***Practicals – 2***

*-BS19B032*

*-R. Vasantha Kumar*

*1) Number of Homo sapiens sequences in different databases are:*

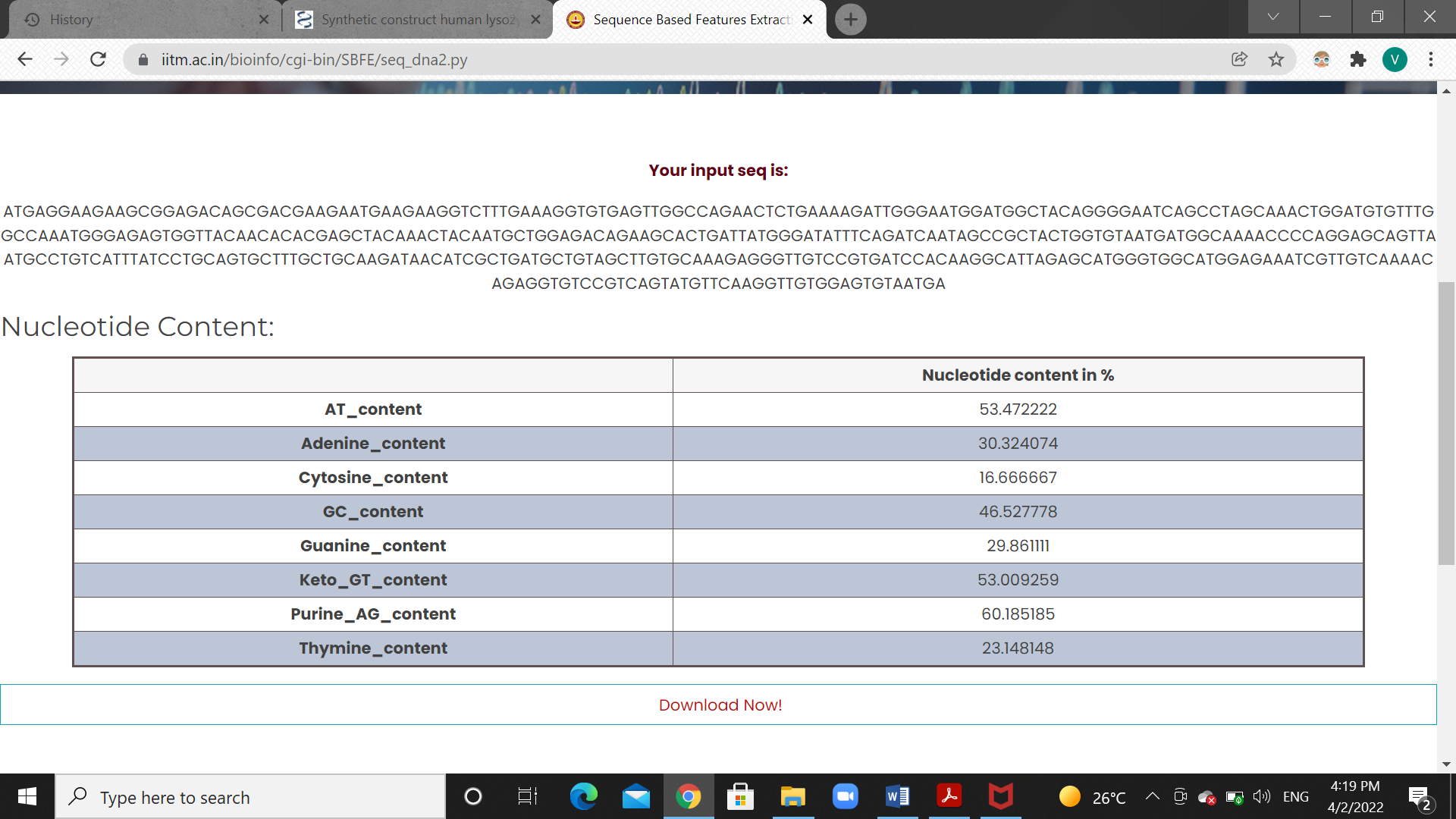
*DDBJ: 18296211*

*GenBank: 189,232,925*

*EMBL: 129,325,718*

*2) AY330867 is the locus for synthetic construct human lysozyme mRNA, which I obtained from ncbi database. Then, by using the seq2feature tool, I found the GC content of the sequence.*

*The GC content of AY330867 is 46.527778.*

**

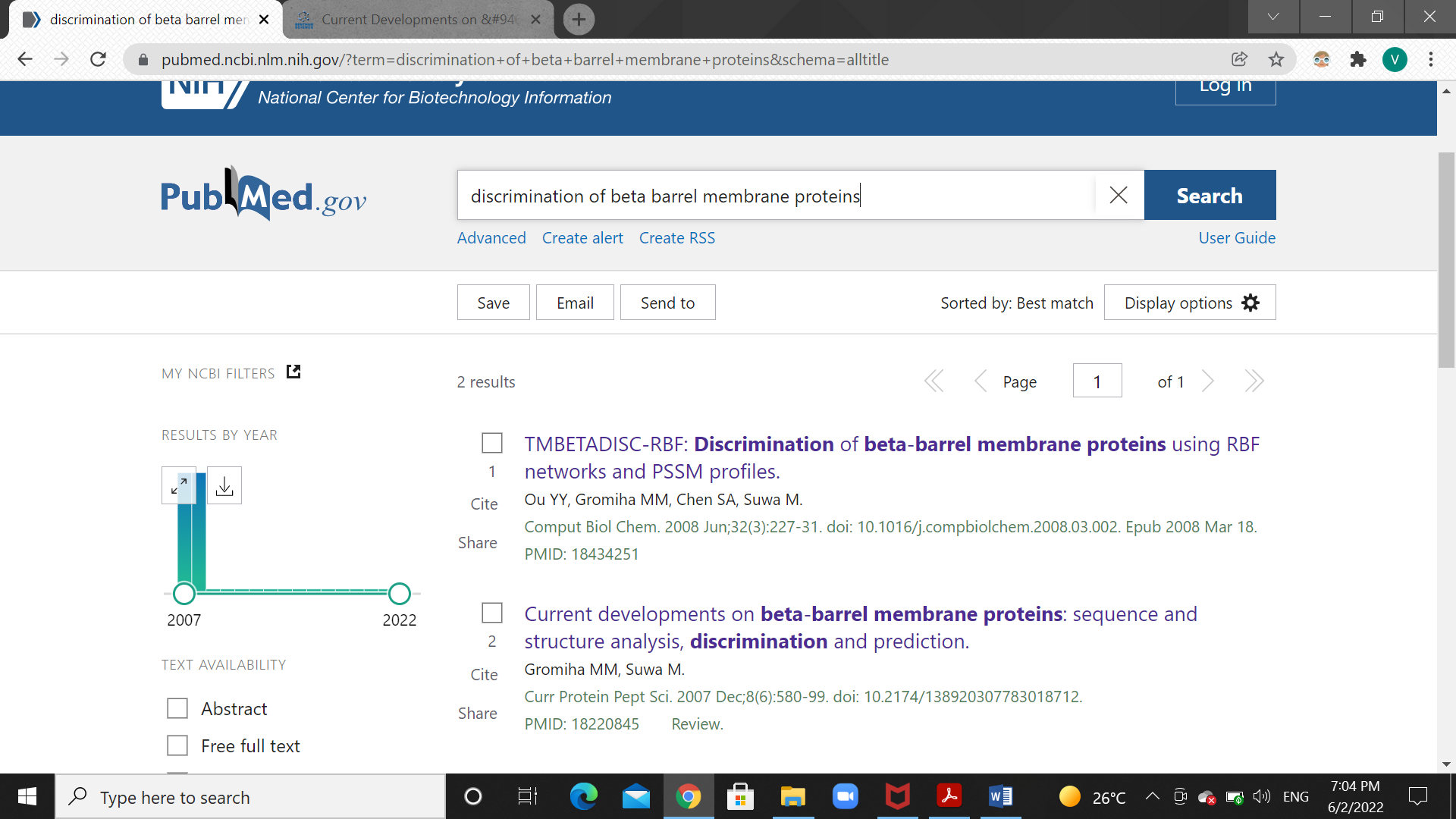
*3) DDBJ, EMBL and GenBank all are DNA databases collecting nucleic acid sequence data.*

