Journal of the Serbian Chemical Society
JSCS@tmf.bg.ac.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. *87* (3) S73–S79 (2022) *Supplementary material*

SUPPLEMENTARY MATERIAL TO

Large-scale comparison between the diffraction-component precision indexes favors Cruickshank's R_{free} function

SORIN AVRAM* and CRISTIAN NEANU**

Department of Computational Chemistry, "Coriolan Dragulescu" Institute of Chemistry, Timişoara, Romania

J. Serb. Chem. Soc. *87* (3) (2022) 321–330

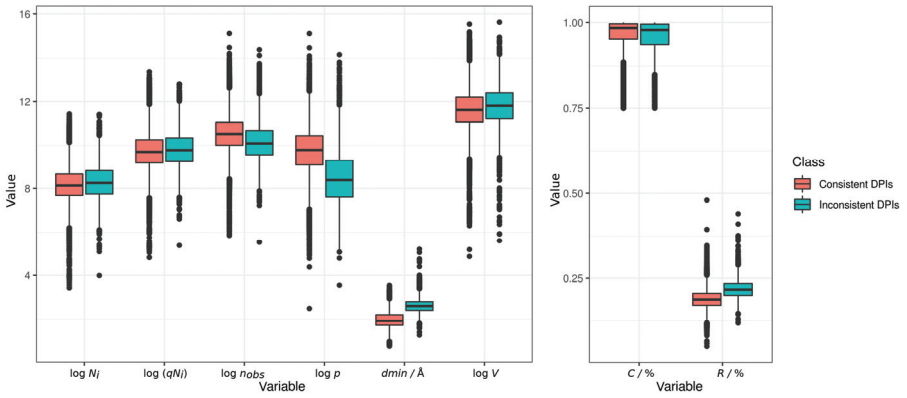

Fig. S-1. Boxplot showing the distribution of variables for consistent and inconsistent *DPI* values in *CR* and *BR* computations (see TABLE I for variables definitions).

*,** Correspondence E-mail: (*)sorin.avram@acad-icht.tm.edu.ro; (**)cristian.neanu@gmail.com

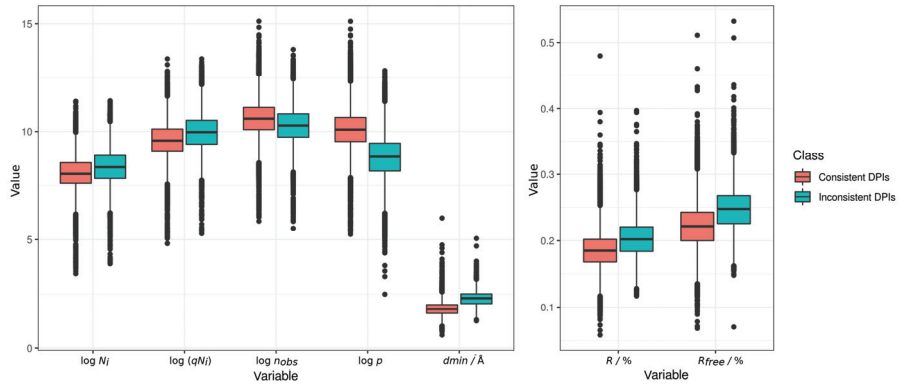

Fig. S-2. Boxplot showing the distribution of variables for consistent and inconsistent *DPI* values in *CR* and *CRF* computations (see Table I for variables definitions).

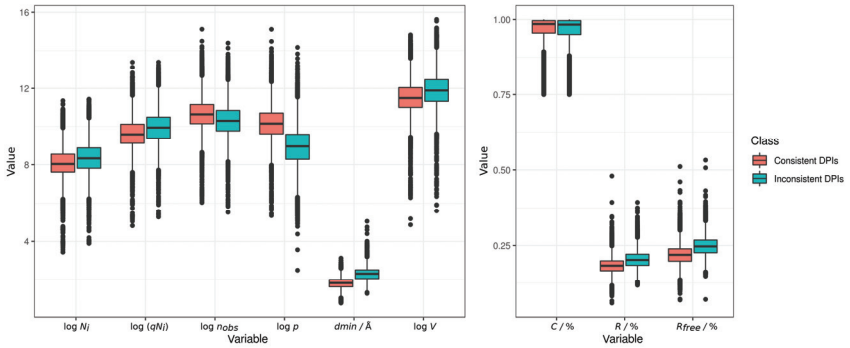

Fig. S-3. Boxplot showing the distribution of variables for consistent and inconsistent *DPI* values in *CR* and *BRF* computations (see Table I for variables definitions).

