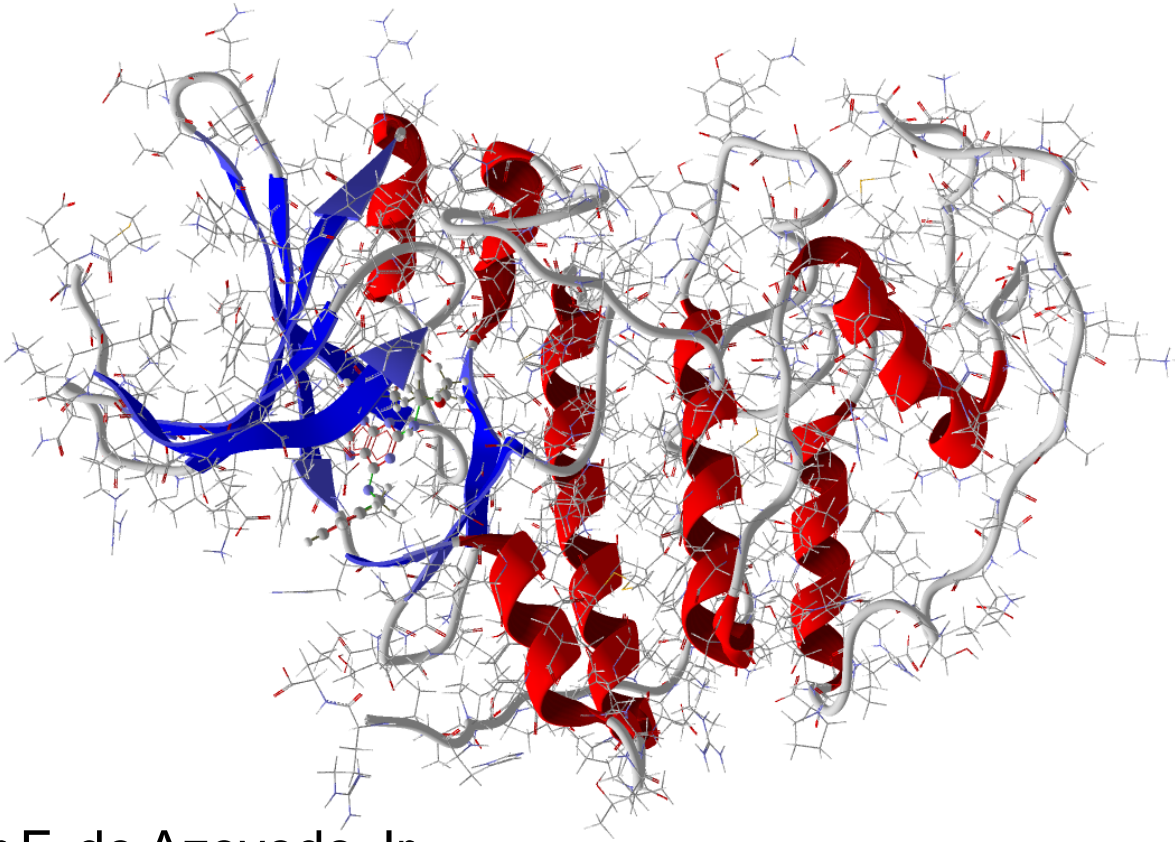
Biofísica Molecular



Prof. Dr. Walter F. de Azevedo Jr.

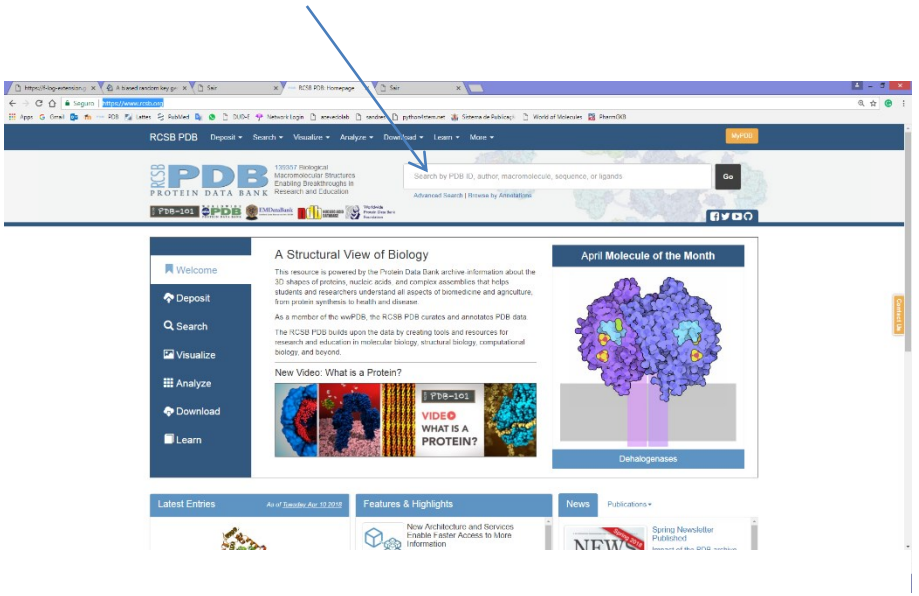
E-mail: walter@azevedolab.net

© 2018 Dr. Walter F. de Azevedo Jr.

Busca de Estruturas no *Protein Data Bank*

As proteínas que tiveram sua estrutura tridimensional determinadas estão disponíveis no site do Protein Data Bank (PDB) (https://www.rcsb.org/). O PDB pode ser pesquisado de forma similar ao PubMed, por autores ou por palavras-chave. Cada estrutura depositada recebe um código de identificação de quatro dígitos alfanuméricos que não têm necessariamente relação com o nome da proteína.

Campo para inserir a palavra para busca

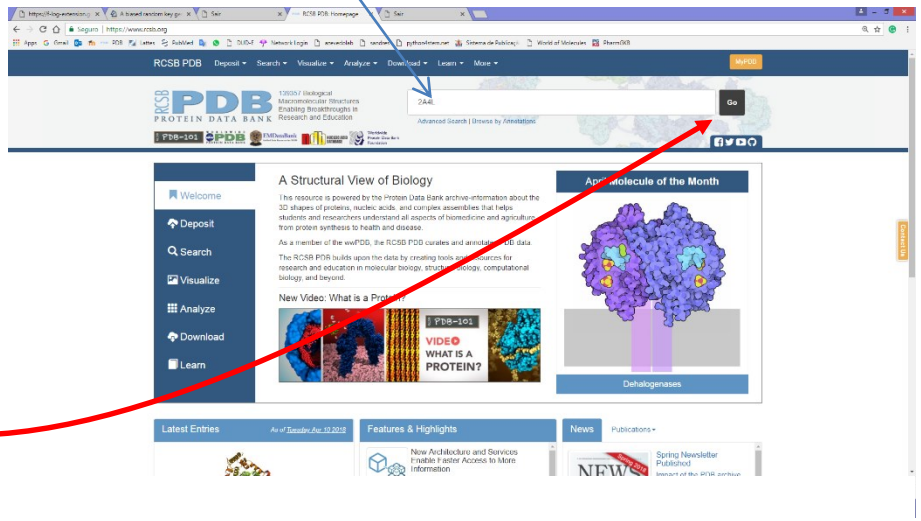


https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

Ao lado ilustramos como acessar as coordenadas com código 2A4L.
Uma vez digitado o código de acesso, normalmente disponibilizado no artigo científico que descreve a estrutura, clicamos no botão Go.

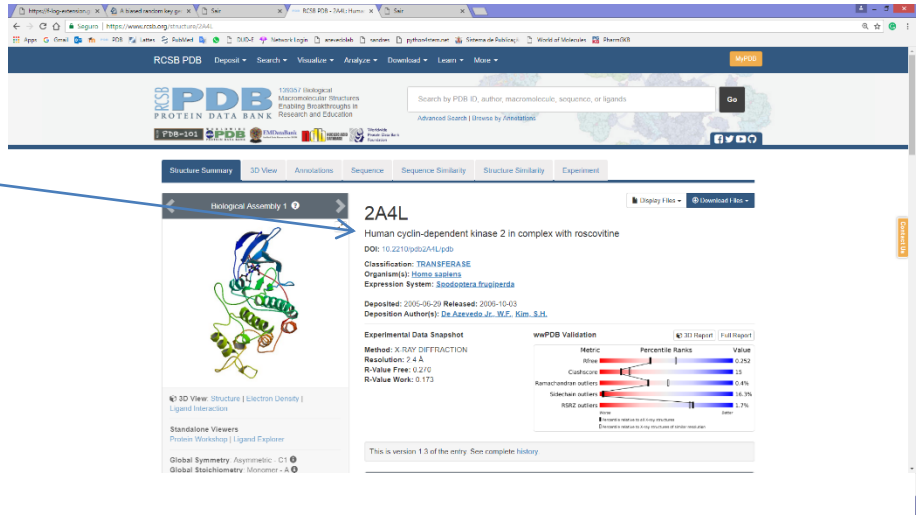
Campo para inserir a palavra para busca



https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

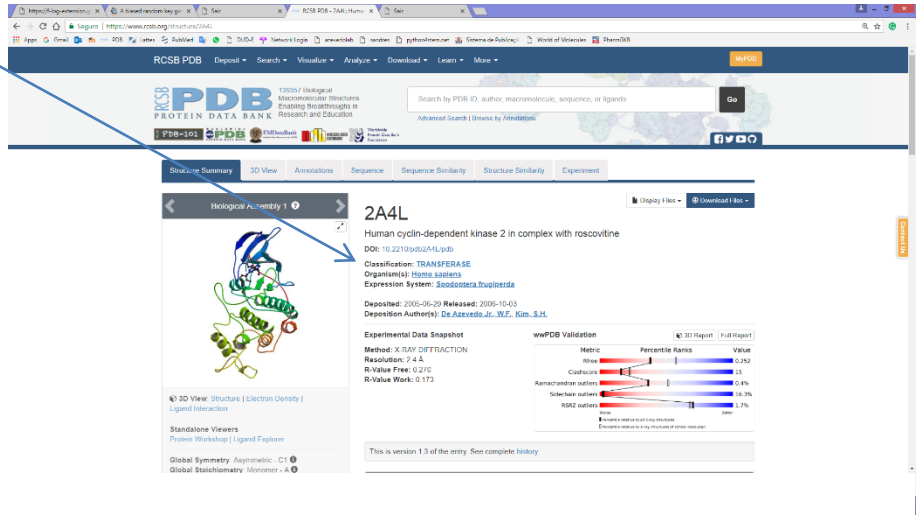
O PDB mostra informações gerais sobre a proteína que está depositada com este código.
Vemos o título da estrutura, no caso: “Human cyclin-dependent kinase 2 in complex with roscovitine”.



