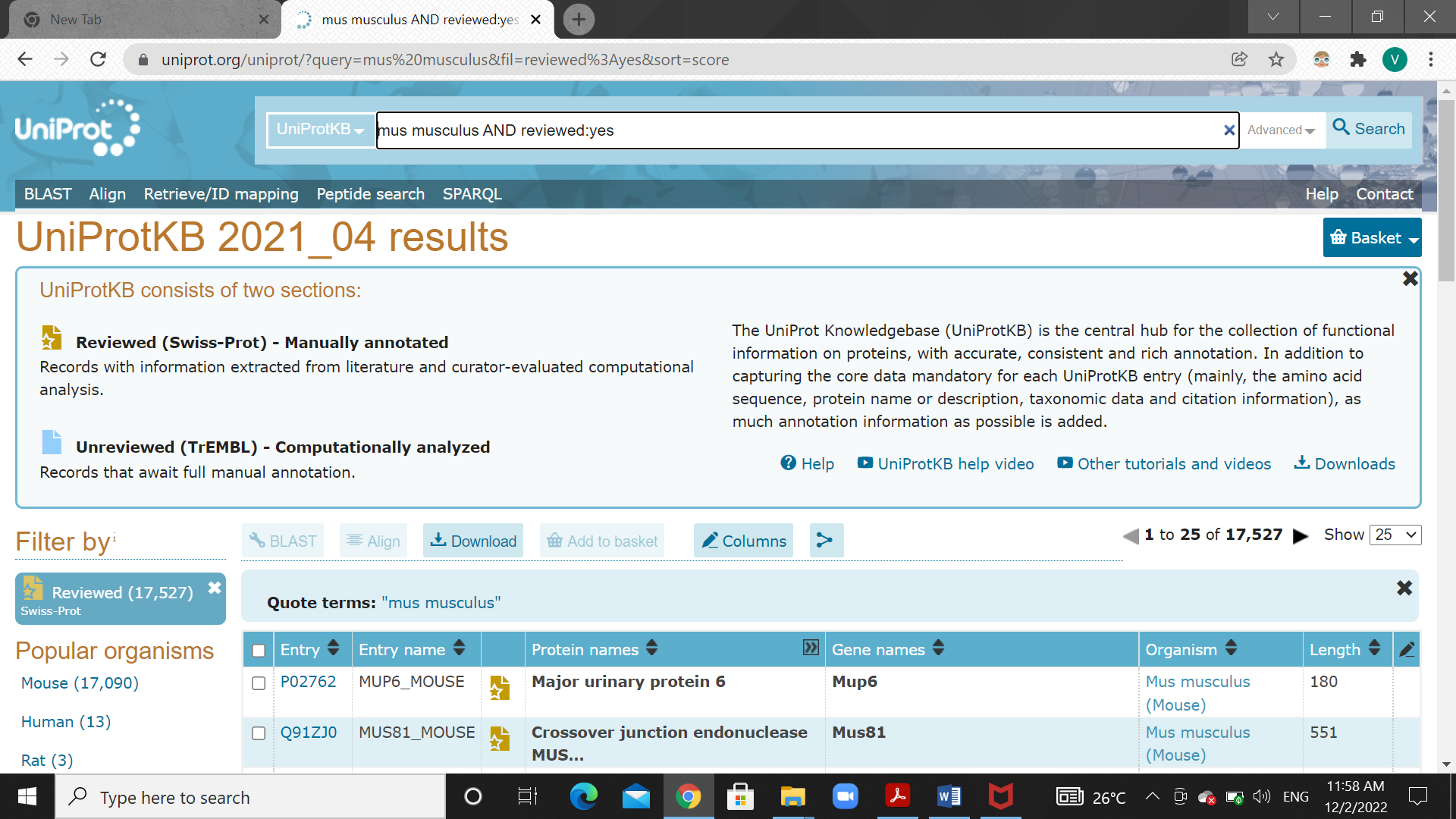
*50%: 68,583*

**

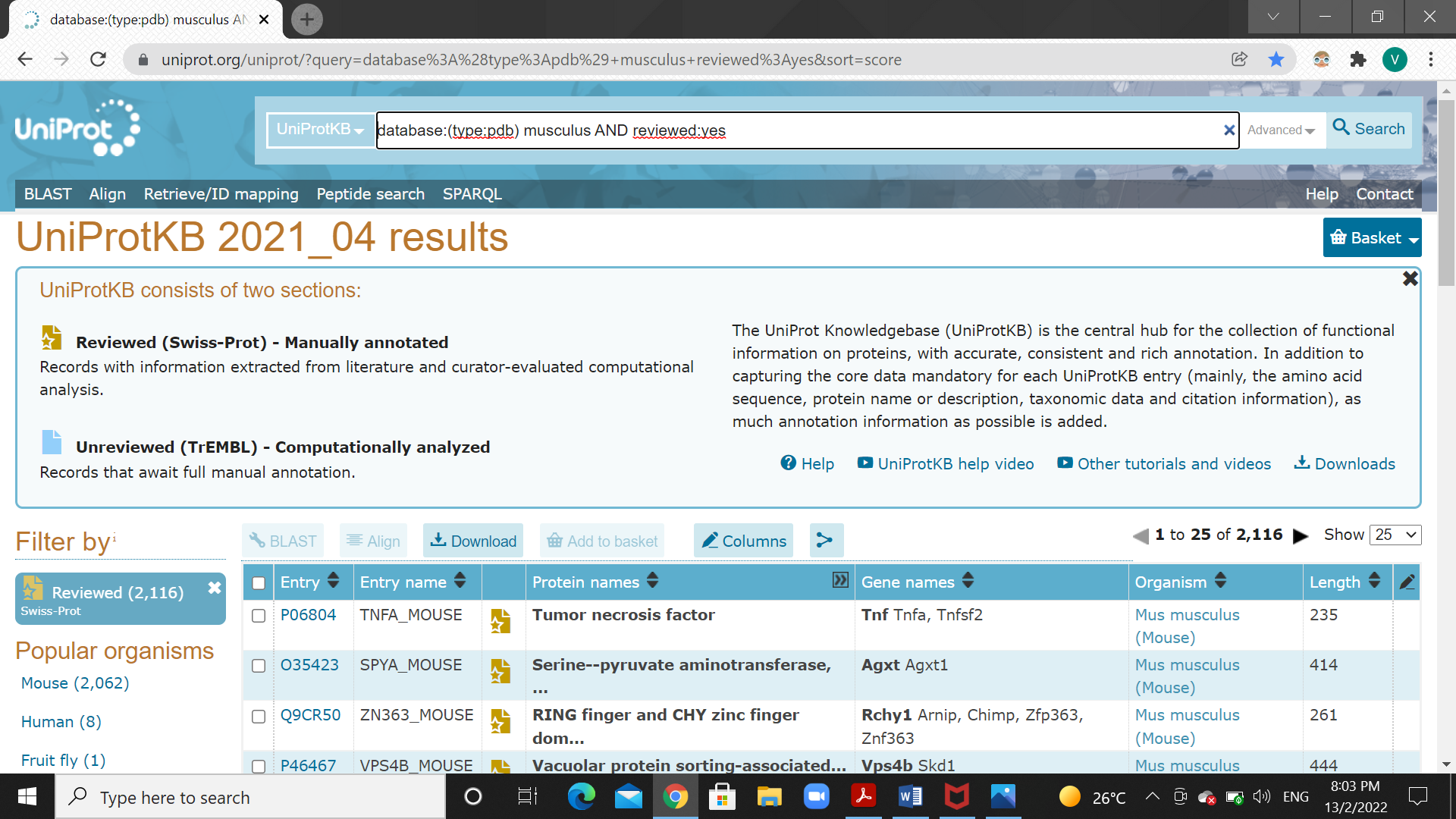
**

**

*4) In UniProt, the number of mouse (Mus musculus) protein sequences that are manually annotated is 17,527.*