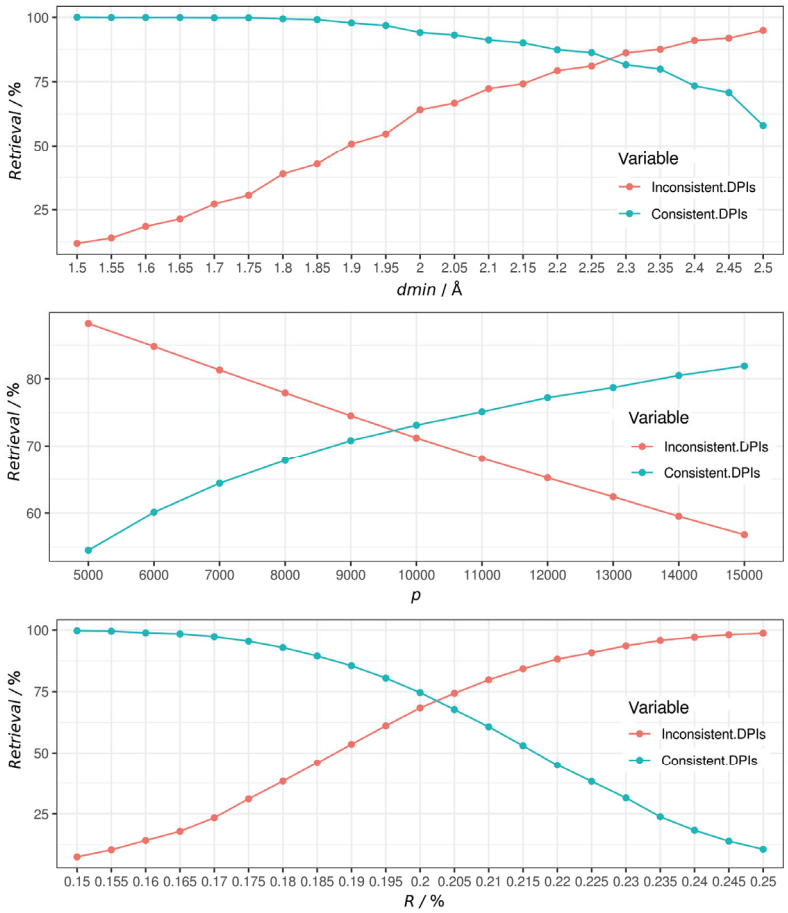

Fig. S-4. Retrieval percentages of consistent and inconsistent *DPI* class instances in *CR* and *BR* computation, at various thresholds for d_{min}, p and R (see TABLE I for variables definitions).