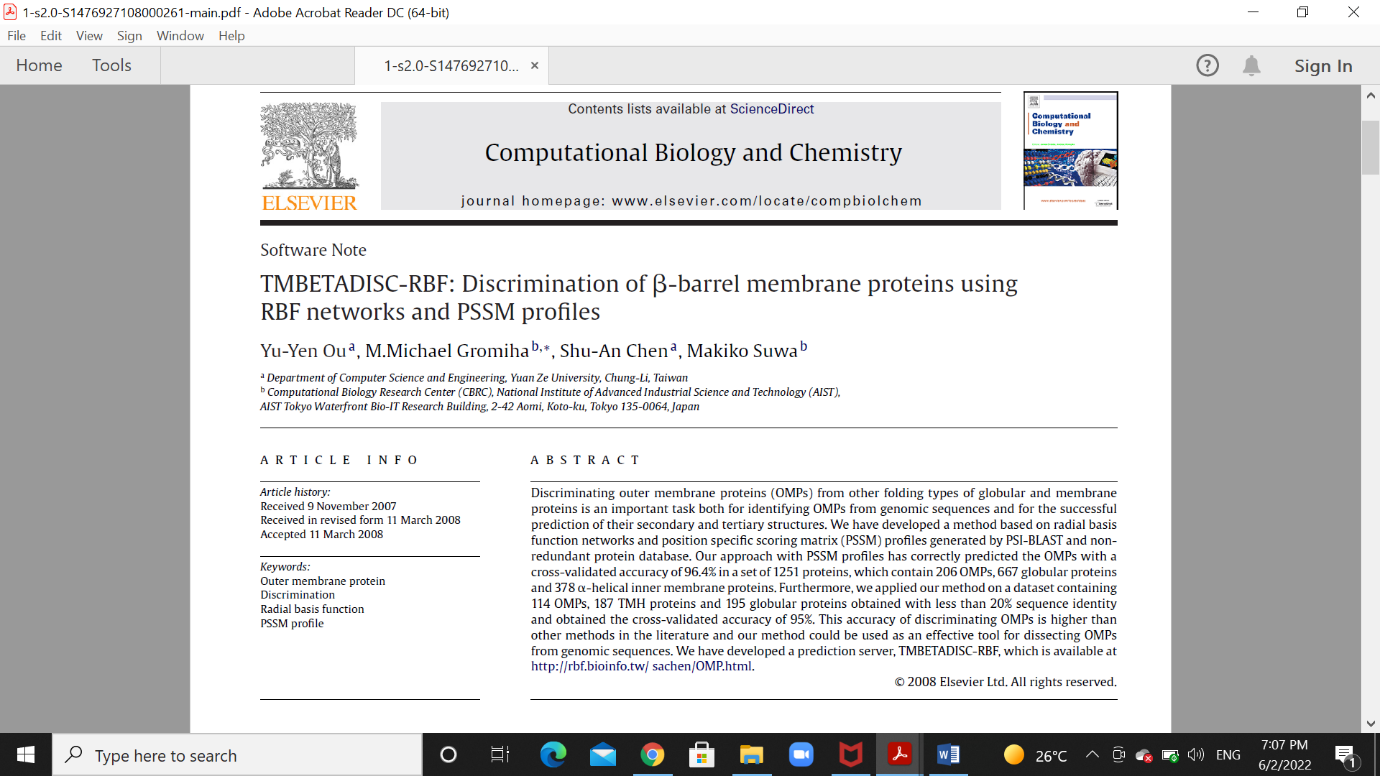
*All the three databases have similar DNA data, but the differences come at the interface and the searching methods.*

*For example, the GenBank database allows us to search for specific sequences in two ways: one is the normal text search and the other is using molecular sequences comparing them using a blast. EMBL also has similar kind of interface.*

*In DDBJ, anyone from any country can add DNA data, but the GenBank database only allows the author to alter or add DNA data.*

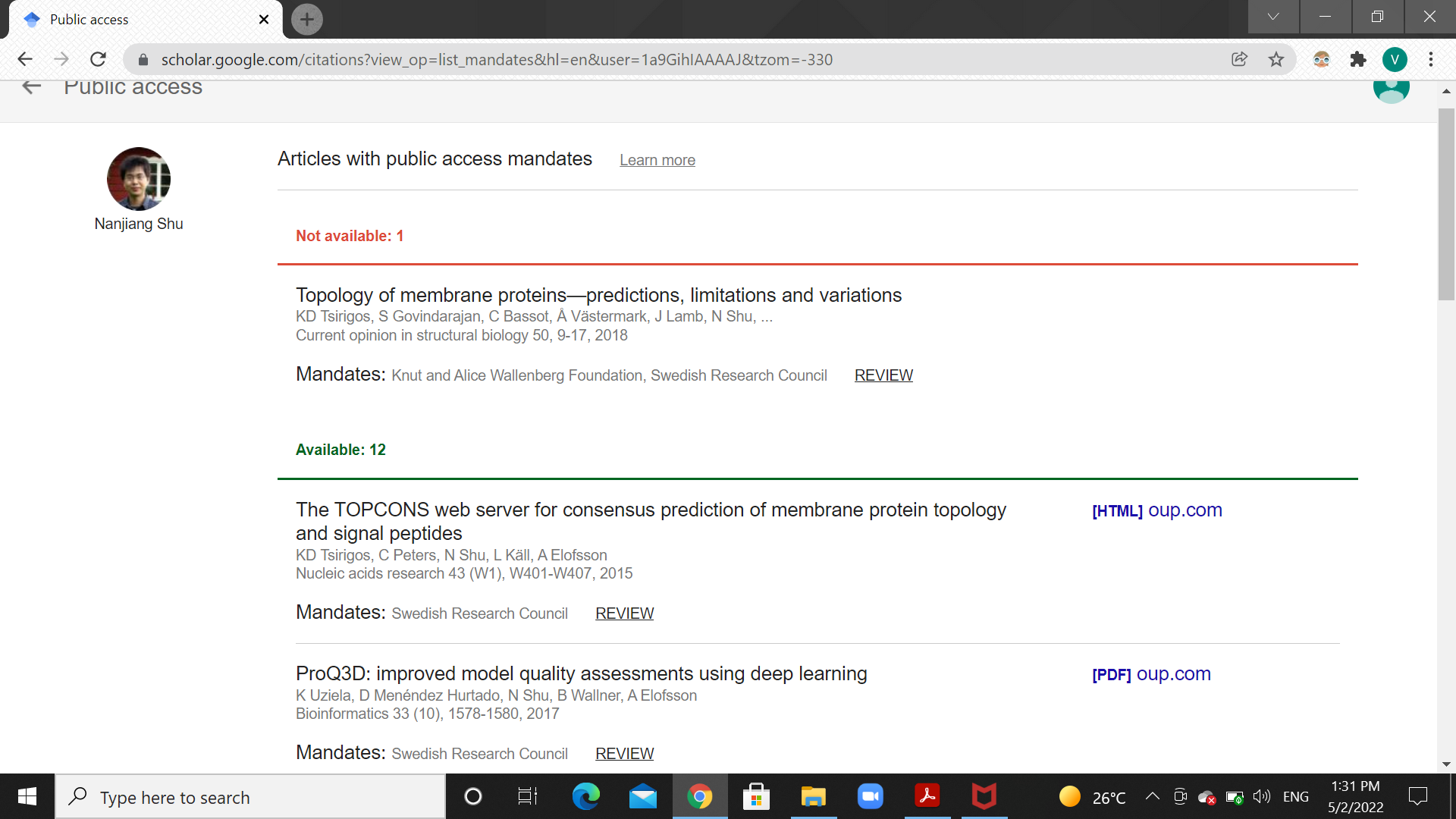
*4) From pubmed, I got the papers about discrimination of beta barrel membrane proteins.*

