Journal of the Serbian Chemical Society
JSCS@tmf.bg.ac.rs • www.shd.org.rs/JSCS

J. Serb. Chem. Soc. *87* (3) 321–330 (2022)
JSCS–5524

Original scientific paper
Published 28 November 2021

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

(Received 18 May, revised 31 August, accepted 25 September 2021)

Abstract: This study aims to provide a first large-scale comparison between the various diffraction-component precision index (*DPI*) equations, assess the applicability of the parameter, and make recommendations on *DPI* computation. The *DPI* estimates the average accuracy of the atomic coordinates obtained by the structural refinement of protein diffraction data, with application in crystallography and cheminformatics. Although, Cruickshank and Blow proposed *DPI* equations based on *R* and R_{free} in order to calculate *DPI* values, which remain scarcely employed in the quality assessment of the Protein Data Base (PDB) files, due to the unclear data extraction protocols (to assign variables), the complex equations, the lack of extensive applicability studies and the limited access to automated computations. In order to address these shortcomings, the entire RCSB PDB database was evaluated using Cruickshank's and Blow's *R* and R_{free} *DPI* variations. Computations of 143070 X-ray structures indicate that R_{free}-based *DPI* equations apply to 30 % more protein structures compared to *R*-based *DPI* equations, with Cruickshank R_{free}--based *DPI* (*CR*F) exceeding the number of successful Blow's R_{free}-based *DPI* (*BR*F) computations. Although our results indicate that, in general, the resolutions < 2 Å assure consistency among the various *DPI*s computations (differences <0.05 Å), we recommend the use of *CR*F *DPI* because of its wider applicability.

Keywords: cheminformatics; drug discovery; protein structure; docking; crystallography.

INTRODUCTION

Cheminformatics has become an indispensable component of modern drug discovery.[1] Three-dimensional protein structures (or ligand-protein complexes) serve as the primary input for the structure-based methods, such as ligand

*,** Correspondence E-mail: (*)sorin.avram@acad-icht.tm.edu.ro; (**)cristian.neanu@gmail.com
https://doi.org/10.2298/JSC200518076A

docking, pharmacophore, and ligand design methods.[1,2] Different types of the PDB data mining analysis have been performed in the past, for example, to facilitate modelling of the synthetic protein structures with desired properties,[3] to study the cation-pi interactions in proteins,[4] *etc*. Additionally, protein-target homology modelling, molecular dynamics, and molecular mechanics simulations are state-of-the-art tools to build protein models, simulate and evaluate the conformational space of proteins, and ultimately detect and understand protein mechanisms and protein-protein and protein-ligand interactions.[2]

Beyond the limitations concerning the algorithms, the output of structure--based computational methods is a subject to the quality of the crystallographic protein structures, *i.e.*, the quality of the models fitting the experimental diffraction data.[5] The deposition of a new macromolecular structure in RCSB PDB (https://www.rcsb.org/)[6,7] is accompanied by multiple indicators describing experimental results and model validation (https://www.wwpdb.org/),[8–10] but lack the unanimous definitions for high-quality structures.[11] With more than 166 thousand macromolecular structures, the PDB database is among the most commonly employed digital resources for protein structures, impacting multiple scientific disciplines, *e.g.*, materials sciences, physics, computer science, chemistry, engineering, and mathematics.[12]

The essential indicators to assess the suitability of crystallographic biomolecules for structure-based methods were extensively reviewed by Warren *et al.*[5] The authors observed that PDB entry selection is often reasoned by the resolution of the X-ray crystal structure, a measure of data (electron density) quantity, misleadingly used as a measure of data quality. Additionally, model quality can be calculated using *R*-factors (R or R_{free})[13] measuring the difference between measured data and data predicted from the model.[5] Warren *et al.*[5] pointed towards the diffraction-component precision index,[14] as a global criterion to computational outputs (such as protein-ligand docking).[5] Depending on the availability of input parameters, Cruickshank proposed an *R*- and R_{free}-based version of *DPI*.[14] To make *DPI* computation more accessible, a few years later, Blow[15] rearranged Cruickshank's formulas in a simplified version (*vide infra*). assess both model and data quality.