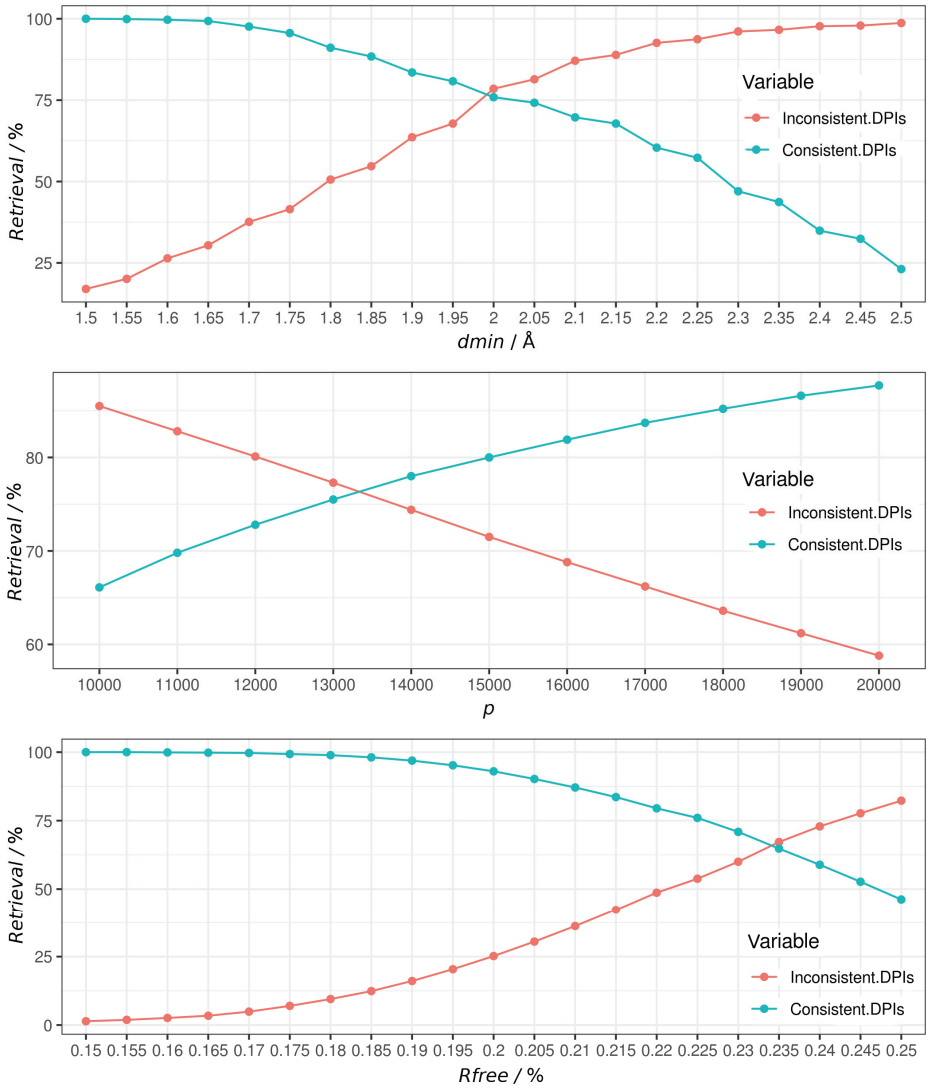

Fig. S-5. Retrieval percentages of consistent and inconsistent *DPI* class instances in *CR* and *CRF* computation, at various thresholds for d_{min}, p, and R_{free} (see Table I for variables definitions).