https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

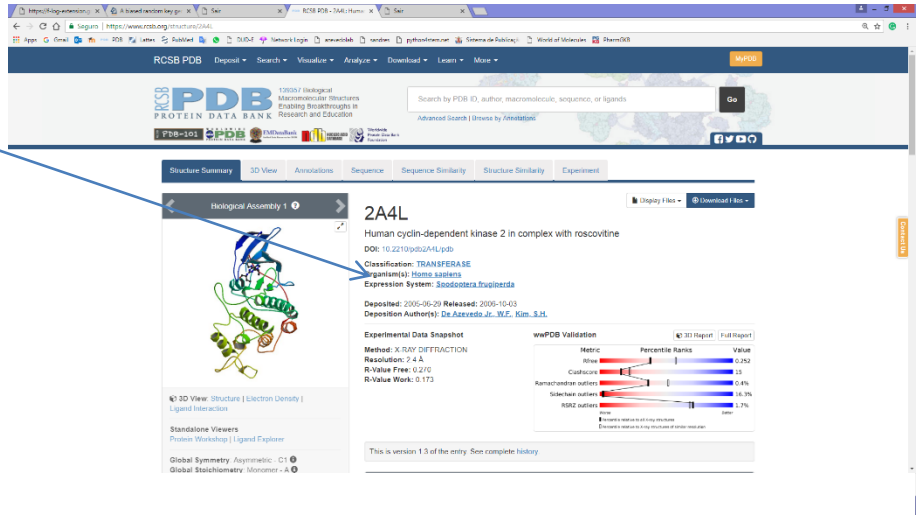
O PDB mostra, no caso de enzimas, a classificação enzimática.
Vemos que a CDK2 é uma transferase.



https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

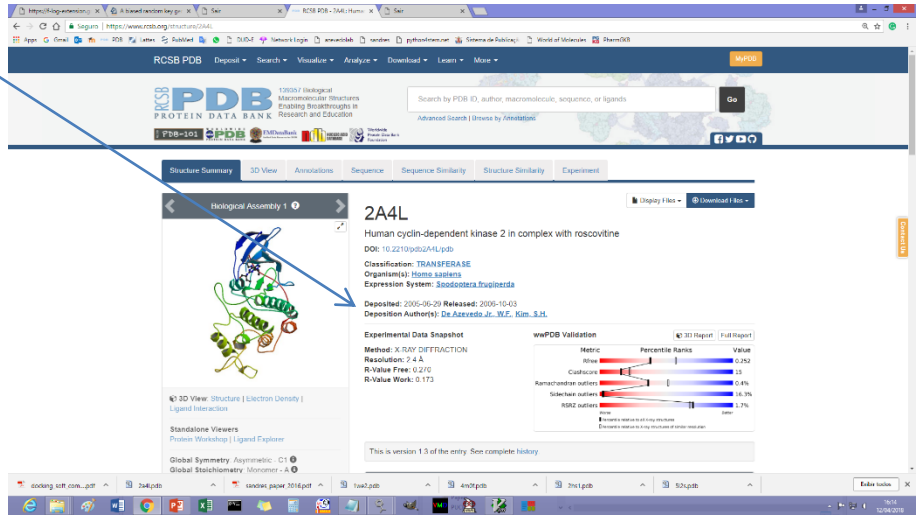
O PDB indica o organismo fonte da proteína.
No caso: *Homo sapiens*



https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

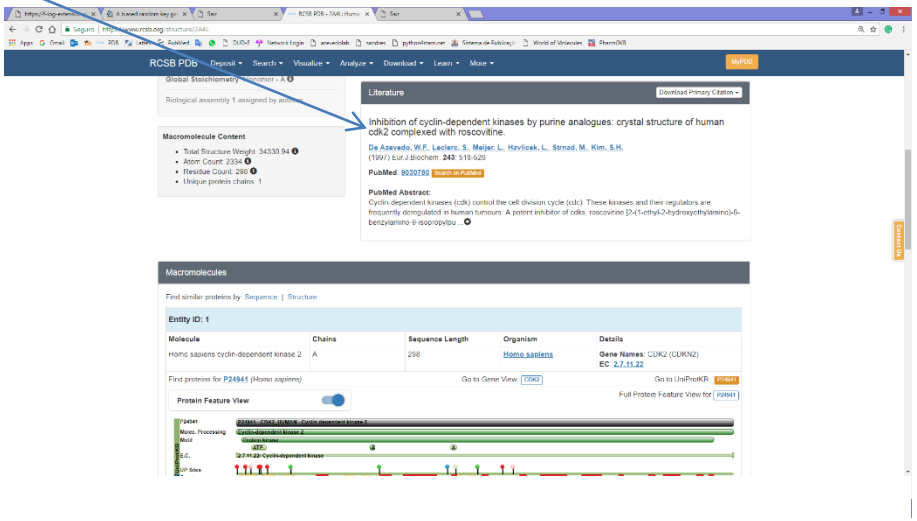
Há informações sobre os autores do depósito da estrutura, ou seja, os cientistas que determinaram a estrutura.



https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

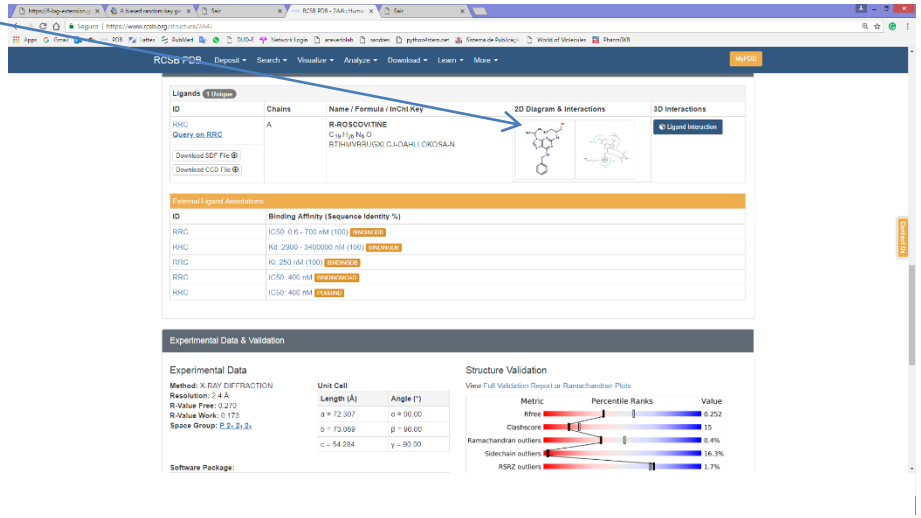
Caso a estrutura seja descrita em um artigo, este é citado no espaço indicado.



https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

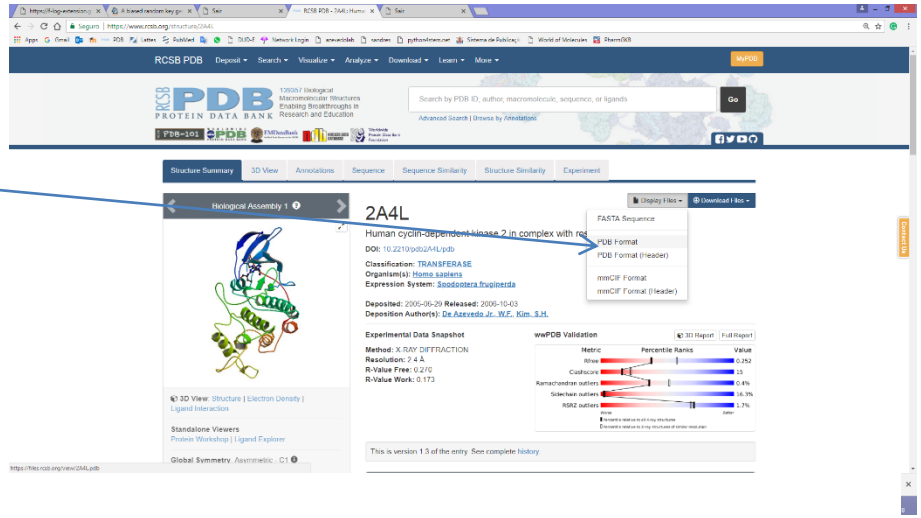
Se rolarmos a página um pouco mais, temos acesso às informações sobre o ligante, caso exista um na estrutura.
Ao lado vemos a estrutura do ligante roscovitine, complexado na estrutura.



https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

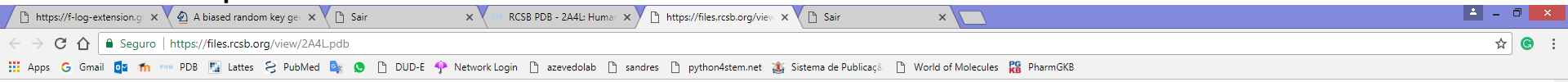
Se retornarmos ao topo da página, podemos ter acesso ao arquivo de coordenadas atômicas da estrutura.
Para visualizar o arquivo com coordenadas clicamos em:
Display Files->PDB Format
Como mostrado ao lado.



https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

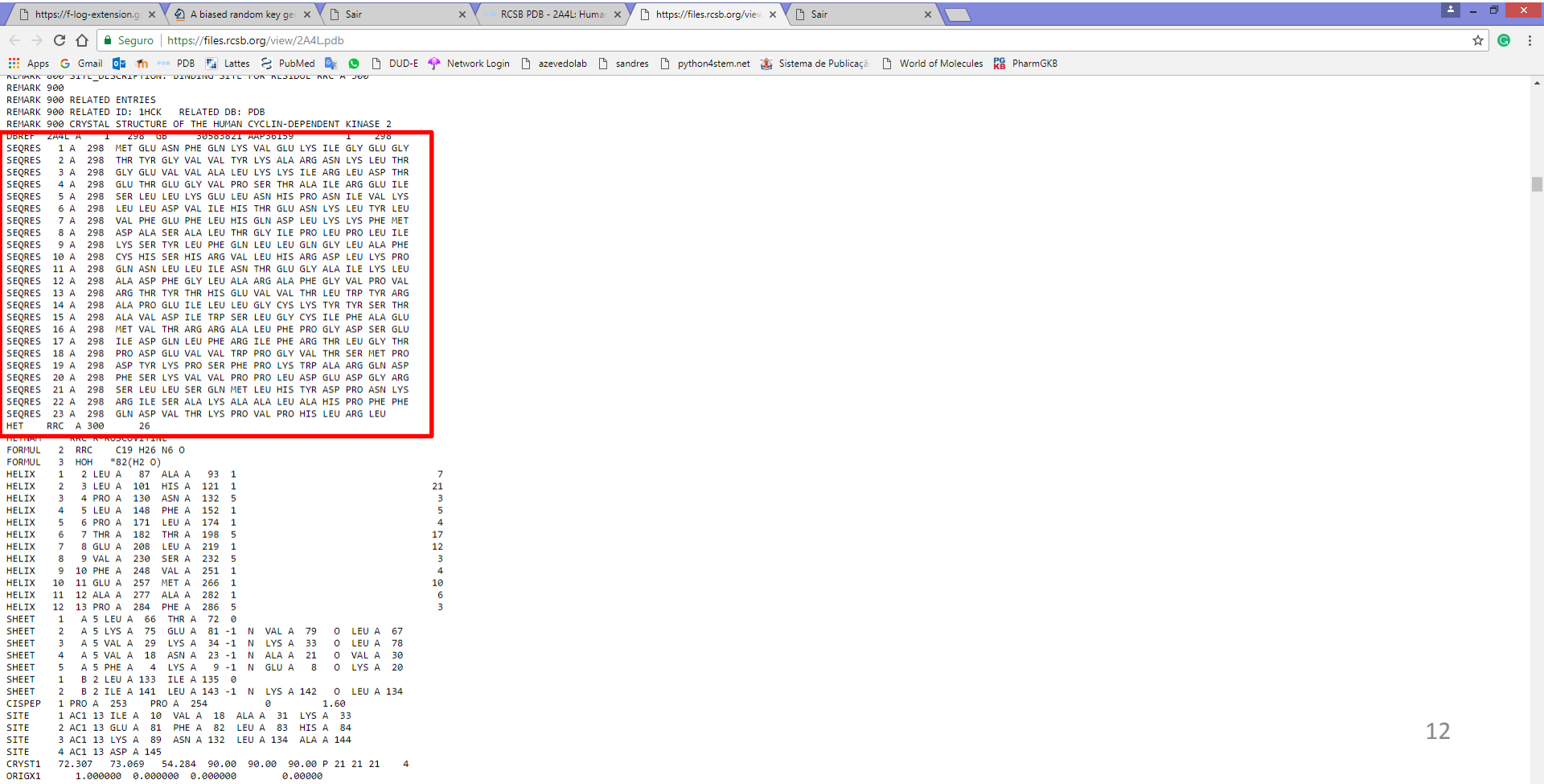
Agora temos acesso ao arquivo PDB, chamado de 2a4l.pdb. É um arquivo texto simples, que você poderia abrir com o wordpad do seu computador. As primeiras linhas trazem informações sobre o nome da proteína, os autores, o artigo e detalhes técnicos de como a proteína foi cristalizada e determinada.