**

**

*5) Using the google scholar database, I found the articles published by Nanjiang Shu.*

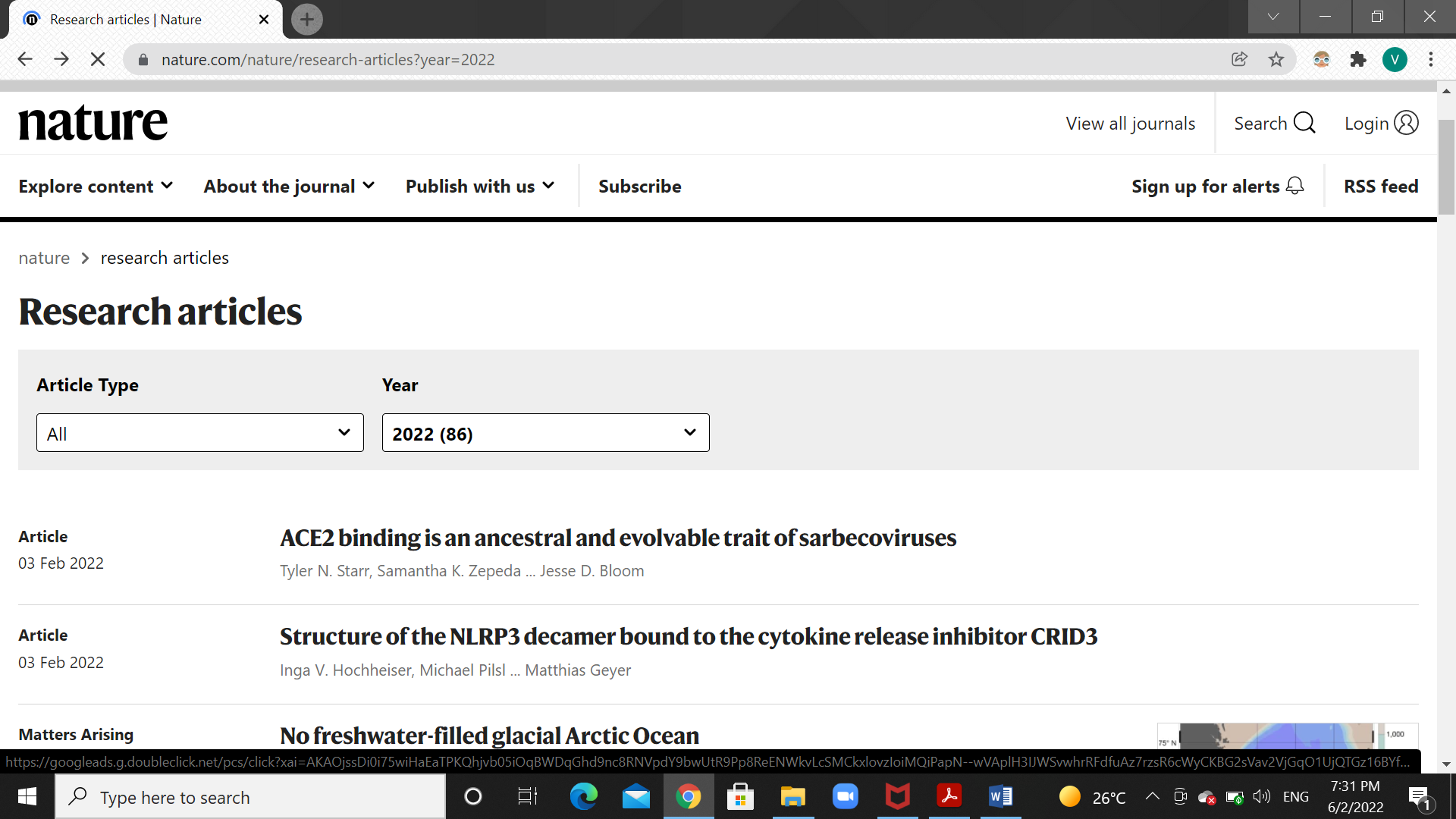
* Topology of membrane proteins—predictions, limitations and variations
* The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides
* ProQ3D: improved model quality assessments using deep learning
* Inclusion of dyad-repeat pattern improves topology prediction of transmembrane β-barrel proteins
* Improved topology prediction using the terminal hydrophobic helices rule
* SubCons: a new ensemble method for improved human subcellular localization predictions
* Improved protein model quality assessments by changing the target function
* Large tilts in transmembrane helices can be induced during tertiary structure formation
* KalignP: Improved multiple sequence alignments using position specific gap penalties in Kalign2
* The evolutionary history of topological variations in the CPA/AT superfamily

**

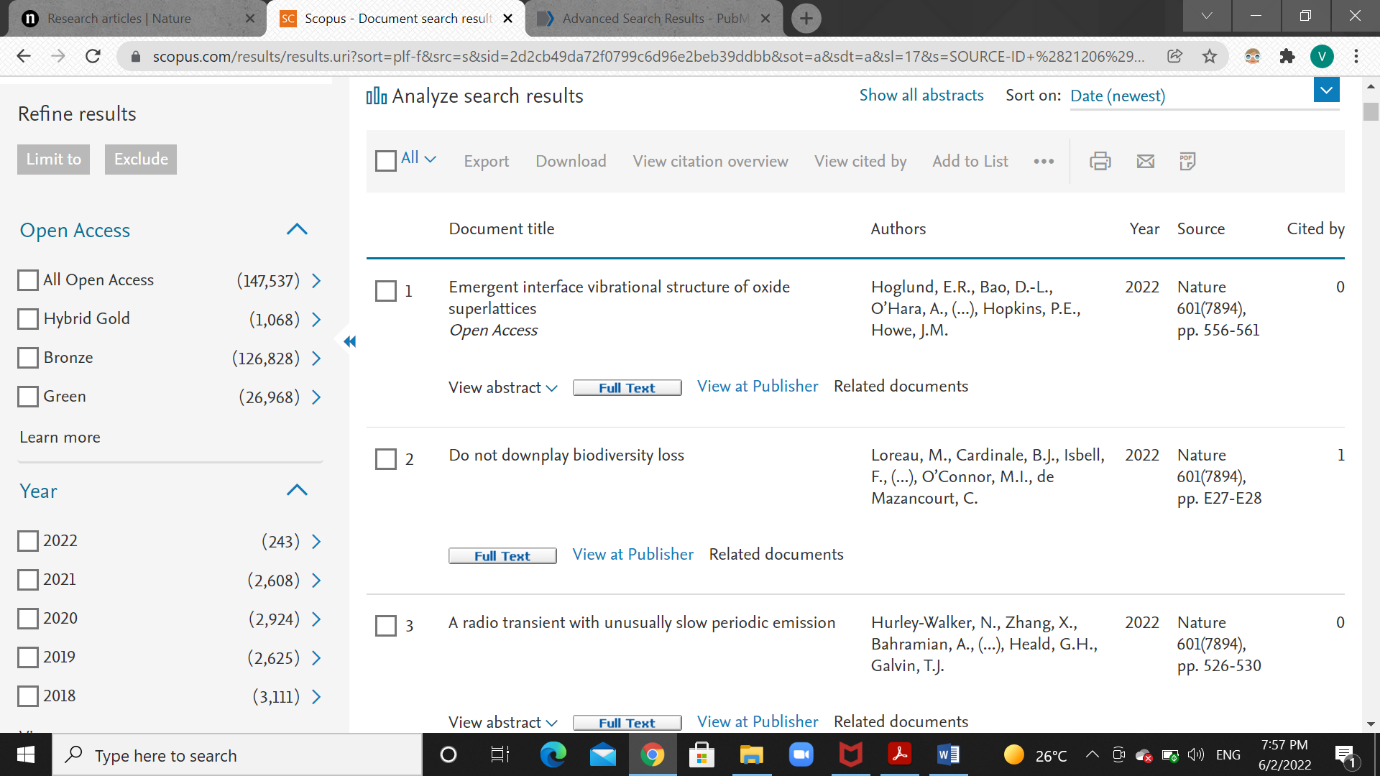
*6) The number of related articles that are listed in PUBMED for the paper - Cell 2008 Dec 26;135(7):1158-9 is 28.*