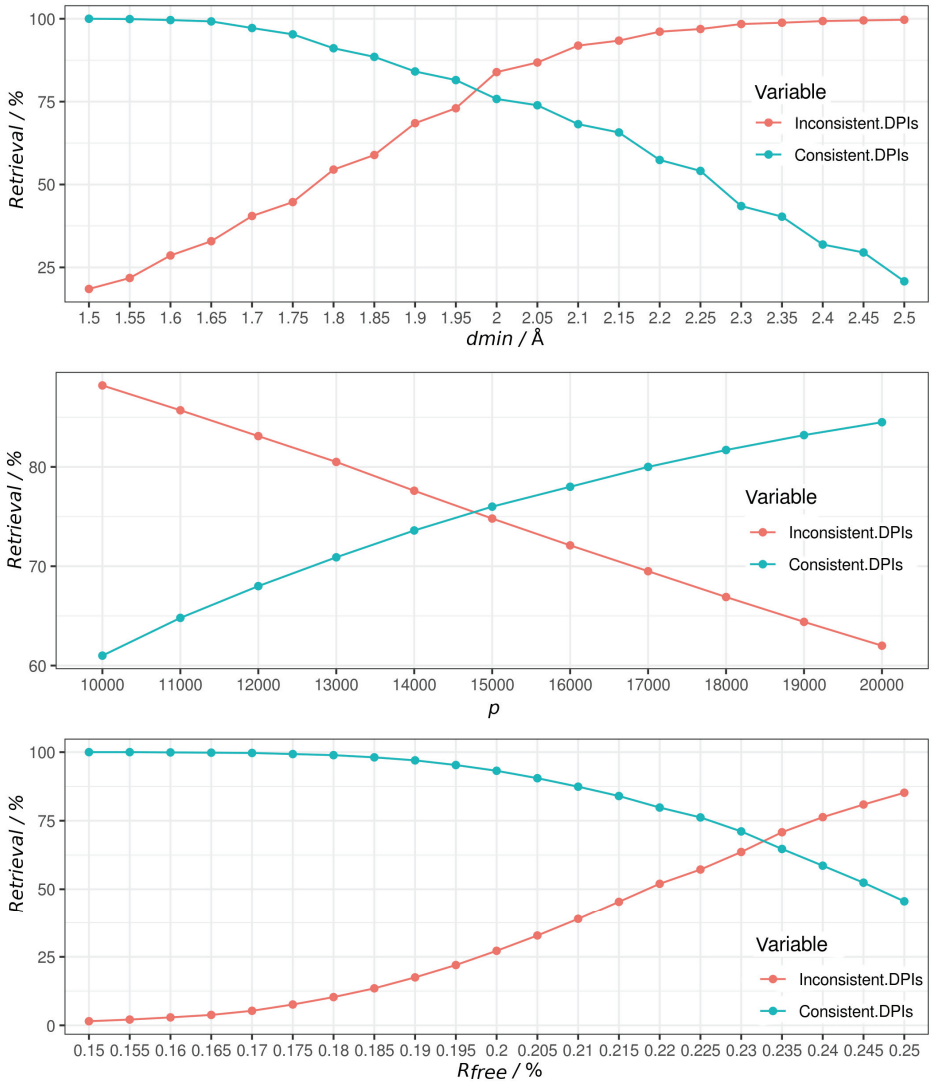

Fig. S-6. Retrieval percentages of consistent and inconsistent *DPI* class instances in *CR* and *BR* computation, at various thresholds for $d_{\min}$, p, and R_{free}. (see Table I for variables definitions).

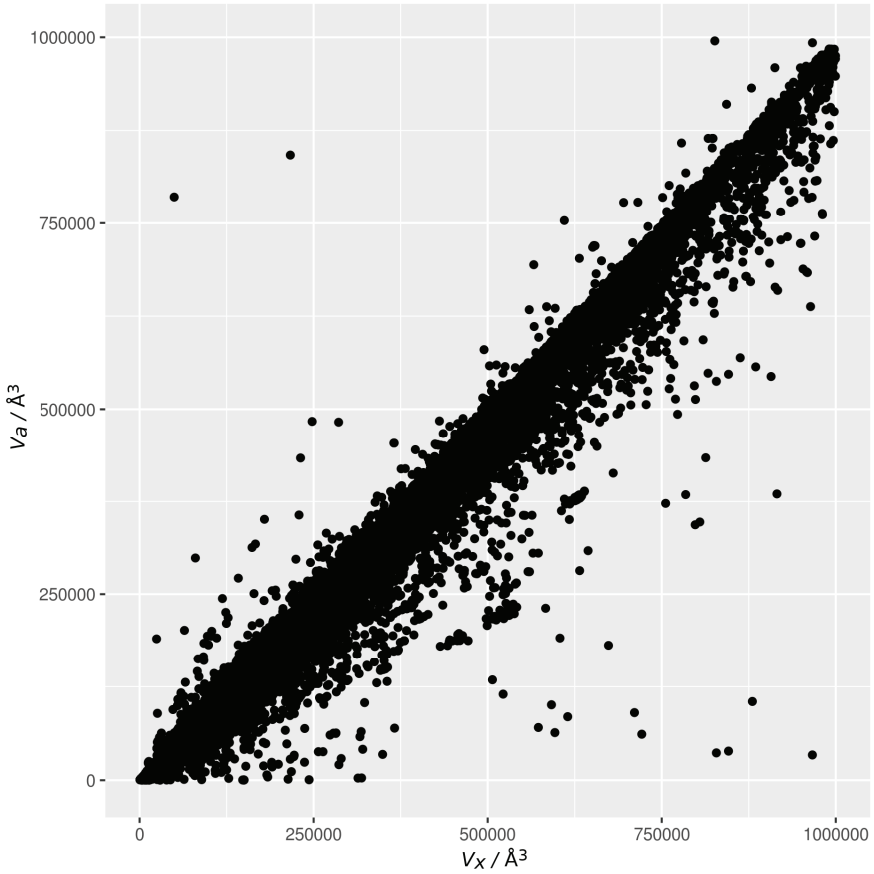

Fig. S-7. Plot of V_a, *i.e.* protein structure volume computed from Matthews coefficient, versus V_x, *i.e.* protein structure volume computed from the crystal asymmetric unit data.