**

*The number of these manually annotated protein sequences that are associated with PDB (3D structures) is 2,116.*

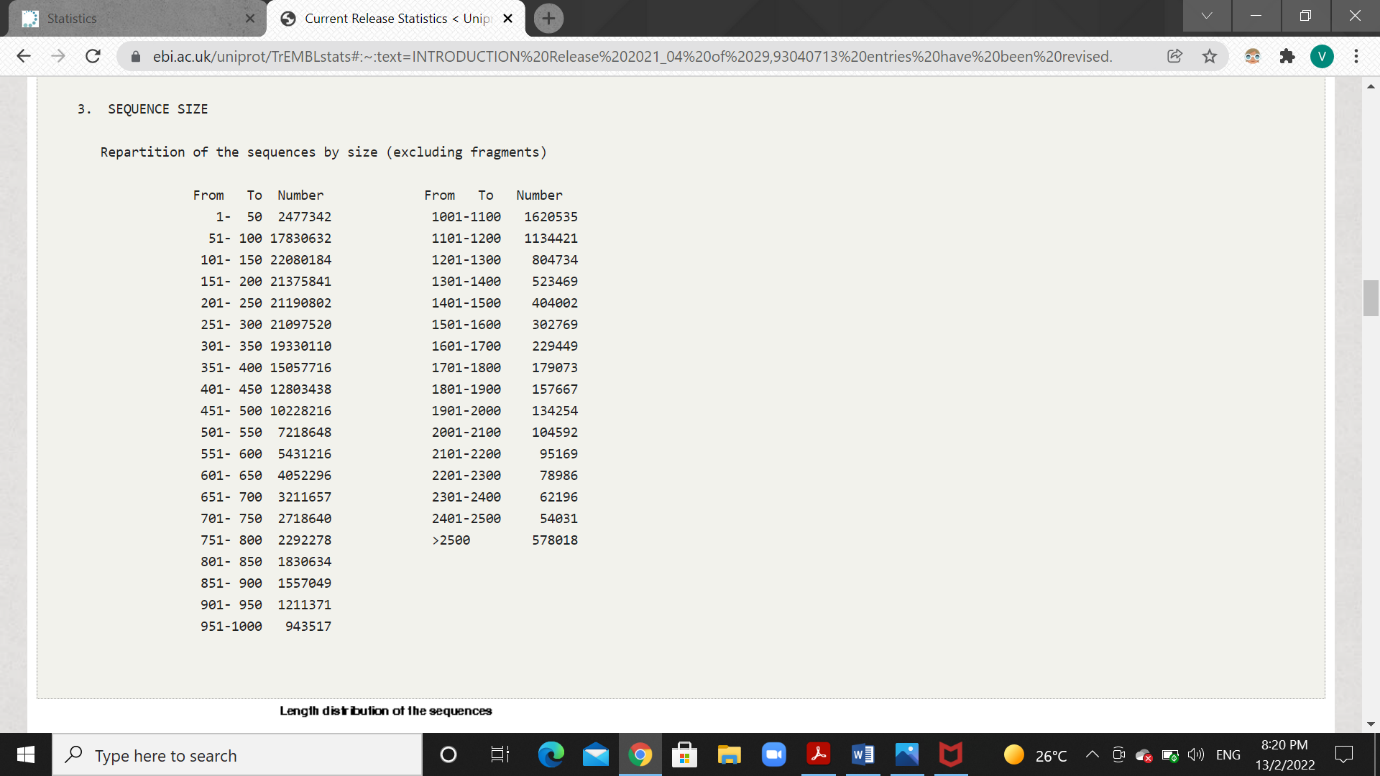
**

*5) Using Retrieve/ID mapping, I mapped the UniProt IDs of the manually curated mouse protein sequences with 3D structures to STRING database.*

*Only one of them successfully mapped.*

**

*6) a) When we look at the distribution of sequence length in UniProt, we could see that most of the sequences are in range 101-150. The average sequence length of the sequences is 348 amino acids.*