**

*7) The number of research articles published by nature journal, according to their website in 2022 is 86.*

**

*According to the SCOPUS website, the number of documents published by nature journal in 2022 is 243.*

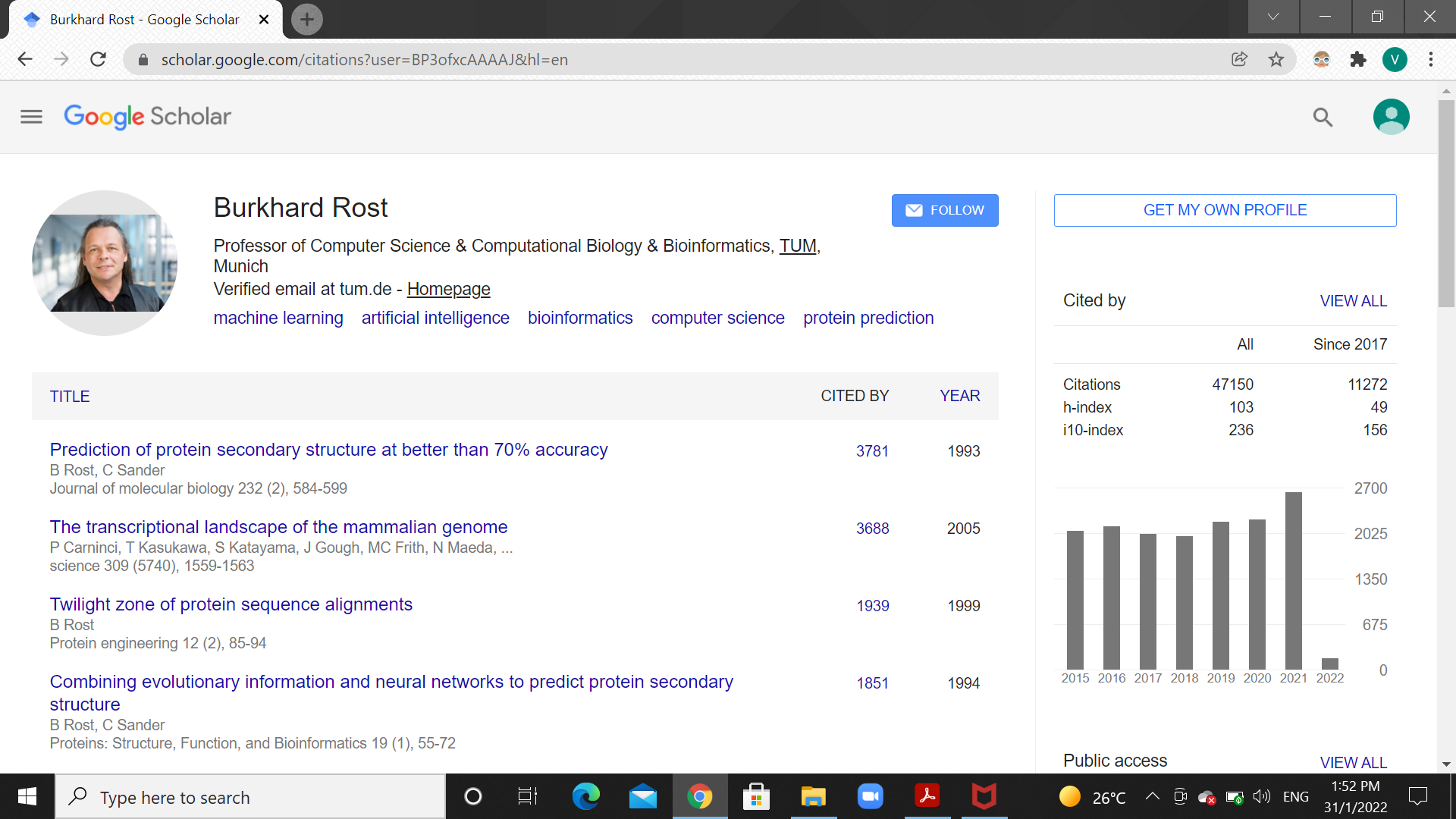
**

*Using the ISSN number of nature journal(1476-4687), I got the articles published by nature journal. Total articles of nature journal for year 2022 is 41.*

*8) In the google scholar website, I found the h-index and number of citations for “Burkhard Rost”.*

*h-index: 103(since 2017-49)*

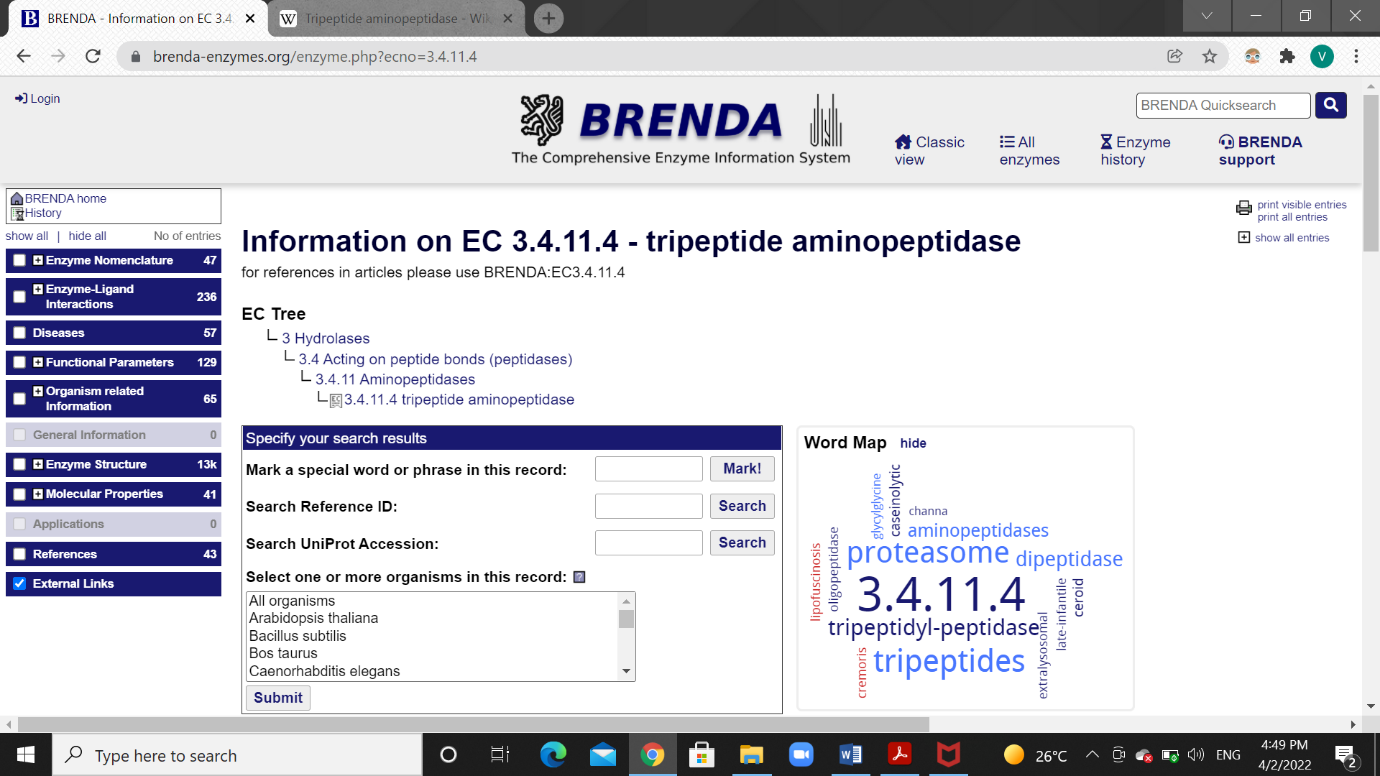
*number of citations: 47150(since 2017-11272)*

**

*9) EC 3.4.11.4 is actually tripeptide aminopeptidase enzyme, in the brenda-enzymes database.*