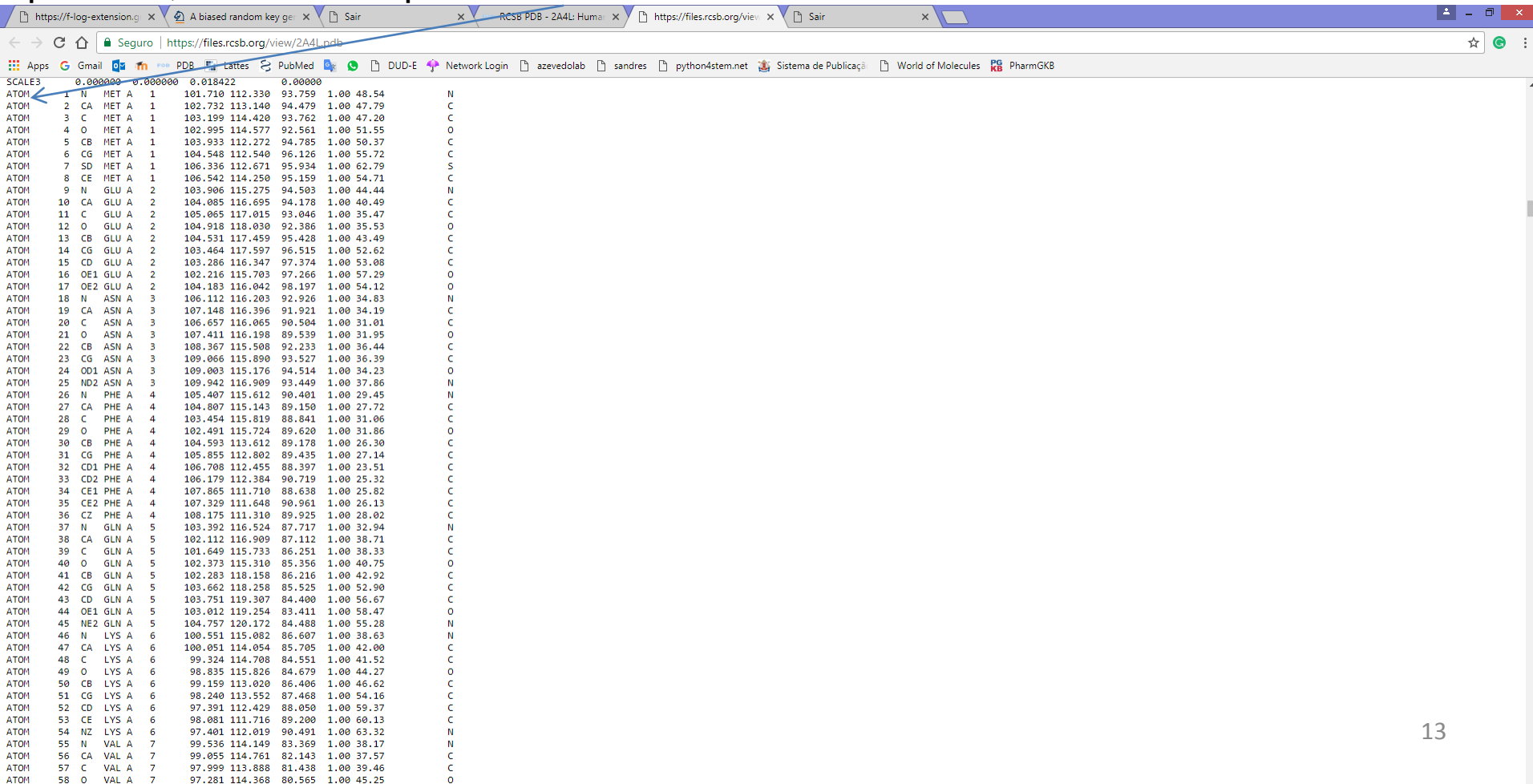

```
HEADER    TRANSFERASE                             29-JUN-05   2A4L
TITLE     HUMAN CYCLIN-DEPENDENT KINASE 2 IN COMPLEX WITH ROSCOVITINE
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: HOMO SAPIENS CYCLIN-DEPENDENT KINASE 2;
COMPND   3 CHAIN: A;
COMPND   4 SYNONYM: CDK2;
COMPND   5 EC: 2.7.1.37;
COMPND   6 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   3 ORGANISM_COMMON: HUMAN;
SOURCE   4 ORGANISM_TAXID: 9606;
SOURCE   5 EXPRESSION_SYSTEM: SPODOPTERA FRUGIPERDA;
SOURCE   6 EXPRESSION_SYSTEM_COMMON: FALL ARMYWORM;
SOURCE   7 EXPRESSION_SYSTEM_TAXID: 7108;
SOURCE   8 EXPRESSION_SYSTEM_CELL_LINE: SF9 CELLS;
SOURCE   9 EXPRESSION_SYSTEM_PLASMID: BACULOVIRUS
KEYWDS    TRANSFERASE, SERINE/THREONINE PROTEIN KINASE, ATP-BINDING, CELL
KEYWDS   2 CYCLE, CELL DIVISION, ROSCOVITINE, PHOSPHORYLATION
EXPDTA    X-RAY DIFFRACTION
AUTHOR    W.F.DE AZEVEDO JR.,S.H.KIM
REVDAT   3   11-OCT-17 2A4L    1       REMARK
REVDAT   2   24-FEB-09 2A4L    1       VERSN
REVDAT   1   03-OCT-06 2A4L    0
JRNL        AUTH   W.F.DE AZEVEDO,S.LECLERC,L.MEIJER,L.HAVLICEK,M.STRNAD,
JRNL        AUTH 2 S.H.KIM
JRNL        TITL   INHIBITION OF CYCLIN-DEPENDENT KINASES BY PURINE ANALOGUES:
JRNL        TITL 2 CRYSTAL STRUCTURE OF HUMAN CDK2 COMPLEXED WITH ROSCOVITINE.
JRNL        REF    EUR.J.BIOCHEM.                V. 243   518 1997
JRNL        REFN                   ISSN 0014-2956
JRNL        PMID   9030780
JRNL        DOI    10.1111/J.1432-1033.1997.0518A.X
REMARK   1
REMARK   1 REFERENCE 1
REMARK   1  AUTH   W.F.DE AZEVEDO JR.,H.J.MUELLER-DIECKMAN,U.SCHULZE-GAHMEN,
REMARK   1  AUTH 2 P.J.WORLAND,E.SAUSVILLE,S.H.KIM
REMARK   1  TITL   STRUCTURAL BASIS FOR SPECIFICITY AND POTENCY OF A FLAVONOID
REMARK   1  TITL 2 INHIBITOR OF HUMAN CDK2, A CELL CYCLE KINASE
REMARK   1  REF    PROC.NATL.ACAD.SCI.USA        V.  93  2735 1996
REMARK   1  REFN                   ISSN 0027-8424
REMARK   1  PMID   8610110
REMARK   1  DOI    10.1073/PNAS.93.7.2735
REMARK   1 REFERENCE 2
REMARK   1  AUTH   H.L.DE BONDT,J.ROSENBLATT,J.JANCARIK,H.D.JONES,D.O.MORGAN,
REMARK   1  AUTH 2 S.H.KIM
REMARK   1  TITL   CRYSTAL STRUCTURE OF CYCLIN-DEPENDENT KINASE 2
REMARK   1  REF    NATURE                        V. 363   595 1993
REMARK   1  REFN                   ISSN 0028-0836
REMARK   1  PMID   8510751
REMARK   1  DOI    10.1038/363595A0
REMARK   2
REMARK   2 RESOLUTION.    2.40 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : X-PLOR 3.1
REMARK   3   AUTHORS     : BRUNGER
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
```

Busca de Estruturas no *Protein Data Bank*

Se rolarmos a página um pouco mais, teremos acesso à estrutura primária da proteína, ou seja, sua sequência de aminoácidos, como destacado no retângulo vermelho.