**

**

*b) The shortest sequence is A0A0G2JLF7\_HUMAN of length 7 amino acids.*

*The longest sequence is A0A5A9P0L4\_9TELE of length 45354 amino acids.*

*c) The amino acid composition in percent for the complete database is:*

*Ala (A) - 9.13 Gln (Q) - 3.77 Leu (L) - 9.88 Ser (S) - 6.71*

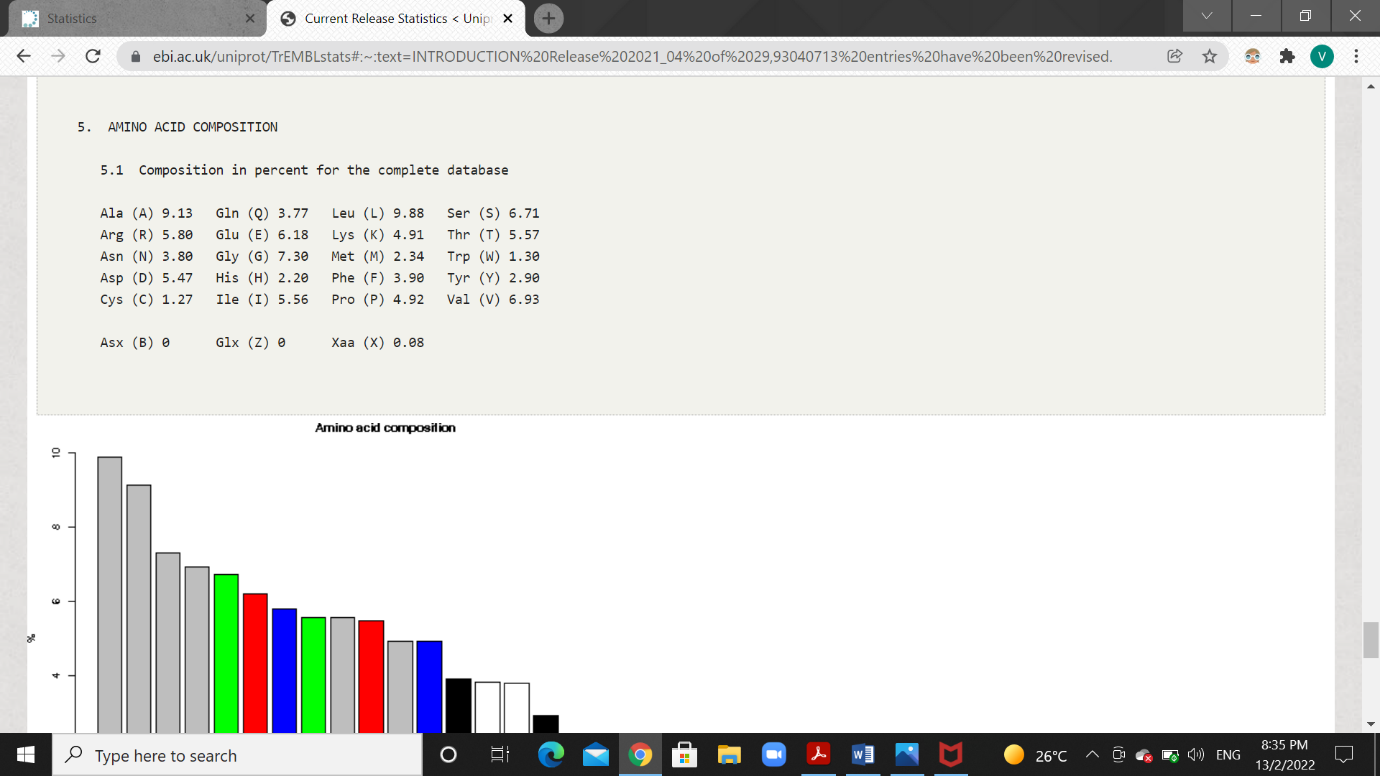
*Arg (R) - 5.80 Glu (E) - 6.18 Lys (K) - 4.91 Thr (T) - 5.57*