The diffraction-component precision index (*DPI*), developed by Cruickshank,[14] estimates the coordinate error of atoms in protein crystal structures, providing means to select highly precise data, as well as to determine the accuracy limit of them.

Several studies have underlined the importance of *DPI* in the evaluation of the quality of PDB structures supplied to computational approaches,[5,16–21] but *DPI* continues to be scarcely employed, due to the unclear data extraction protocols (to assign variables), the complex equations, the lack of extensive applicability studies, and the limited access to automated computations. To date,

we identified only the free online resources offered by K. Sekar's laboratory facilitating the computation of Cruickshank *DPI*.[19] Moreover, to our knowledge, no comparative studies on large numbers of PDB files between Cruickshank and Blow *DPI* equations (for R and R_{free}) were reported in the literature.

This study aims to provide a first large-scale comparison between the various *DPI* equations, assess the applicability of the parameter, and make recommendations on *DPI* computation.

EXPERIMENTAL

PDB data and variables

The entire RCSB PDB database was downloaded (159678 PDB files; accessed on Jan 28, 2020) and 143070 X-ray structures were retained (NMR and 3D electron microscopy-based determined structures were discarded). The PDB files were used to extract (and compute) the variables in Table I.

TABLE I. Variables extracted from Protein Data Base (PDB) files employed for diffraction-component precision index (*DPI*) calculations.

Name[a]	Description	Observations
N_i	Number of occupied atoms	Count of non-hydrogen atoms with occupancy of 1.0 (from ATOM and HETATM)
N_{prot}	Number of protein atoms	Count of non-hydrogen protein atoms in the PDB file (from ATOM)
n_{obs}	Number of observations/reflections	from REMARK 3
q	Number of parameters to be refined for each non-H	For fully anisotropic refinement $q = 9$ For isotropic refinement, $q = 4$[14]
p	Number of free parameters	$p = n_{obs} - n_{param}$[14]
n_{param}	Number of parameters to be refined	$n_{param} = N_i q$[15]
C / %	Completeness	from REMARK 3
R / %	The normalized linear residual between observed and calculated structure factor amplitudes	Protein atomic model refinement R factor (from REMARK 3
R_{free} / %	Cross-validation R value (R value calculated for the cross-validation data excluded from refinement)[13]	Protein atomic model refinement R_{free} factor (from REMARK 3)
V / Å^3	Crystal asymmetric unit volume	i. $V = V_M\, 14.12\, N_{prot}$,[15] V_M is Matthews coefficient (from REMARK 280) ii. volume of the asymmetric unit computed from the crystal unit description iii. (from *CR*YST1)
d_{min} / Å	Resolution	from REMARK 2 or REMARK 3: RESOLUTION RANGE HIGH

[a]A list of variables, extracted (and computed) from PDB files, to calculate *DPI* values is displayed

DPI computations

Based on the parameters described in Table I, the *DPI* was computed according to the Cruickshank and Blow equations as described in Table II.

TABLE II. Diffraction-component precision index (*DPI*) equations computed using the variables described in Table I; *CR*: *DPI* computed with *R*-based *DPI* Cruickshank formula, *CRF*: *DPI* computed with R_{free}-based *DPI* Cruickshank formula, *BR*: *DPI* computed with *R*-based *DPI* Blow formula; *BRF*: *DPI* computed with Rfree-based *DPI* Blow formula (see Table I for variable definitions)

DPI Equation[a]		Ref.
$CR = \left(\frac{N_i}{p}\right)^{-1/2} C^{-1/3} R d_{min}$	(1)	14
$CRF = \left(\frac{N_i}{n_{obs}}\right)^{-1/2} C^{-1/3} R_{free} d_{min}$	(2)	14
$BR = 1.28 N_i^{1/2} \left(1 - \frac{q N_i}{n_{obs}}\right)^{-1/2} V^{1/3} n_{obs}^{-5/6} R$	(3)	15
$BRF = 128 N_i^{1/2} V_{-1/3} n_{obs}^{-5/8} R_{free}$	(4)	15

For proteins containing several unique protein chains (*e.g.*, 5DIS – 4 unique protein chains: A, B, C, D) we have taken in to account and processed only the data from the "A" chain.

