A WEB TOOL TO COMPUTE THE DIFFRACTION PRECISION INDEX OF PROTEIN STRUCTURES

Cristian NEANU*, Sorin AVRAM

Department of Computational Chemistry, "Coriolan Drăgulescu" Institute of Chemistry Timisoara,

24 Mihai Viteazul Av., 300223 Timisoara, Romania

**** e-mail:** cristi@acad-icht.tm.edu.ro*

The Protein Data Bank (PDB) [1] is one of the main open access digital data resource in biology and medicine. Data collected in the PDB archive on biological macromolecules contain general information about the structures stored in the archive, as well as information specific to the types of methods of determining the structure (X-rays or NMR) [2].

The diffraction-component precision index (DPI) estimates the average accuracy of atomic coordinates obtained by the structural refinement of protein diffraction data, with application in crystallography and computational chemistry.

The heterogeneous nature of PDB structure files makes automated DPI computation difficult, limiting its use.

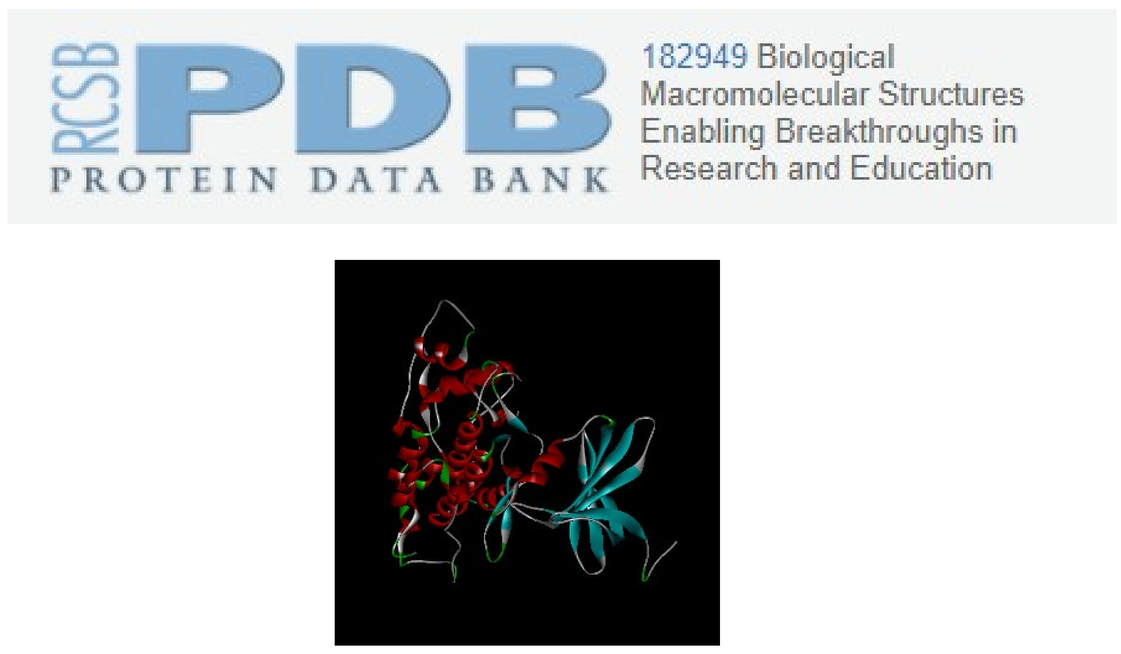

Figure 1. **1J1B,** protein sample structure

```
HEADER    OXYGEN TRANSPORT                        24-DEC-97   112M
TITLE     SPERM WHALE MYOGLOBIN D122N N-PROPYL ISOCYANIDE AT PH 9.0
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: MYOGLOBIN;
COMPND   3 CHAIN: A;
COMPND   4 ENGINEERED: YES;
COMPND   5 MUTATION: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: PHYSETER CATODON;
SOURCE   3 ORGANISM_COMMON: SPERM WHALE;
SOURCE   4 ORGANISM_TAXID: 9755;
SOURCE   5 ORGAN: SKELETAL;
SOURCE   6 TISSUE: SKELETAL MUSCLE;
SOURCE   7 CELLULAR_LOCATION: CYTOPLASM;
SOURCE   8 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE   9 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE  10 EXPRESSION_SYSTEM_STRAIN: PHAGE RESISTANT TB1;
SOURCE  11 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE  12 EXPRESSION_SYSTEM_PLASMID: PEMBL 19+
KEYWDS    LIGAND BINDING, OXYGEN STORAGE, OXYGEN BINDING, HEME, OXYGEN
KEYWDS   2 TRANSPORT
EXPDTA    X-RAY DIFFRACTION
AUTHOR    R.D.SMITH,J.S.OLSON,G.N.PHILLIPS JR.
REVDAT   5   07-MAR-18 112M    1       REMARK
REVDAT   4   24-FEB-09 112M    1       VERSN
REVDAT   3   03-MAY-05 112M    1       AUTHOR
REVDAT   2   17-MAY-99 112M    1       JRNL   HELIX
REVDAT   1   08-APR-98 112M    0
JRNL        AUTH   R.D.SMITH
JRNL        TITL   CORRELATIONS BETWEEN BOUND N-ALKYL ISOCYANIDE ORIENTATIONS
JRNL        TITL 2 AND PATHWAYS FOR LIGAND BINDING IN RECOMBINANT MYOGLOBINS
JRNL        REF    THESIS, RICE                               1999
JRNL        REFN
REMARK   2
REMARK   2 RESOLUTION.    2.34 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : X-PLOR 3.851
REMARK   3   AUTHORS     : BRUNGER