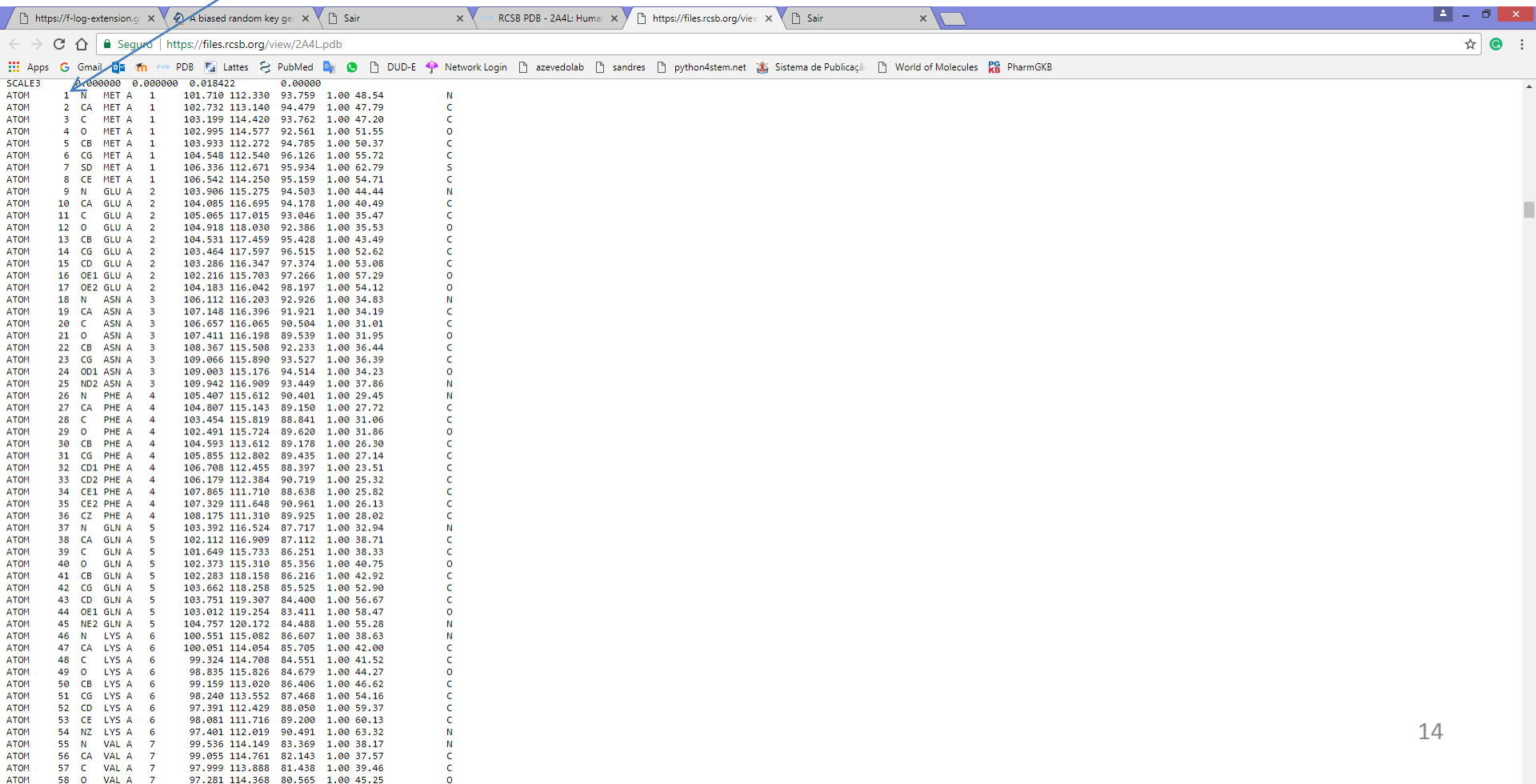


Descendo a página um pouco mais, teremos acesso às coordenadas atômicas, como mostrado abaixo. Na verdade há milhares delas, uma para cada átomo na estrutura da proteína. Cada linha que indica informações sobre as coordenadas dos átomos da proteína, inicia com a palavra “ATOM”.



Busca de Estruturas no *Protein Data Bank*

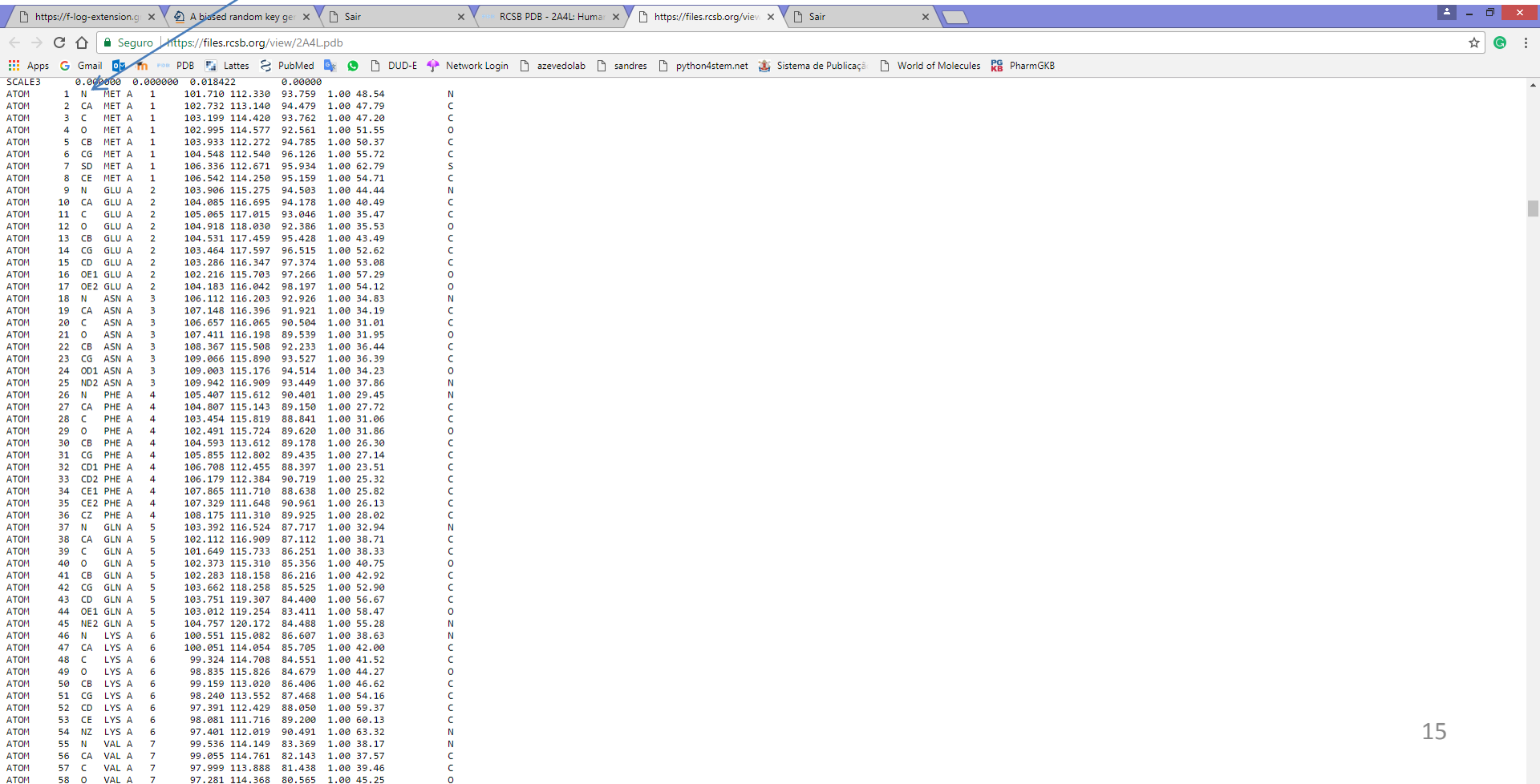
Em seguida temos o número do átomo.




```
SCALE3      0.000000  0.000000  0.018422        0.00000
ATOM      1  N   MET A   1     101.710 112.330  93.759  1.00 48.54           N
ATOM      2  CA  MET A   1     102.732 113.140  94.479  1.00 47.79           C
ATOM      3  C   MET A   1     103.199 114.420  93.762  1.00 47.20           C
ATOM      4  O   MET A   1     102.995 114.577  92.561  1.00 51.55           O
ATOM      5  CB  MET A   1     103.933 112.272  94.785  1.00 50.37           C
ATOM      6  CG  MET A   1     104.548 112.540  96.126  1.00 55.72           C
ATOM      7  SD  MET A   1     106.336 112.671  95.934  1.00 62.79           S
ATOM      8  CE  MET A   1     106.542 114.250  95.159  1.00 54.71           C
ATOM      9  N   GLU A   2     103.906 115.275  94.503  1.00 44.44           N
ATOM     10  CA  GLU A   2     104.085 116.695  94.178  1.00 40.49           C
ATOM     11  C   GLU A   2     105.065 117.015  93.046  1.00 35.47           C
ATOM     12  O   GLU A   2     104.918 118.030  92.386  1.00 35.53           O
ATOM     13  CB  GLU A   2     104.531 117.459  95.428  1.00 43.49           C
ATOM     14  CG  GLU A   2     103.464 117.597  96.515  1.00 52.62           C
ATOM     15  CD  GLU A   2     103.286 116.347  97.374  1.00 53.08           C
ATOM     16  OE1 GLU A   2     102.216 115.703  97.266  1.00 57.29           O
ATOM     17  OE2 GLU A   2     104.183 116.042  98.197  1.00 54.12           O
ATOM     18  N   ASN A   3     106.112 116.203  92.926  1.00 34.83           N
ATOM     19  CA  ASN A   3     107.148 116.396  91.921  1.00 34.19           C
ATOM     20  C   ASN A   3     106.657 116.065  90.504  1.00 31.01           C
ATOM     21  O   ASN A   3     107.411 116.198  89.539  1.00 31.95           O
ATOM     22  CB  ASN A   3     108.367 115.508  92.233  1.00 36.44           C
ATOM     23  CG  ASN A   3     109.066 115.890  93.527  1.00 36.39           C
ATOM     24  OD1 ASN A   3     109.003 115.176  94.514  1.00 34.23           O
ATOM     25  ND2 ASN A   3     109.942 116.909  93.449  1.00 37.86           N
ATOM     26  N   PHE A   4     105.407 115.612  90.401  1.00 29.45           N
ATOM     27  CA  PHE A   4     104.807 115.143  89.150  1.00 27.72           C
ATOM     28  C   PHE A   4     103.454 115.819  88.841  1.00 31.06           C
ATOM     29  O   PHE A   4     102.491 115.724  89.620  1.00 31.86           O
ATOM     30  CB  PHE A   4     104.593 113.612  89.178  1.00 26.30           C
ATOM     31  CG  PHE A   4     105.855 112.802  89.435  1.00 27.14           C
ATOM     32  CD1 PHE A   4     106.708 112.455  88.397  1.00 23.51           C
ATOM     33  CD2 PHE A   4     106.179 112.384  90.719  1.00 25.32           C
ATOM     34  CE1 PHE A   4     107.865 111.710  88.638  1.00 25.82           C
ATOM     35  CE2 PHE A   4     107.329 111.648  90.961  1.00 26.13           C
ATOM     36  CZ  PHE A   4     108.175 111.310  89.925  1.00 28.02           C
ATOM     37  N   GLN A   5     103.392 116.524  87.717  1.00 32.94           N
ATOM     38  CA  GLN A   5     102.112 116.909  87.112  1.00 38.71           C
ATOM     39  C   GLN A   5     101.649 115.733  86.251  1.00 38.33           C
ATOM     40  O   GLN A   5     102.373 115.310  85.356  1.00 40.75           O
ATOM     41  CB  GLN A   5     102.283 118.158  86.216  1.00 42.92           C
ATOM     42  CG  GLN A   5     103.662 118.258  85.525  1.00 52.90           C
ATOM     43  CD  GLN A   5     103.751 119.307  84.400  1.00 56.67           C
ATOM     44  OE1 GLN A   5     103.012 119.254  83.411  1.00 58.47           O
ATOM     45  NE2 GLN A   5     104.757 120.172  84.488  1.00 55.28           N
ATOM     46  N   LYS A   6     100.551 115.082  86.607  1.00 38.63           N
ATOM     47  CA  LYS A   6     100.051 114.054  85.705  1.00 42.00           C
ATOM     48  C   LYS A   6      99.324 114.708  84.551  1.00 41.52           C
ATOM     49  O   LYS A   6      98.835 115.826  84.679  1.00 44.27           O
ATOM     50  CB  LYS A   6      99.159 113.020  86.406  1.00 46.62           C
ATOM     51  CG  LYS A   6      98.240 113.552  87.468  1.00 54.16           C
ATOM     52  CD  LYS A   6      97.391 112.429  88.050  1.00 59.37           C
ATOM     53  CE  LYS A   6      98.081 111.716  89.200  1.00 60.13           C
ATOM     54  NZ  LYS A   6      97.401 112.019  90.491  1.00 63.32           N
ATOM     55  N   VAL A   7      99.536 114.149  83.369  1.00 38.17           N
ATOM     56  CA  VAL A   7      99.055 114.761  82.143  1.00 37.57           C
ATOM     57  C   VAL A   7      97.999 113.888  81.438  1.00 39.46           C
ATOM     58  O   VAL A   7      97.281 114.368  80.565  1.00 45.25           O
```