Therefore, in order to calculate the Blow-*DPI* (Eqs. (3) and (4)), the crystal volume for all proteins was evaluated by processing the number of atoms from the "A" chain.

Following Gurusaran's *et al.* suggestion,[20] we retained 120594 structures which met the following 2 criteria: completeness C >75 % and the percentage of fully occupied atoms in the protein structure >90 %.

RESULTS AND DISCUSSION

Comparative analysis of DPI *computations*

The results indicate that for 120484 PDB IDs *(i.e.*, 84 % of the number of X-ray entries available), at least one *DPI* value was calculated according to the equations in Table II using the parameters extracted from the PDB files as described in Table I. In the case of 87451 entries (*i.e.*, 62 % of the number of X-ray structures) the *DPI* values were successfully determined by all four equations. Following Gurusaran *et al.*,[20] we focused on *DPI* values <1 Å, which we considered a reasonable cutoff for the analysis of relevant *DPI* computations. Thereby, 3338 entries (2.8 %) with *DPI* values >1 Å, obtained through either of the equations, were discarded.

R_{free}-based *DPI*s showed larger applicability (117562 PDB entries) compared to *R*-based *DPI*s (88620 PDB entries). The distribution of the values followed very similar trends with 90 % showing *DPI* values <0.5 Å. Cruickshank *DPI*s indicated median values of 1.7 Å, similar to Blow *DPI*s (1.6 Å), as it can be seen in Fig. 1A.

Further, the pairwise correlation between *DPI* computations at *DPI* cutoffs ranging stepwise from < 1 to < 0.1 Å (Fig. 1B) were calculated. Very strong correlations (Pearson >0.97) were found in (*CR*, *BR*) pairs as well as (*CRF*, *BRF*) pairs independent of the *DPI* cutoffs. Contrary to that, the *R*-based and R_{free}-based *DPI* outputs coincided more at smaller *DPI* values (Fig. 1B), *e.g.*, the correlation of (*CR*, *CRF*) and (*BR*, *BRF*) pairs increased from 0.85 for *DPI* values <1 Å, to 0.96 Å at *DPI* values < 0.1 Å. These results suggested that, in general, either Cruikshank or Blow equations should be used, and differences between *R* and R_{free} *DPI*s values are more likely to be encountered at lower *DPI* values.