*This enzyme belongs to the class of enzymes called Hydrolases.*

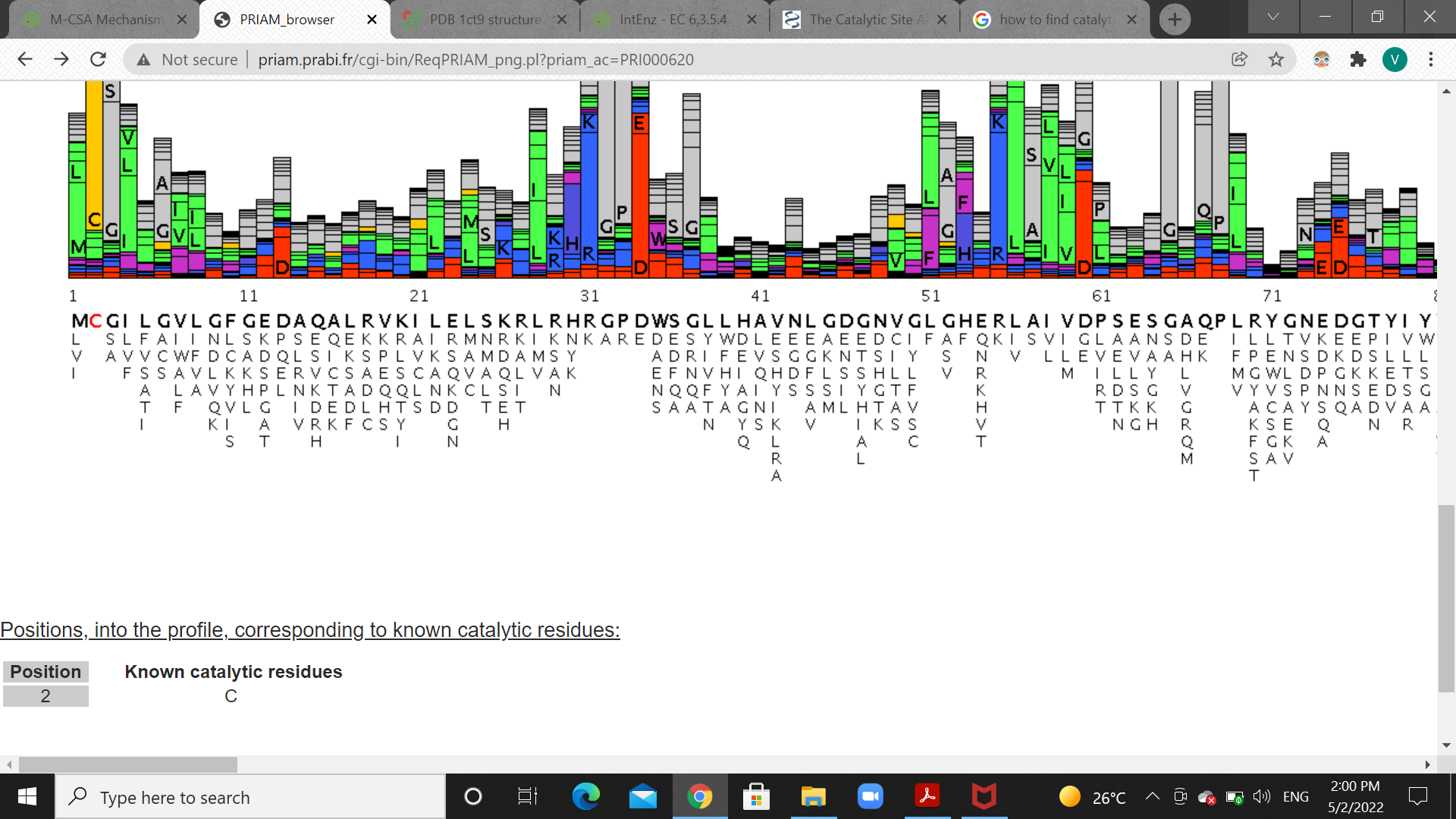
*This enzyme acts on peptide bonds. Its main function is to release the N-terminal residue from the tripeptide.*

**

*10) From the link given, I found the EC number of Asparagine synthetase, that is, 6.3.5.4.*

*Then in the catalytic site atlas I found the catalytic residues:*

*It is found in site 2, as catalytic residue C.*

**

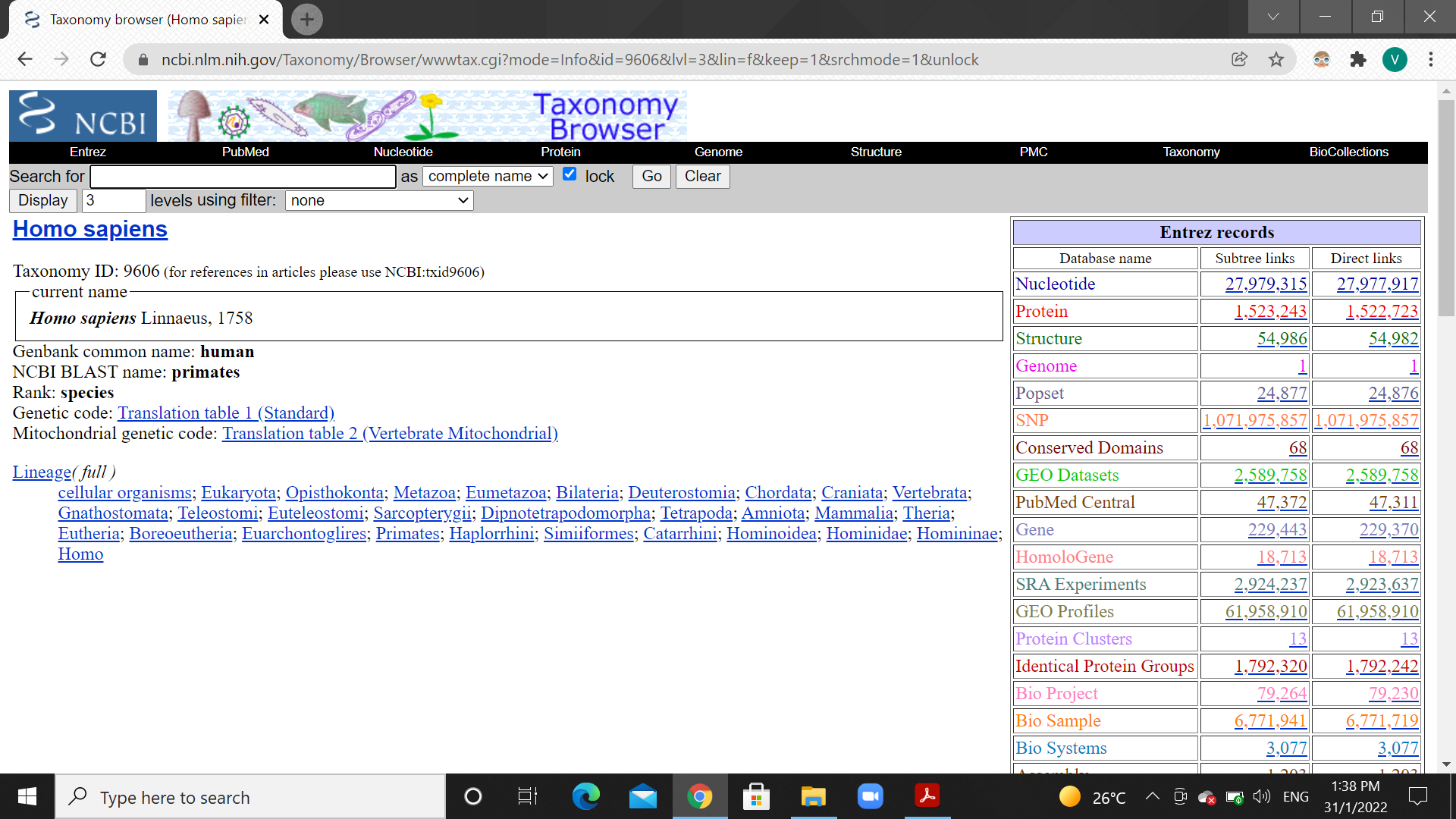
*11) From the ncbi database, and using its taxonomy browser, I found the taxonomy IDs of given organisms.*