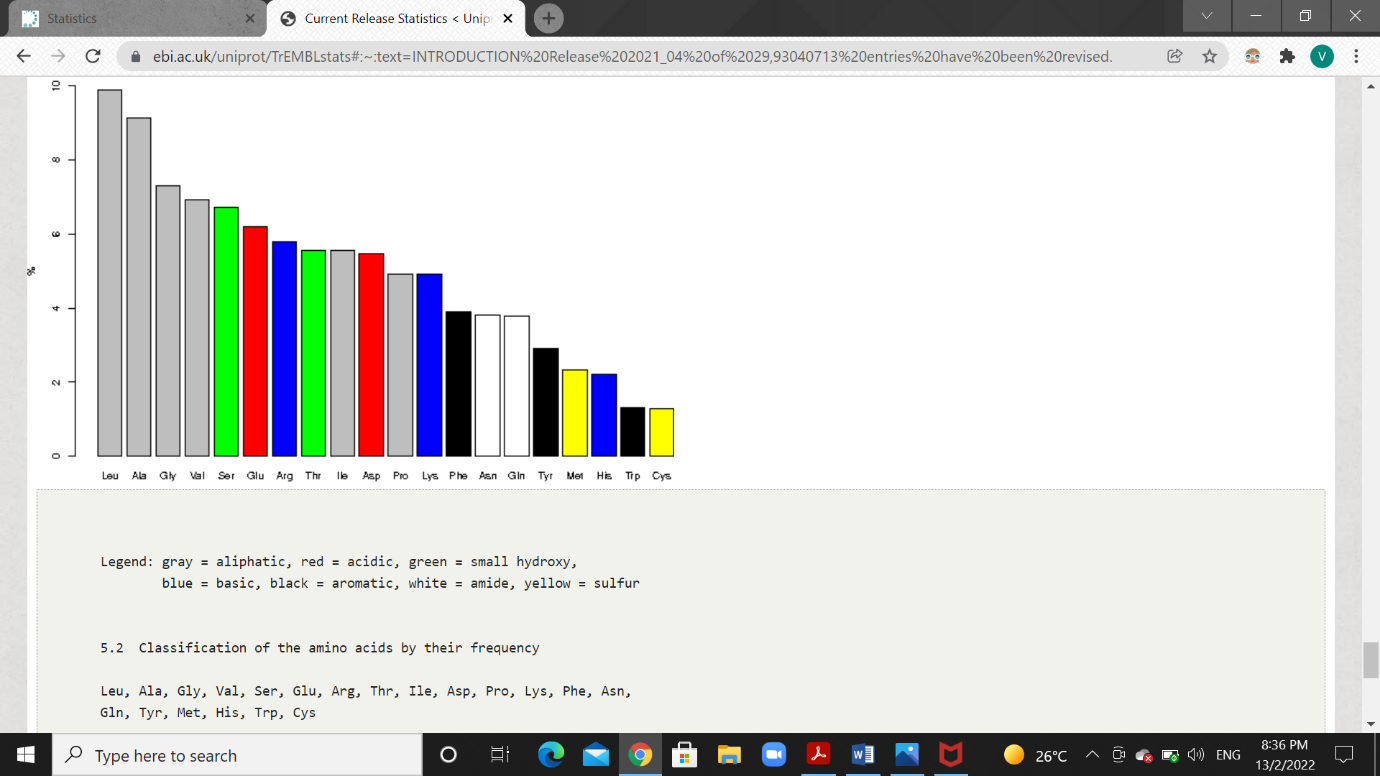
*Asn (N) - 3.80 Gly (G) - 7.30 Met (M) - 2.34 Trp (W) - 1.30*

*Asp (D) - 5.47 His (H) - 2.20 Phe (F) - 3.90 Tyr (Y) - 2.90*

*Cys (C) - 1.27 Ile (I) - 5.56 Pro (P) - 4.92 Val (V) - 6.93*

*Asx (B) - 0 Glx (Z) - 0 Xaa (X) - 0.08*

**

**