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.34
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 5.00
REMARK   3   DATA CUTOFF            (SIGMA(F)) : 0.000
REMARK   3   DATA CUTOFF HIGH         (ABS(F)) : 10000000.000
REMARK   3   DATA CUTOFF LOW          (ABS(F)) : 0.0010
REMARK   3   COMPLETENESS (WORKING+TEST)   (%) : 81.1
REMARK   3   NUMBER OF REFLECTIONS             : 6769
```

```
ATOM      2  CA  MET A   0      24.486   9.577  -9.952  1.00 31.81           C
ATOM      3  C   MET A   0      25.896  10.021 -10.325  1.00 30.76           C
ATOM      4  O   MET A   0      26.872   9.321 -10.045  1.00 31.02           O
ATOM      5  CB  MET A   0      24.106  10.130  -8.576  1.00 33.06           C
ATOM      6  CG  MET A   0      24.869   9.511  -7.423  1.00 33.56           C
ATOM      7  SD  MET A   0      24.239  10.050  -5.829  1.00 36.21           S
ATOM      8  CE  MET A   0      24.760  11.803  -5.852  1.00 33.39           C
ATOM      9  N   VAL A   1      25.992  11.182 -10.969  1.00 29.08           N
ATOM     10  CA  VAL A   1      27.278  11.729 -11.392  1.00 27.24           C
ATOM     11  C   VAL A   1      27.377  13.228 -11.126  1.00 26.03           C
ATOM     12  O   VAL A   1      26.480  13.992 -11.483  1.00 26.40           O
ATOM     13  CB  VAL A   1      27.518  11.496 -12.902  1.00 27.23           C
ATOM     14  CG1 VAL A   1      28.871  12.061 -13.325  1.00 26.87           C
ATOM     15  CG2 VAL A   1      27.428  10.016 -13.230  1.00 27.78           C
ATOM     16  N   LEU A   2      28.465  13.639 -10.480  1.00 23.95           N
ATOM     17  CA  LEU A   2      28.701  15.050 -10.195  1.00 21.90           C
ATOM     18  C   LEU A   2      29.400  15.698 -11.383  1.00 20.88           C
ATOM     19  O   LEU A   2      30.114  15.029 -12.133  1.00 21.39           O
ATOM     20  CB  LEU A   2      29.591  15.220  -8.961  1.00 21.24           C
ATOM     21  CG  LEU A   2      29.013  15.019  -7.563  1.00 21.14           C
ATOM     22  CD1 LEU A   2      28.781  13.544  -7.291  1.00 20.43           C
ATOM     23  CD2 LEU A   2      29.977  15.604  -6.547  1.00 20.35           C
ATOM     24  N   SER A   3      29.195  16.995 -11.563  1.00 19.21           N
ATOM     25  CA  SER A   3      29.850  17.706 -12.649  1.00 18.89           C
ATOM     26  C   SER A   3      31.234  18.132 -12.161  1.00 18.39           C
ATOM     27  O   SER A   3      31.504  18.117 -10.959  1.00 18.30           O
ATOM     28  CB  SER A   3      29.032  18.937 -13.059  1.00 18.59           C
ATOM     29  OG  SER A   3      28.957  19.891 -12.010  1.00 18.04           O
ATOM     30  N   GLU A   4      32.113  18.482 -13.093  1.00 16.27           N