* *Human:*

*Scientific name: Homo sapiens*

*Taxonomy ID : 9606*

*No. of chromosomes: 46(23 pairs)*

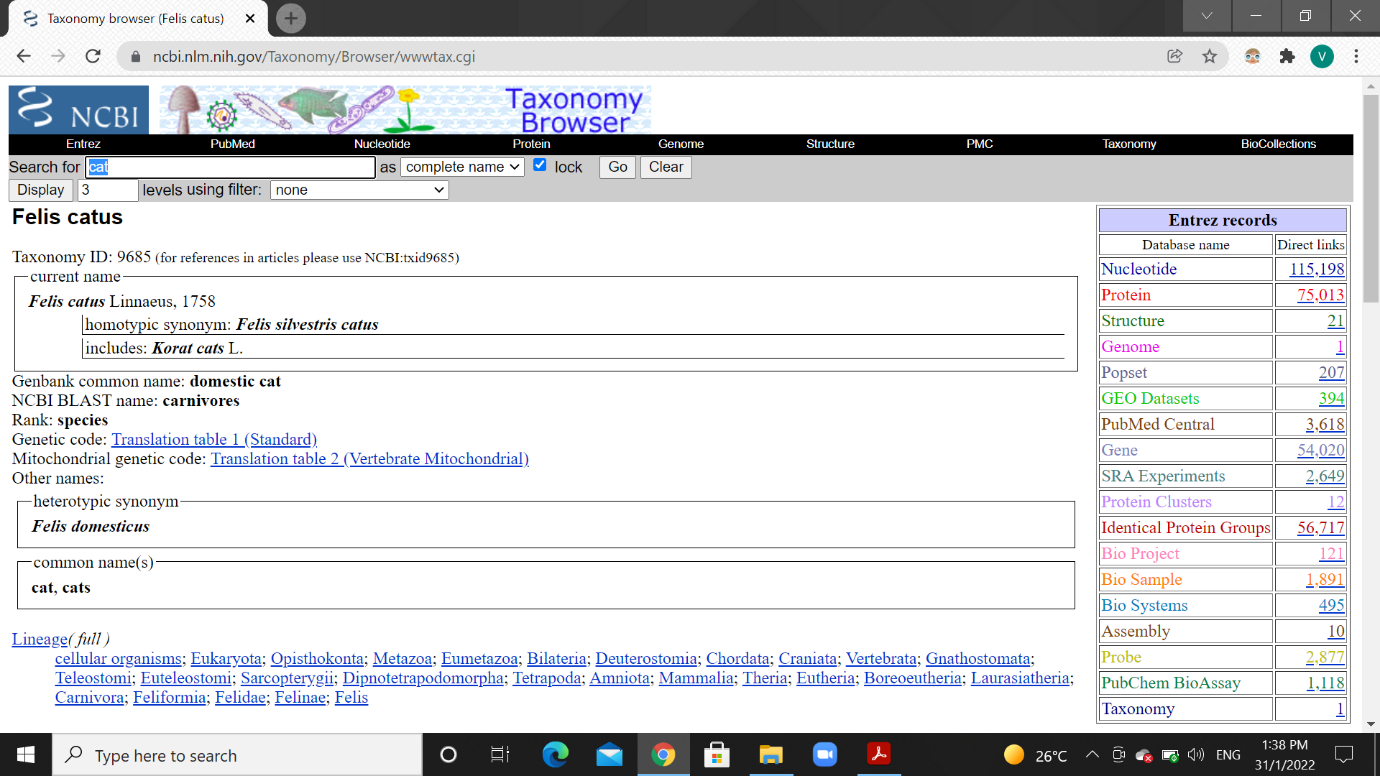


* *Cat:*

*Scientific name: Felis catus*

*Taxonomy ID : 9685*

*No. of chromosomes: 38(18 pairs, 1X, 1Y=20)*

**

* *Dog:*

*Scientific name: Canis lupus familiaris*

*Taxonomy ID : 9615*

*No. of chromosomes: 78(38 pairs, 1X, 1Y=40)*

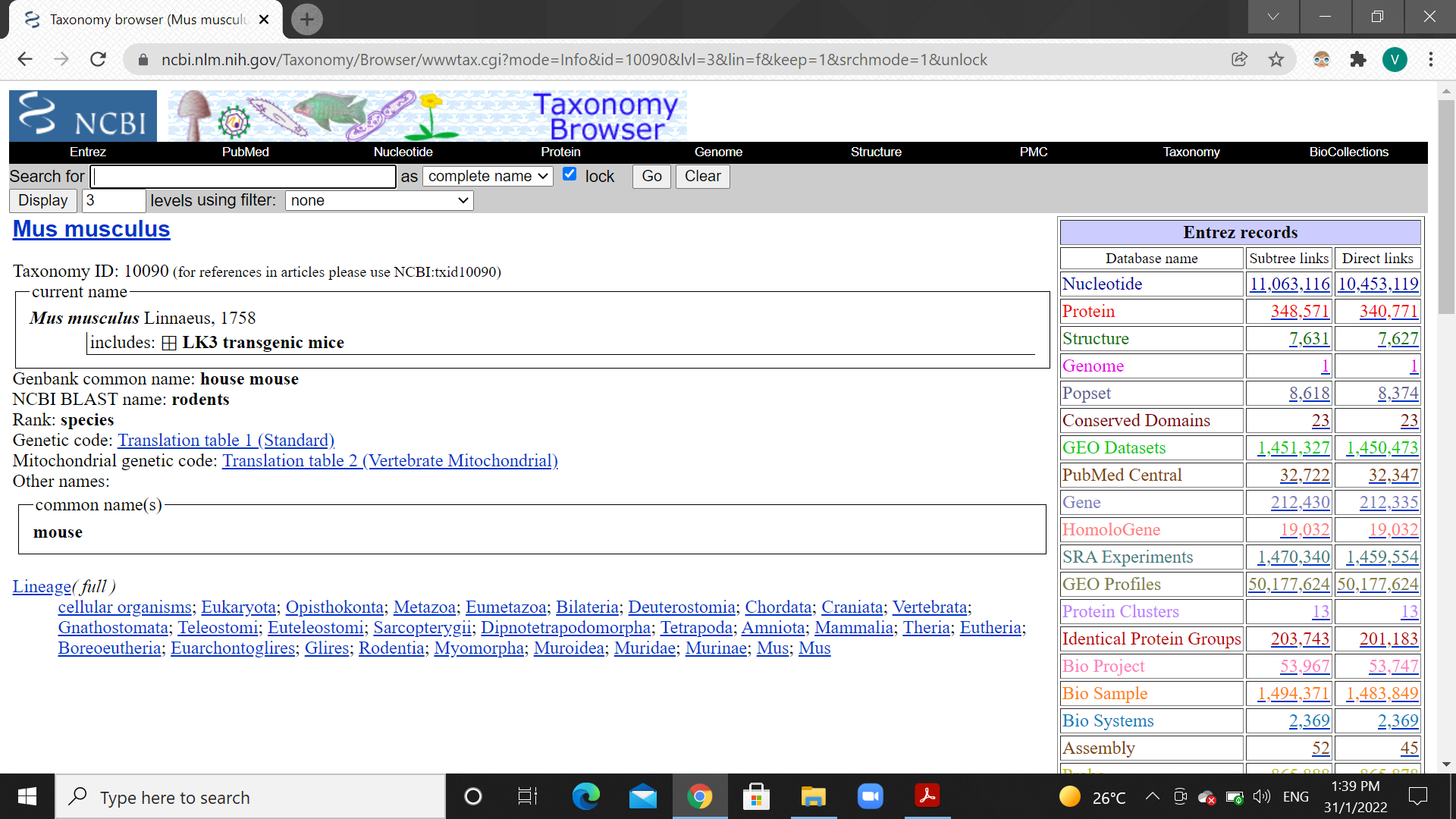
**

* *House mouse:*

*Scientific name: Mus musculus*

*Taxonomy ID : 10090*

*No. of chromosomes: 40(20 pairs)*

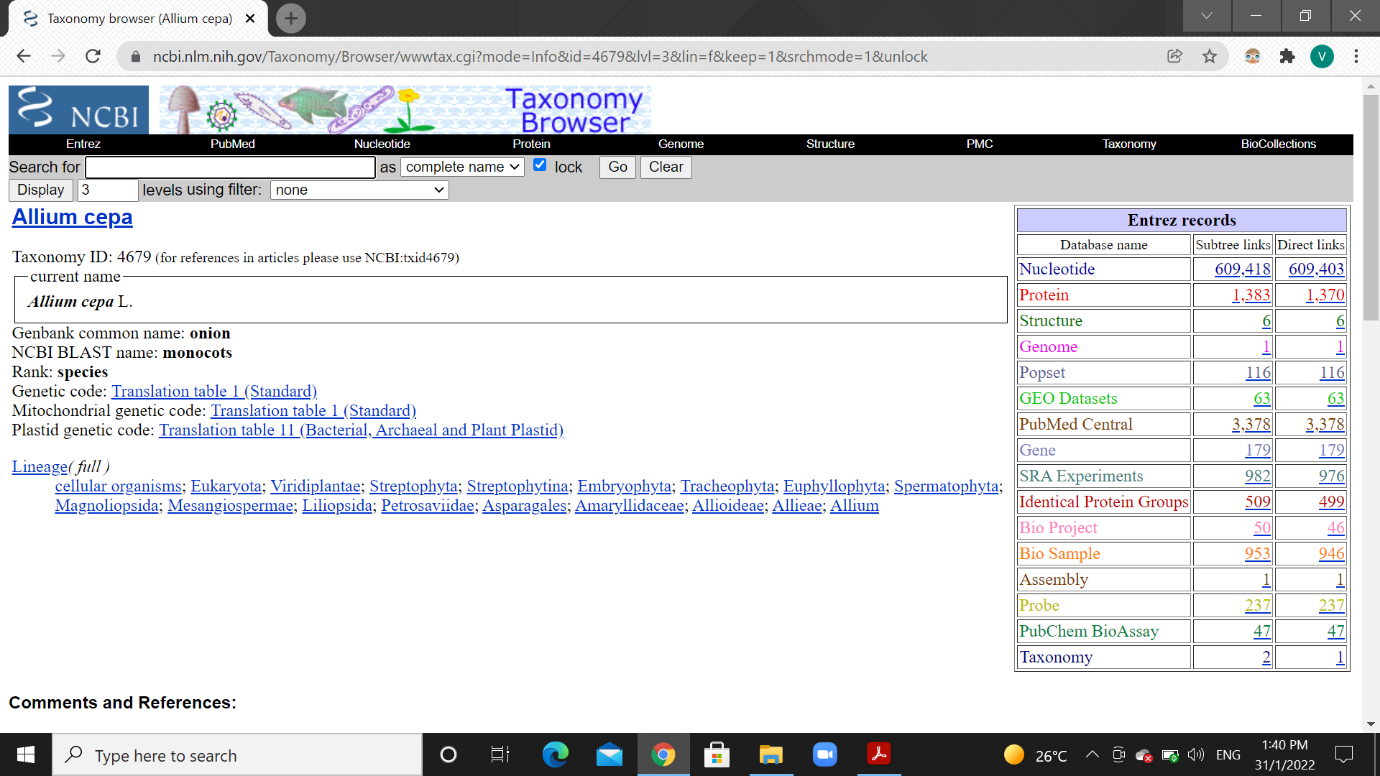
**

* *Onion:*

*Scientific name: Allium cepa*

*Taxonomy ID : 4679*

*No. of chromosomes: 16(8 pairs)*

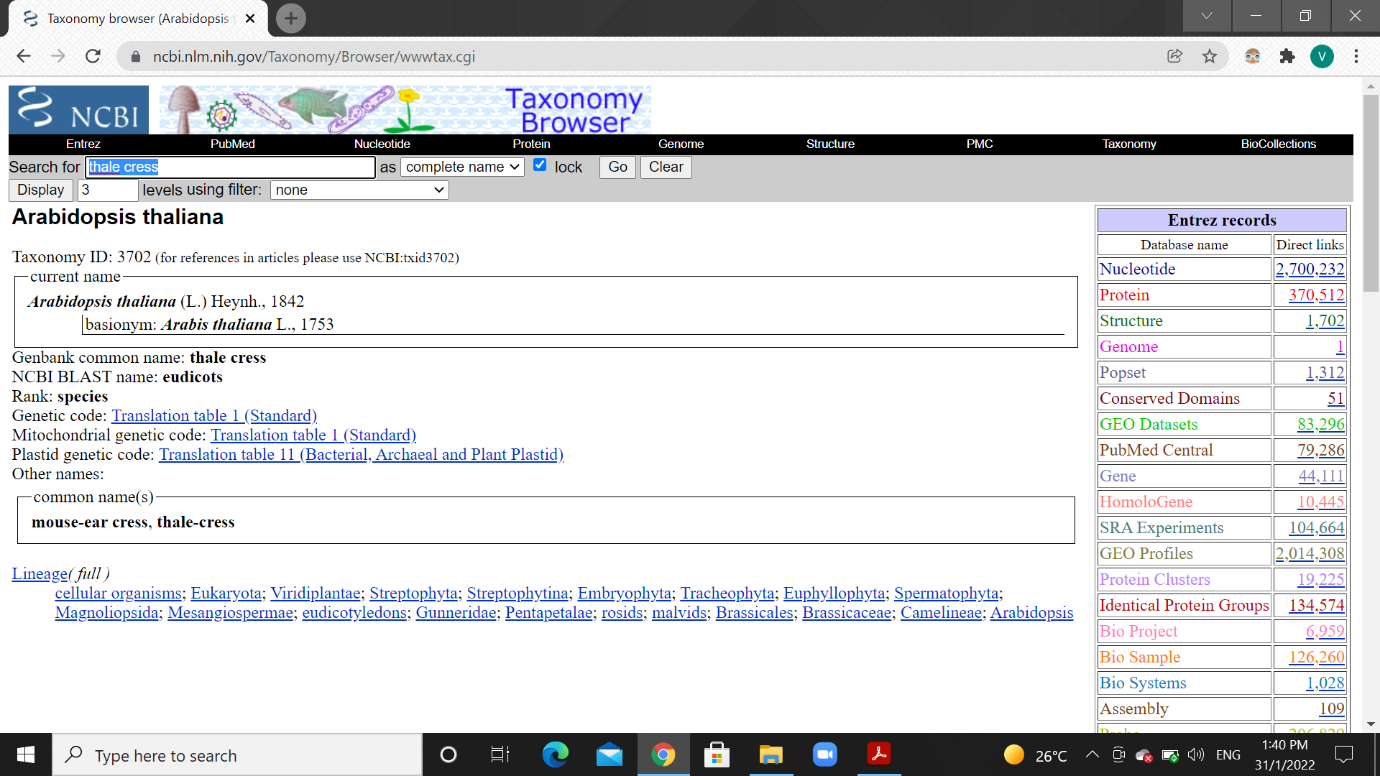
**

* *Thale cress:*

*Scientific name: Arabidopsis thaliana*

*Taxonomy ID : 3702*

*No. of chromosomes: 5*

**

*12) NCBI E-utilities are nine set of server side programs, which provide stable interface for Entrez query and for accessing the database in NCBI. The E-utilities are set of simple server syntax codes which helps us to retrieve or fetch the requested data.*

*This is an example syntax for fetching homo sapiens related sequences, from ncbi in fasta format.*

*esearch -db ncbi -query "homo sapiens sequences"*

*elink -receptor gene*

*efetch -format fasta*

*13) Using Nucleic Acids Research (NAR), database category list, I found different databases for given categories:*