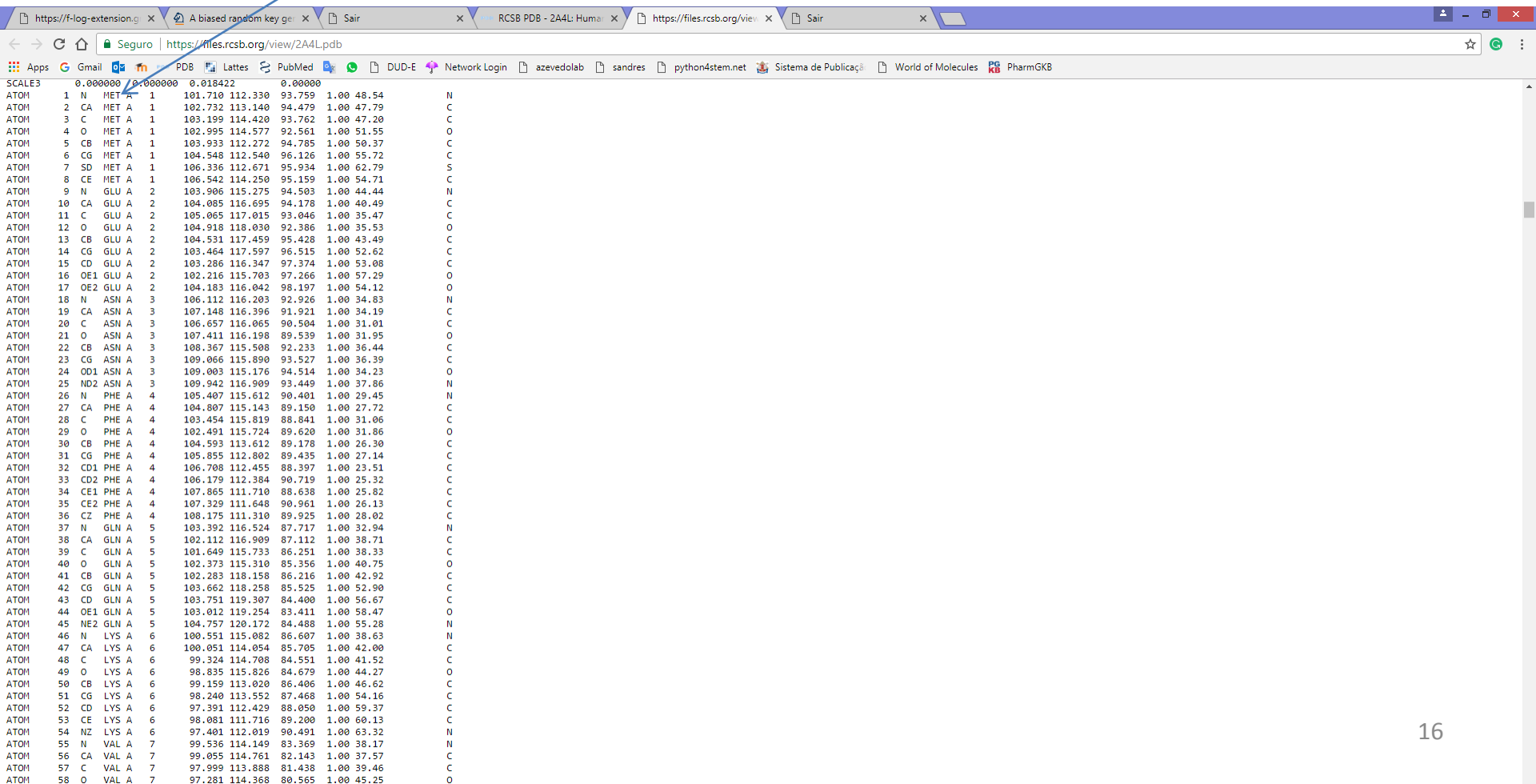
Busca de Estruturas no *Protein Data Bank*

Depois o identificador do átomo.

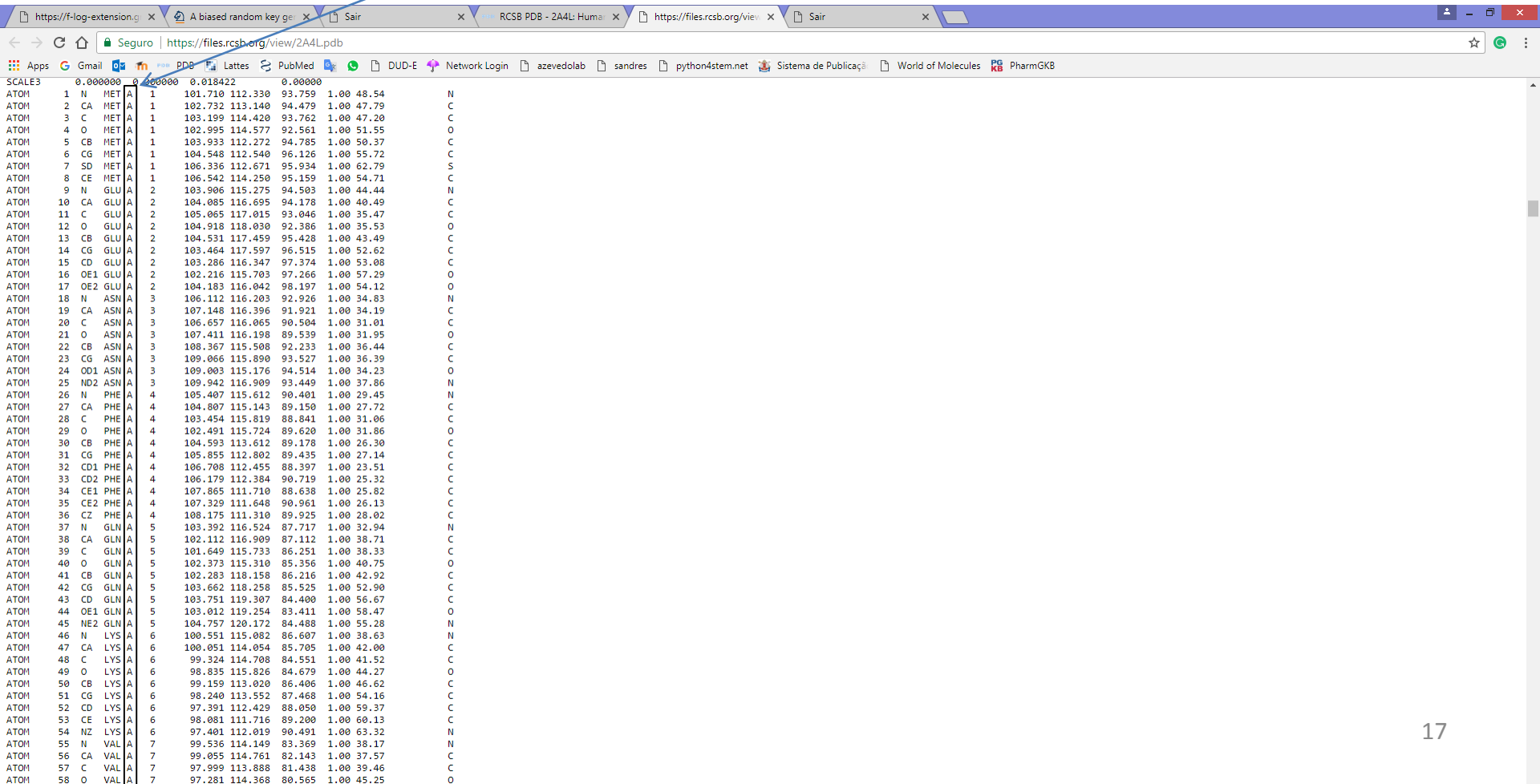


Busca de Estruturas no *Protein Data Bank*

Em seguida o código de 3 letras do aminoácido onde o átomo está.