ATOM     31  CA  GLU A   4      33.463  18.929 -12.757  1.00  7.99           C
ATOM     32  C   GLU A   4      33.394  20.157 -11.851  1.00 15.05           C
ATOM     33  O   GLU A   4      34.167  20.273 -10.901  1.00 24.03           O
ATOM     34  CB  GLU A   4      34.245  19.272 -14.032  1.00 21.44           C
ATOM     35  CG  GLU A   4      35.649  19.844 -13.795  1.00 18.22           C
ATOM     36  CD  GLU A   4      36.594  18.872 -13.095  1.00 21.03           C
ATOM     37  OE1 GLU A   4      36.475  17.649 -13.311  1.00 26.16           O
ATOM     38  OE2 GLU A   4      37.473  19.336 -12.338  1.00 21.82           O
ATOM     39  N   GLY A   5      32.460  21.061 -12.151  1.00  8.85           N
ATOM     40  CA  GLY A   5      32.293  22.268 -11.358  1.00  8.36           C
ATOM     41  C   GLY A   5      31.954  21.951  -9.913  1.00  8.62           C
ATOM     42  O   GLY A   5      32.462  22.595  -8.997  1.00  9.12           O
ATOM     43  N   GLU A   6      31.107  20.943  -9.713  1.00  9.33           N
ATOM     44  CA  GLU A   6      30.712  20.515  -8.376  1.00  9.68           C
ATOM     45  C   GLU A   6      31.908  19.866  -7.667  1.00  7.41           C
ATOM     46  O   GLU A   6      32.102  20.049  -6.465  1.00  8.69           O
ATOM     47  CB  GLU A   6      29.527  19.547  -8.457  1.00  9.10           C
ATOM     48  CG  GLU A   6      28.248  20.192  -9.006  1.00 23.34           C
ATOM     49  CD  GLU A   6      27.084  19.214  -9.163  1.00  9.26           C
ATOM     50  OE1 GLU A   6      27.297  18.092  -9.653  1.00 18.36           O
```

Figure 2. Sample data to be processed from .pdb files

We are developing a free DPI-online server to compute DPIs in real-time for the prioritization of high-quality PDB structures and quality-assessment of docking results.

The proposed computational tool compute DPI using Cruickshank, Blow and Goto DPI equations [3], [4], [5], [6] and will be available online on (https://www.chembioinf.ro).

REFERENCES

1. RCSB Protein Data Bank (PDB), http://www.rcsb.org/

2. Berman et al, The Protein Data Bank, Acta Cryst. (2002). D58, 899-907.

*3. Cruickshank, D. W. J., Remarks about protein structure precision. Acta Cryst. (1999) D**55**, 583–601.*

*4. Blow, D. M. (2002). Rearrangement of Cruickshank's formulae for the diffraction-component precision index , Acta Cryst. D**58**, 792–797.*

5. Goto J, Kataoka R, Hirayama N (2004) Ph4Dock: pharmacophorebased protein-ligand docking. J Med Chem 47:6804–6811.

6. Halip, L., Avram, S. & Neanu, C. The B-factor index for the binding site (BFIbs) to prioritize crystal protein structures for docking. Struct Chem 32, 1693–1699 (2021).