Analysis of the DPI *data and corresponding variables for the subsets of alpha- and beta-helical PDB structures*

The general workflow which was used to assess the the alpha helix (AP) and beta sheets (BP) proteins was the following (Fig. S-8):

From our PDB downloaded database, 117,829 PDB structures passed the following criteria:

- Completeness ≥ 75 %
- Number of occupied atoms ≥90 %
- *DPI_CRF* computed and > 0

The lists of alpha helix (AP) and beta sheets (BP) proteins were downloaded from the RCSB PDB database (https://www.rcsb.org/)

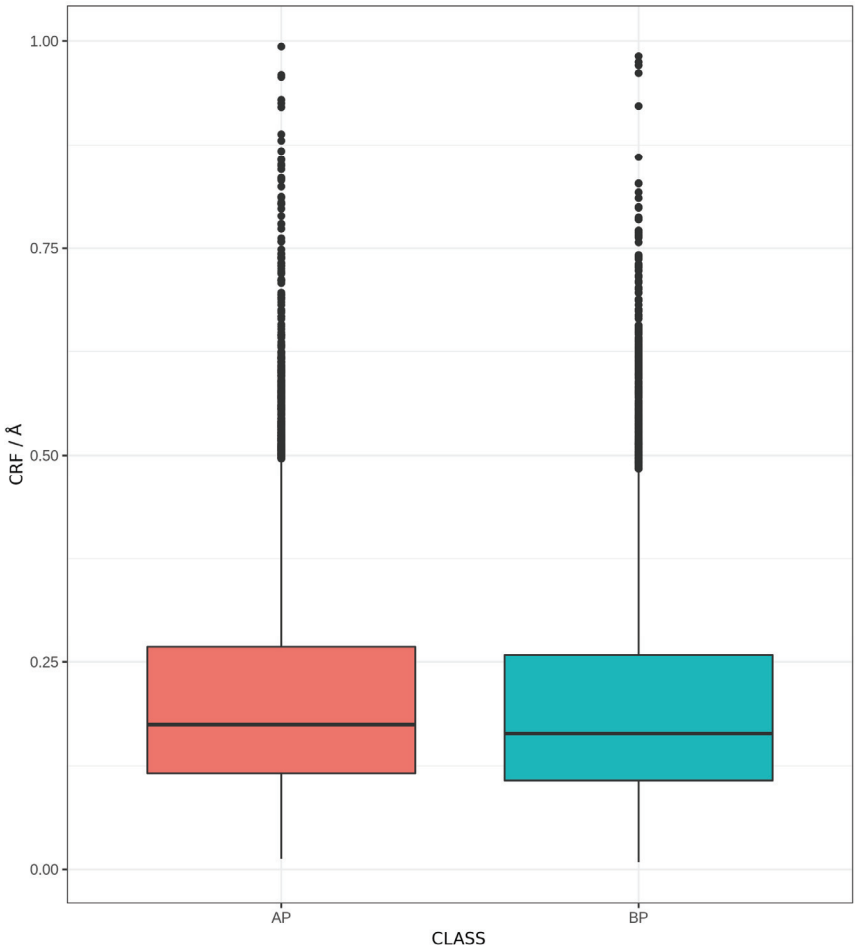

Fig. S-8. Boxplot showing the distribution of variables for AP (alpha helix proteins) and BP (beta helix proteins) *DPI* values in *CRF* computations.

The overlapping PDBs (1,648) between the APs (17,034 PDBs) and BP (26,893 PDBs) were removed from both sets, leaving a total of 40,631 PDB (37.9 % APs and 62.1 % BPs). For 29,802 PDBs *CRF_DPI* values were successfully computed (12,077 APs and 17,725 BPs).

The comparative analysis between alfa and beta proteins in terms of *DPI* revealed no significant differences (Figure S-8). This could be explained by the fact that the technology applied to obtain crystals and models is the same for lipo-soluble, water-soluble, alpha-helical or beta-chain structures, and consequently the accuracy of the positions of atoms is similar.