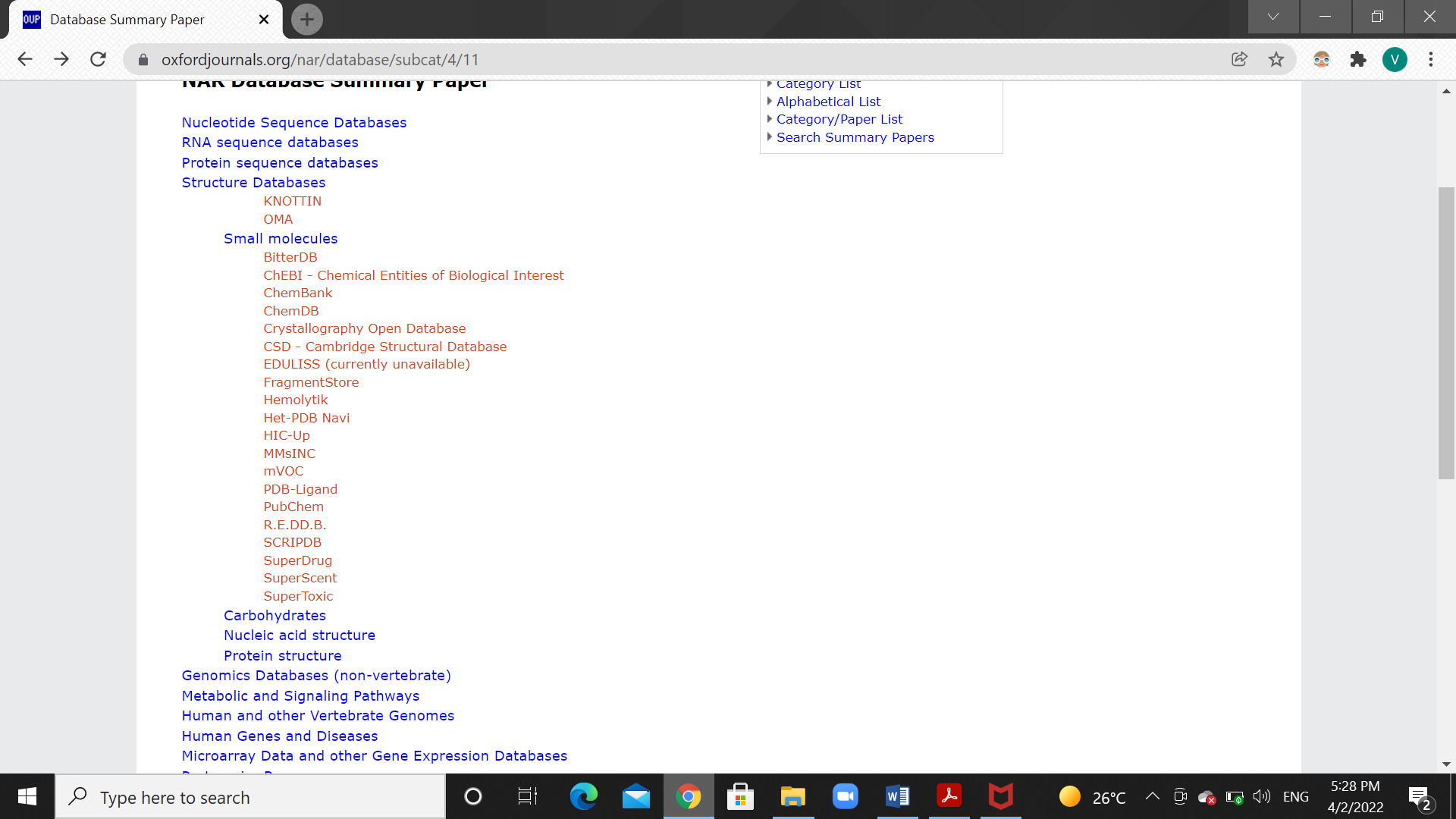
*a) Protein properties:*

* + - *HHMD – Human Histone Modification Database*
    - *PIDD – Database for protein modelling*

**

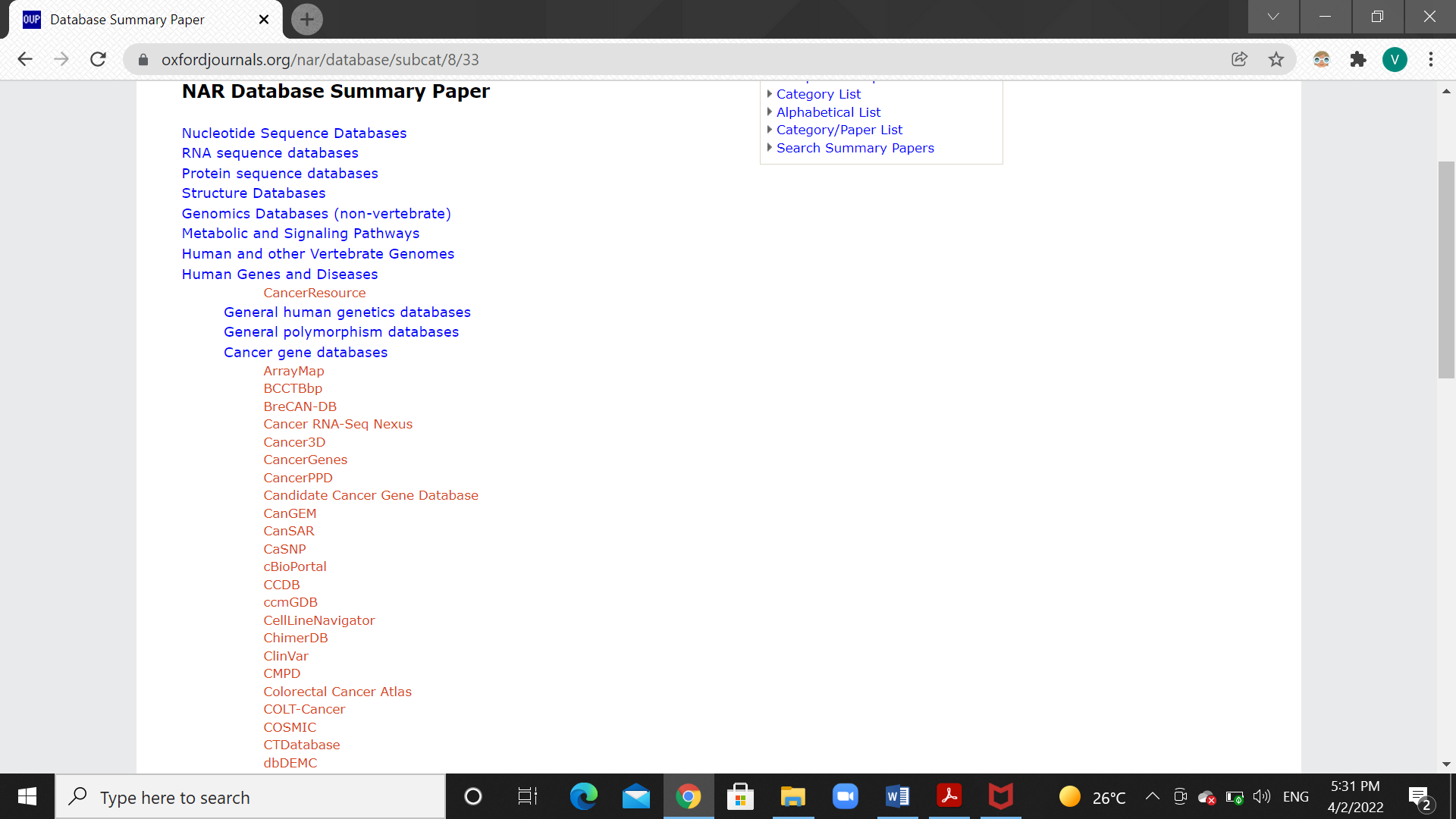
*b) Small molecules (Structure related):*

* + - *BitterDB – For bitter taste molecules and receptors.*
    - *Hemolytik – For hemolytic and non-hemolytic peptides.*

**

*c) Cancer gene databases:*

* + - *BCCTBbp - Breast Cancer Campaign Tissue Bank bioinformatics portal*
    - *CanGEM - Gene copy number changes in cancer*

**