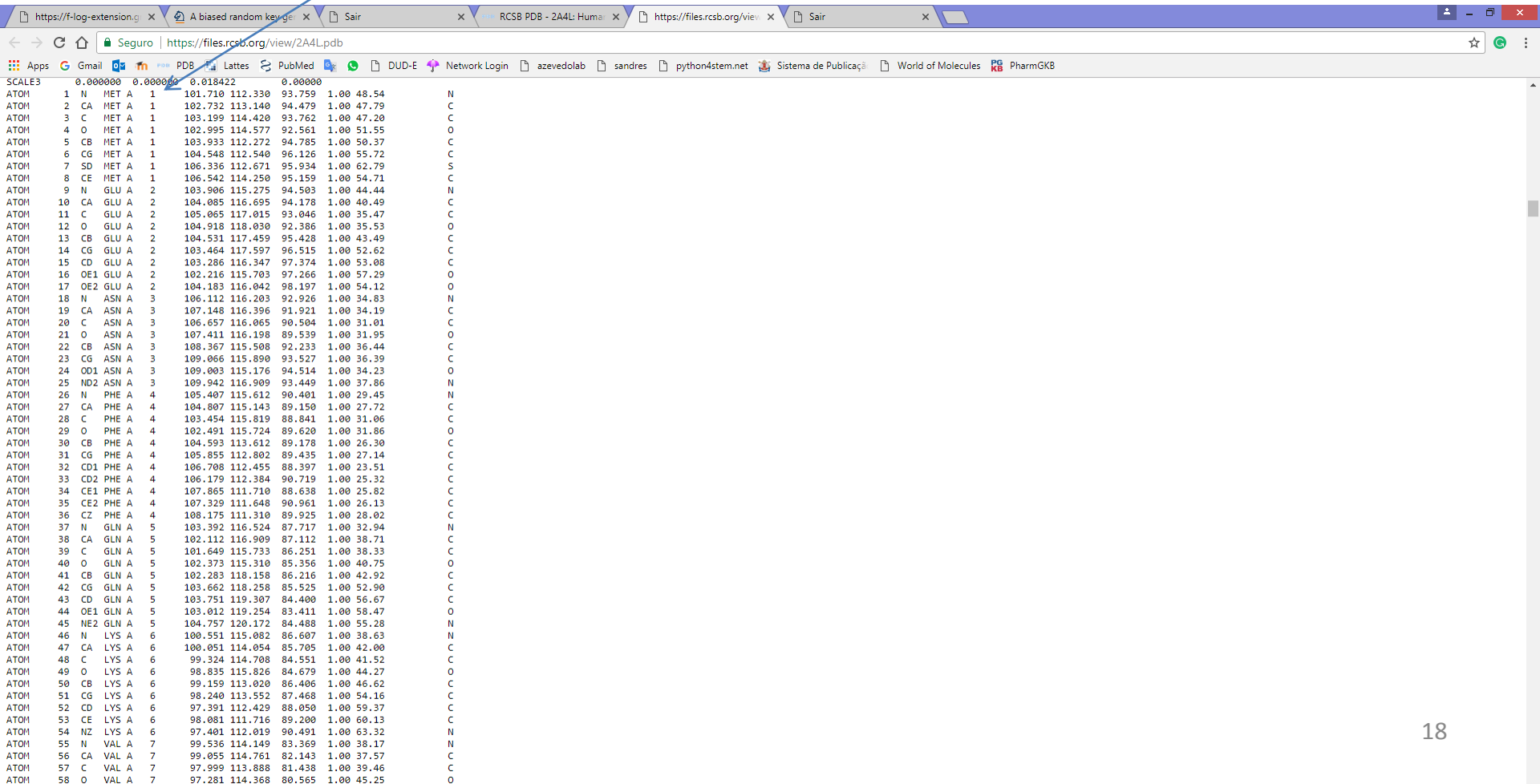


A letra em destaque indica a cadeia polipeptídica, no caso de termos mais de uma, usamos letras diferentes.



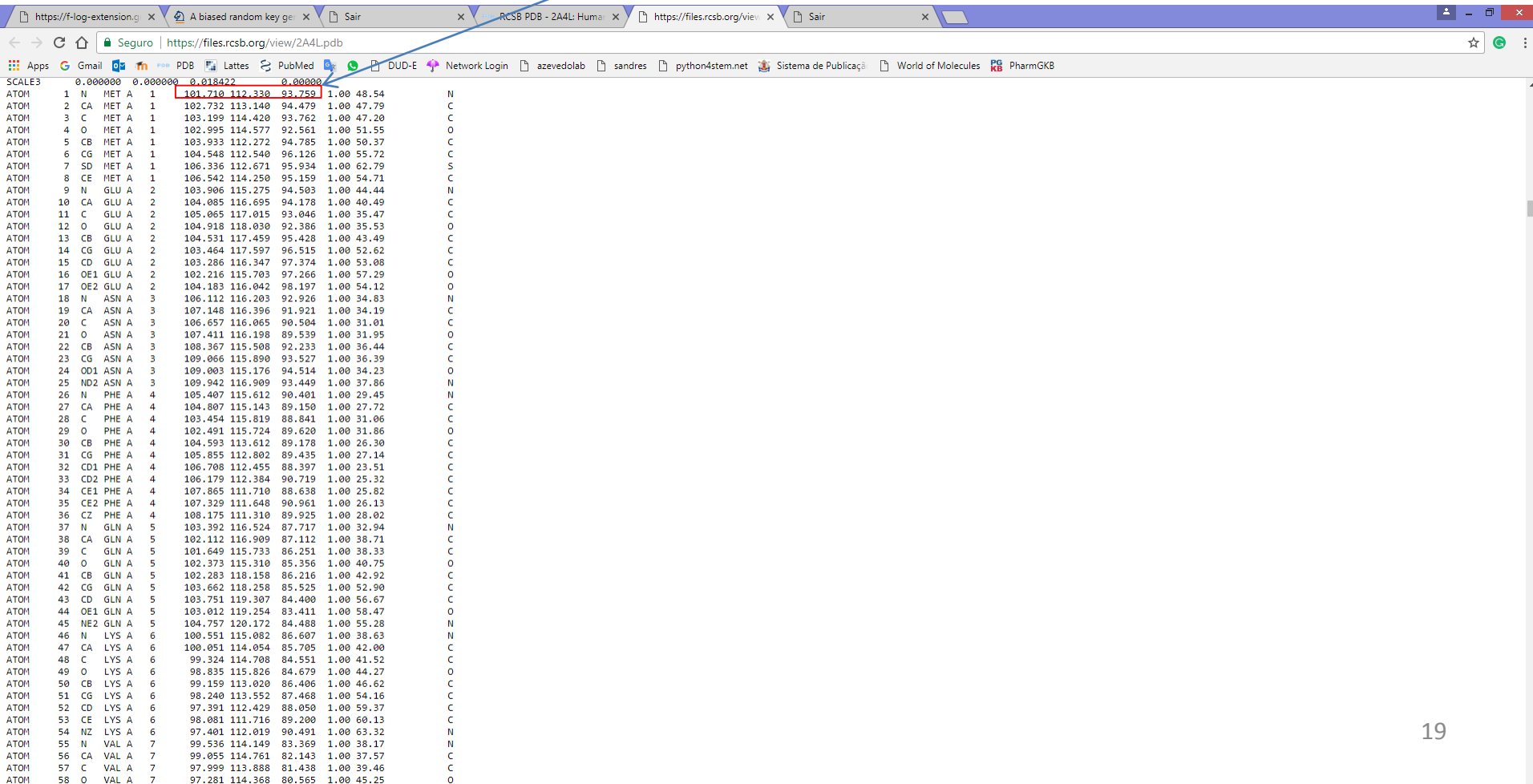
Busca de Estruturas no *Protein Data Bank*

Depois temos o número do aminoácido. Veja que temos mais de um átomo para cada aminoácido.



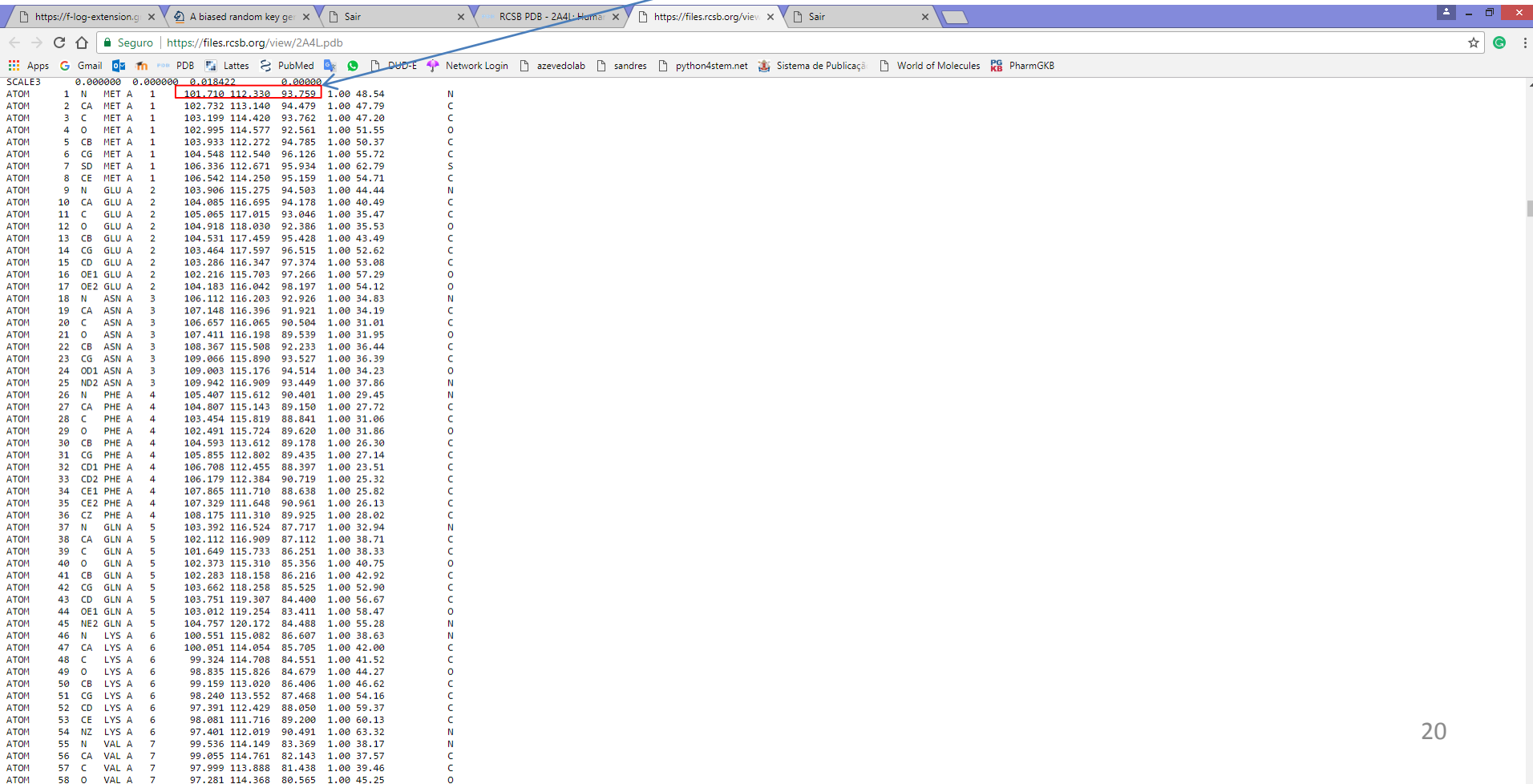
Busca de Estruturas no *Protein Data Bank*

Os três números seguintes indicam as coordenadas x,y,z do átomo, em Ângstrom (1 Å = 10^{-10} m).



Busca de Estruturas no *Protein Data Bank*

Os programas que desenham estruturas de proteína usam esta informação para representar os átomos na tela gráfica.



Se retornarmos ao topo da página, podemos fazer download do arquivo de coordenadas atômicas da estrutura.
Para baixar o arquivo com coordenadas clicamos em:
Download Files->PDB Format
Como mostrado ao lado.
Depois escolhemos a pasta onde queremos deixar o arquivo.
O arquivo baixado chama-se 2a4l.pdb, os quatro primeiros dígitos indicam o código da proteína, a extensão, .pdb, indica que é um arquivo no formato PDB.
Podemos visualizar este arquivo com o programa VMD.

https://www.rcsb.org/
Acesso em: 12 de abril de 2018.

Busca de Estruturas no *Protein Data Bank*

Nos próximos slides, trazemos as páginas de entrada do PDB para cada uma das proteínas escolhidas pelos grupos. Identifique a proteína escolhida pelo grupo e anote as seguintes informações:

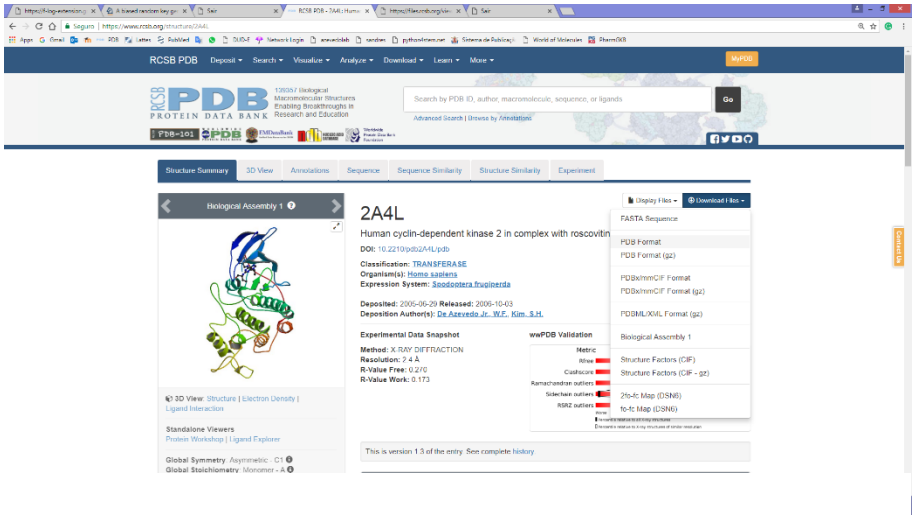
1) Código PDB da estrutura

2) Classificação enzimática

3) Organismo

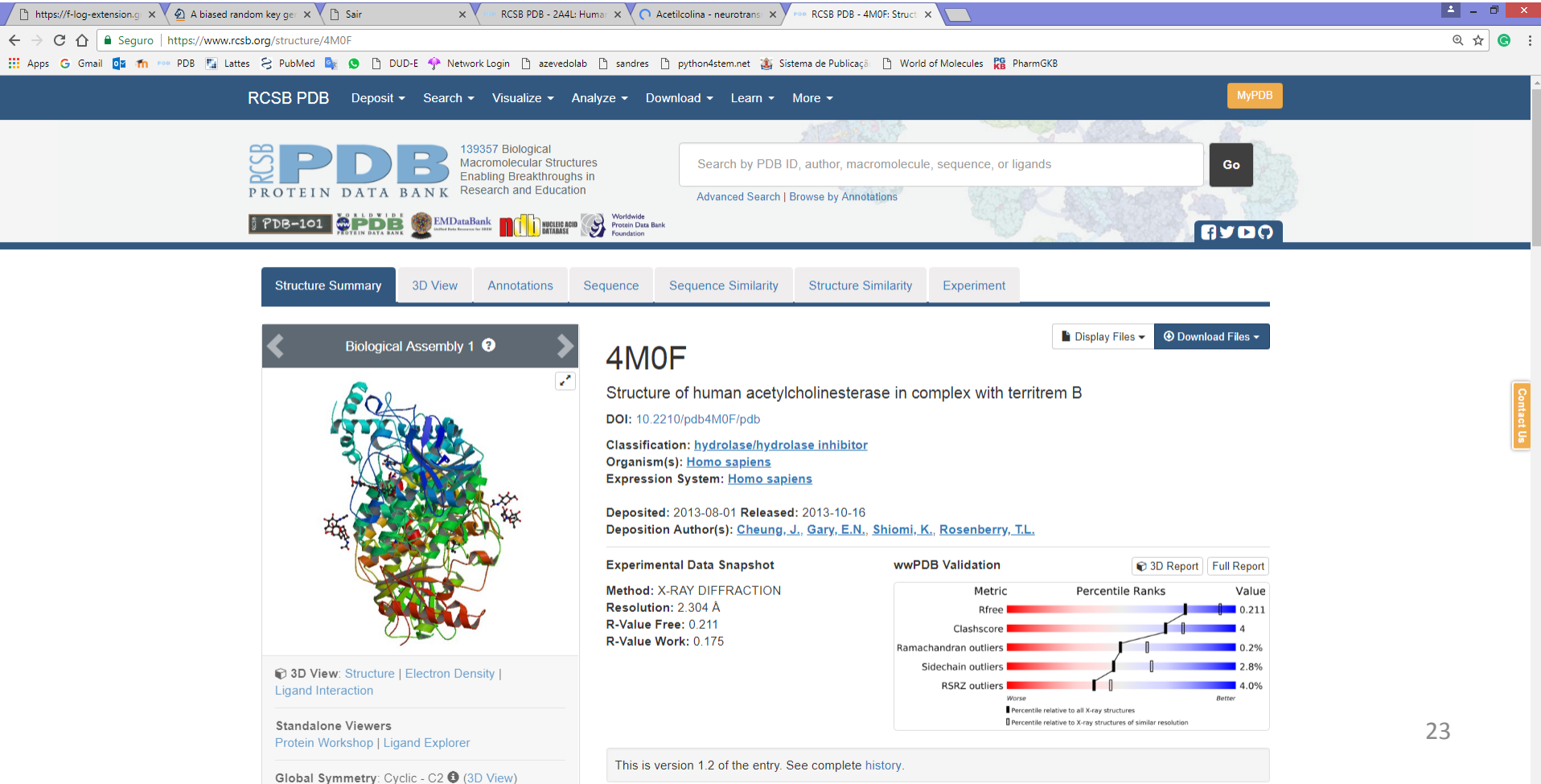
4) Autores

5) Referência do artigo principal que descreve a estrutura



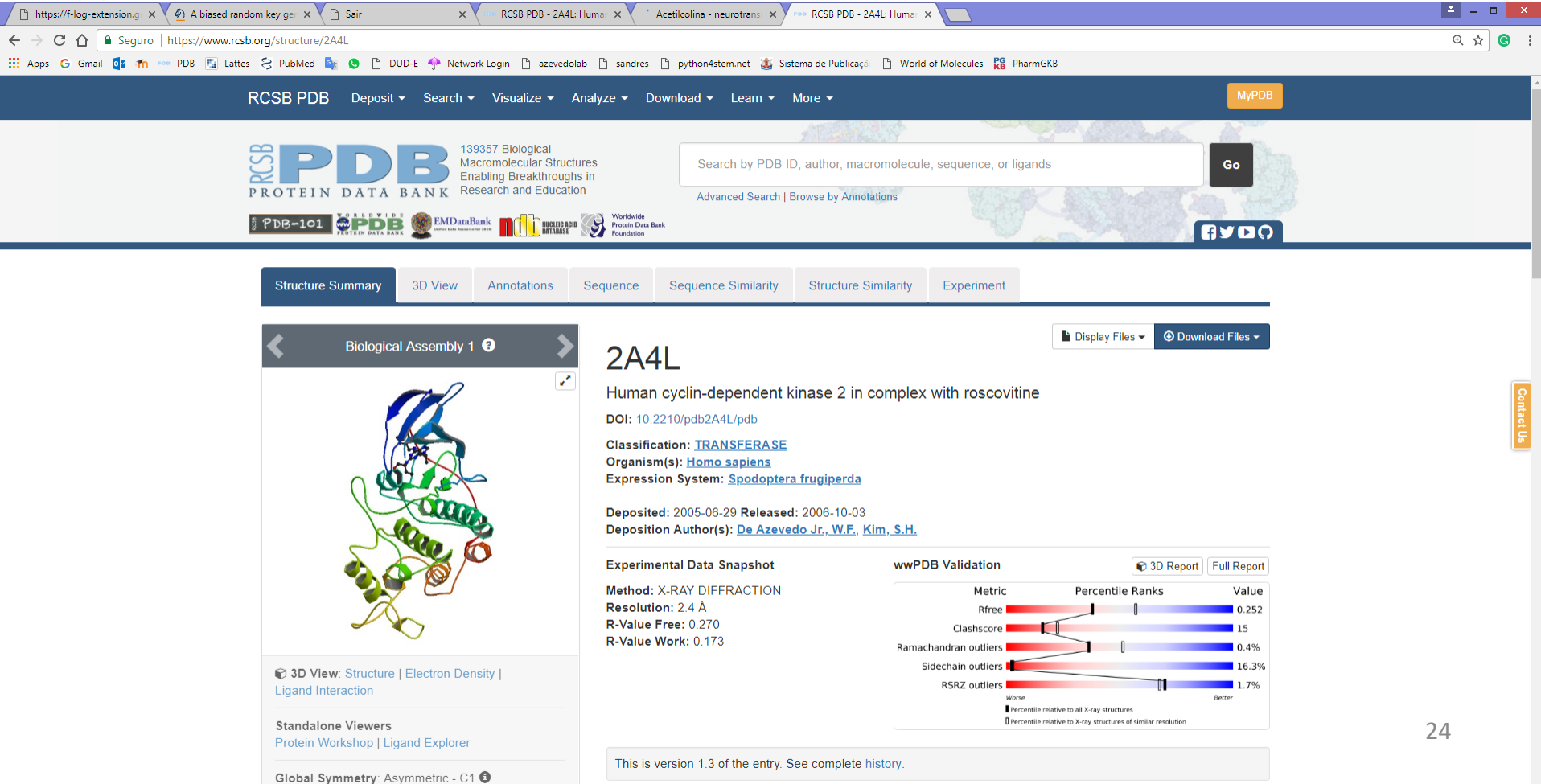
PDB da Acetilcolinesterase (AChE)

Usaremos o PDB de código 4M0F para a estrutura da Acetilcolinesterase (AChE), consulte o site do PDB (https://www.rcsb.org/) para obter as informações solicitadas.



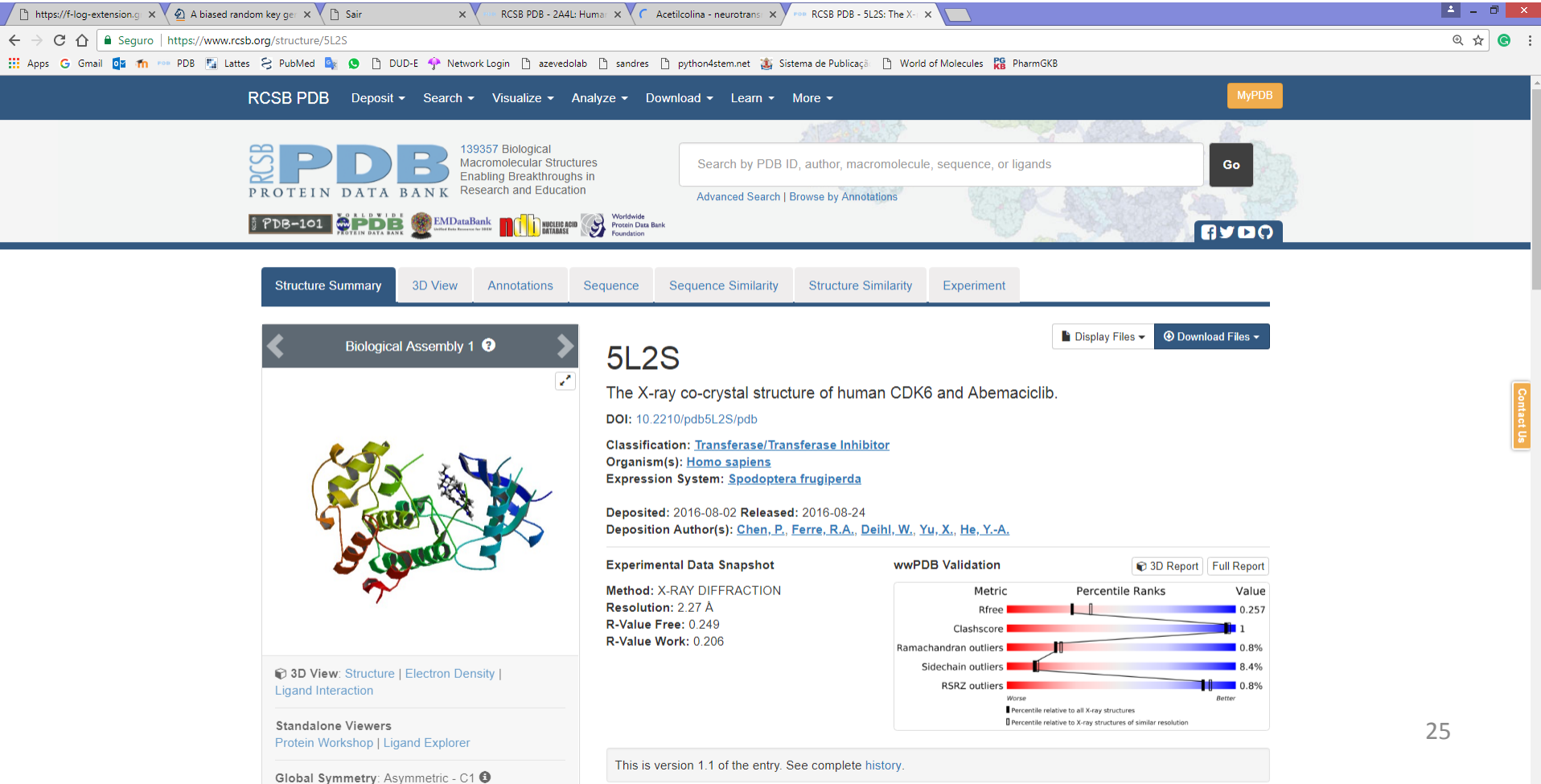
PDB da Quinase Dependente de Ciclina 2 (CDK2)

Usaremos o PDB de código 2A4L para a estrutura da quinase dependente de ciclina 2 *(cyclin-dependente kinase)(*CDK2), consulte o site do PDB (https://www.rcsb.org/) para obter as informações solicitadas.



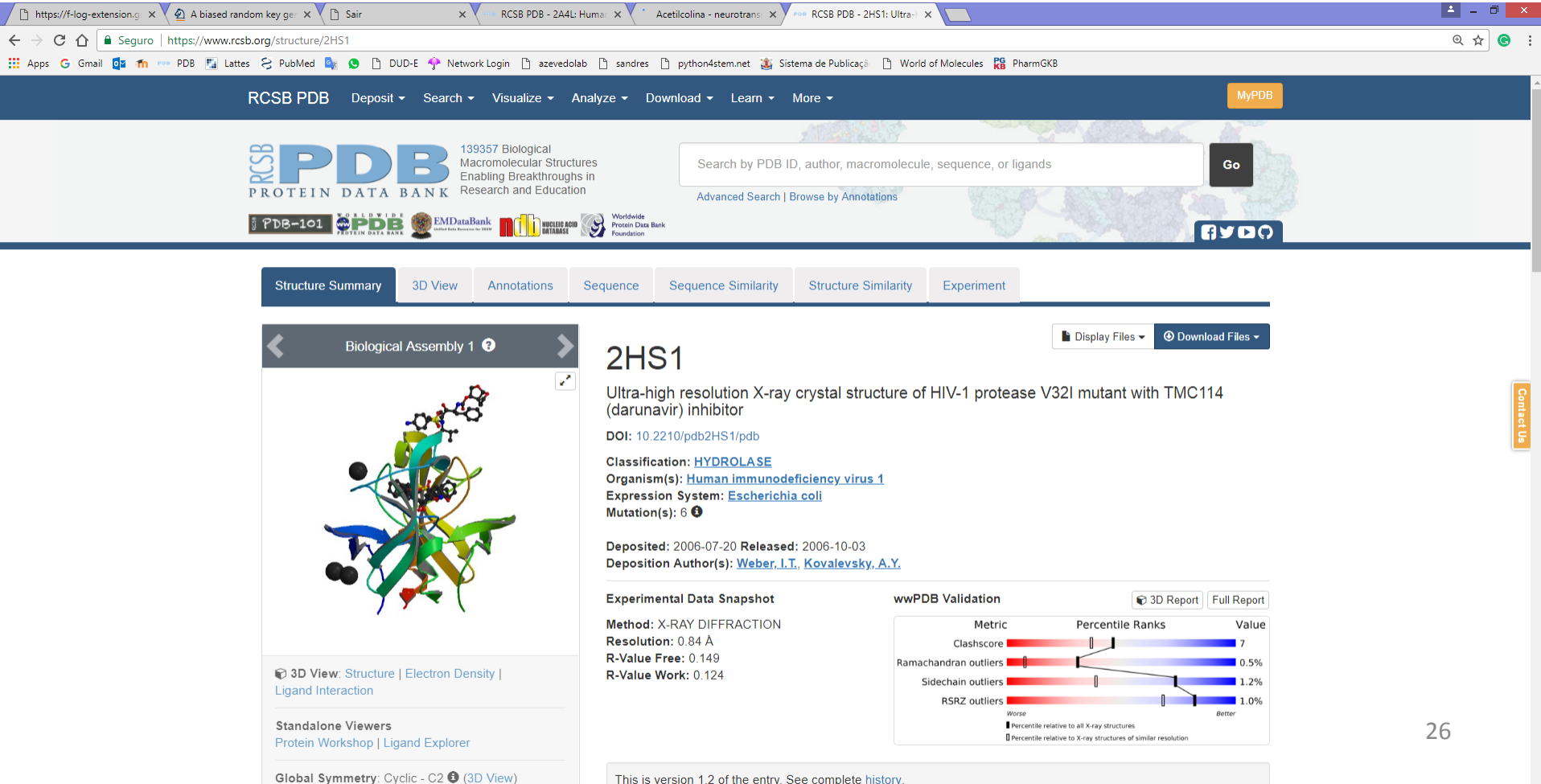
PDB da Quinase Dependente de Ciclina 6 (CDK6)

Usaremos o PDB de código 5L2S para a estrutura da quinase dependente de ciclina 6 *(cyclin-dependente kinase)(*CDK6), consulte o site do PDB (https://www.rcsb.org/) para obter as informações solicitadas.



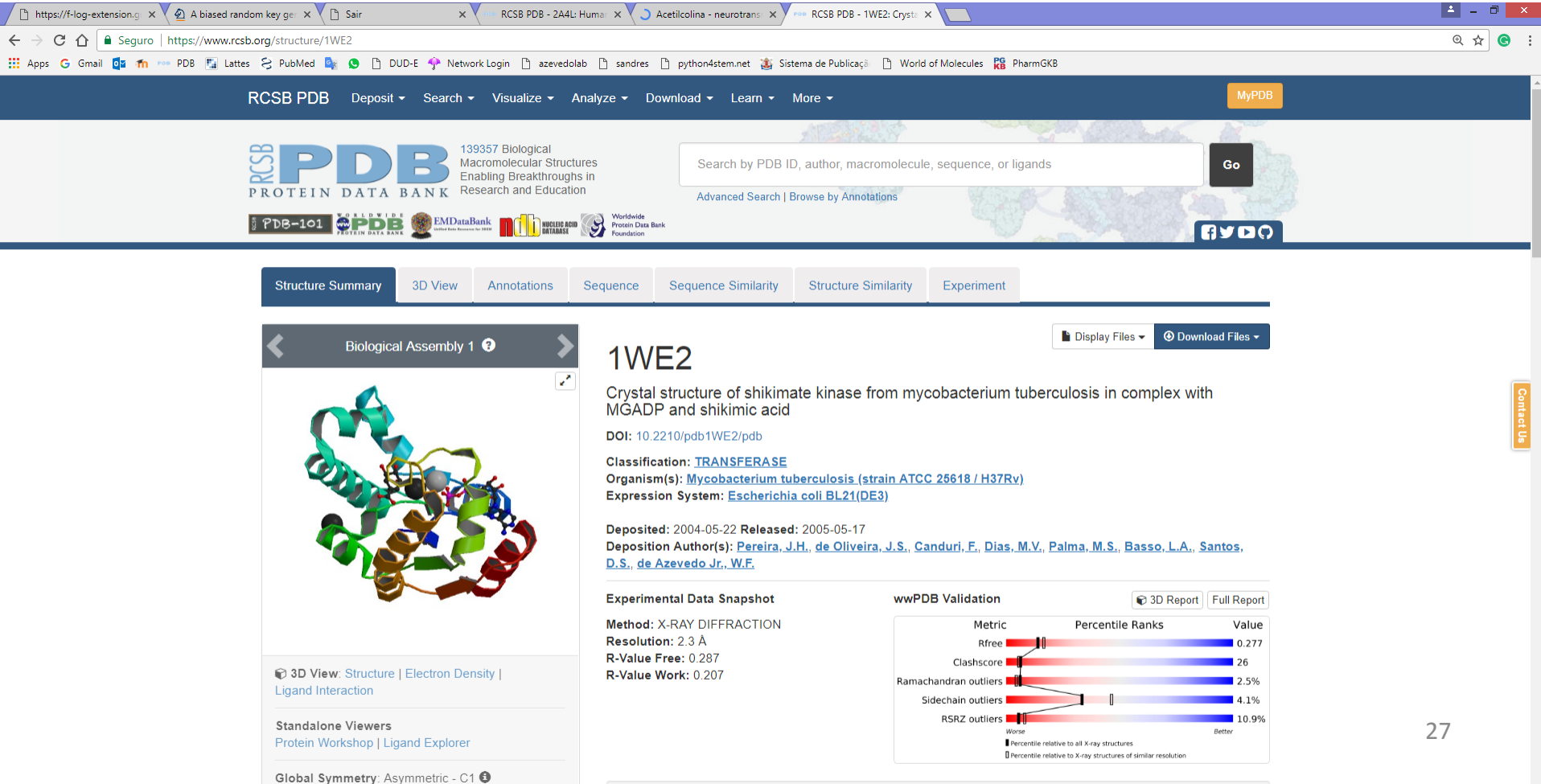
PDB da Protease do HIV-1

Usaremos o PDB de código 2HS1 para a estrutura da protease do HIV-1, consulte o site do PDB (https://www.rcsb.org/) para obter as informações solicitadas.



PDB da Chiquimato Quinase

Usaremos o PDB de código 1WE2 para a estrutura da chiquimato quinase, consulte o site do PDB (https://www.rcsb.org/) para obter as informações solicitadas.



Site do Protein Data Bank:
https://www.rcsb.org/
Acesso em 12 de Abril de 2018.