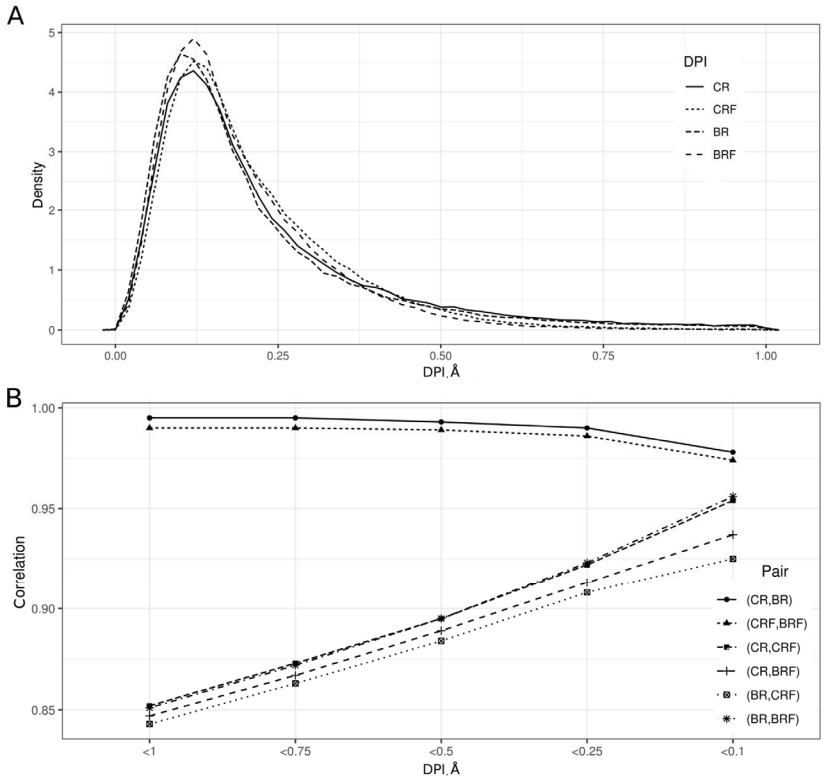

Fig. 1. A) Distribution of *DPI* values computed by 4 equations. B) Pairwise correlation between *DPI* computations at *DPI* cutoffs ranging stepwise from 1 to <0.1 Å.

The R_{free} equations were calculated for ~30 % more protein structures compared to *R*-based *DPI*, a significant advantage. Surprisingly, more proteins (2685 PDB structures) allowed *CRF* calculations compared to *BRF*. Although this difference is small, it was expected that Blow's equation would be more applicable due to fewer input variables, commonly contained in the header of a PDB file.[16]

In the next sections, it was aimed to find the source of inconsistency between the outputs of the *DPI* computations and further the limiting factors in *DPI* calculations.

Key variables for DPI *consistency*

To identify the parameters responsible for the variations in *DPI* calculations (obtained with the equations in Table II), *CR* was used as a reference and the PDB entries for which *CR* < 1 Å were explored. The absolute difference (Δ) between *CR* values and the outputs of each *DPI* of the other equations was computed pair-wisely. Consistency has been defined around a threshold of 0.05. Accordingly, the data were split into two groups: *DPI* differences ≤0.05 Å were considered "Consistent *DPI*s" and *DPI* differences >0.05 Å "Inconsistent *DPI*s". For example, $\Delta CRF_i = |CR_i - CRF_i|$ defines the difference between CR_i and CRF_i *DPI* values obtained for PDB entry *i*. If $\Delta CRF_i \leq 0.05$ Å, *i* is labeled as "consistent *DPI*s", and otherwise as "inconsistent *DPI*s" in the Δ*CRF* set. Similarly, Δ*BR* and Δ*BRF* using *CR* were computed as a reference. The size of the sets and the number of class representatives are reported in Table III.

TABLE III. Important variables for diffraction-component precision index (*DPI*) consistency in Δ*BR*, Δ*CRF* and Δ*BRF* sets (see Table I for variable definitions)

DPI Comparison	Inconsistent/ consistent *DPI* counts	Variable	Threshold value[a,b]	Retrieval, %		
				Inconsistent *DPI*[c]	Consistent *DPI*[c]	Overall
Δ*BR*	5027/81289	d_{min} / Å	2.3	86.2	81.6	85.9
		p	10000	71.2	73.1	71.3
		R / %	0.2	68.4	74.6	68.7
Δ*CR*F	28645/58143	d_{min} / Å	2	78.5	75.9	77.6
		p	13000	77.3	75.5	76.8
		R_{free} / %	0.235	67.2	64.8	66.4
Δ*BR*F	32850/51800	d_{min} / Å	2	83.9	75.8	80.8
		p	15000	74.8	76.0	75.3
		R_{free} / %	0.235	70.8	64.7	68.4

[a]Inconsistent *DPI* identified for d_{min} >, R_{free} >, *p* < threshold values (see Figs. S-1–S-3 of the Supplementary material to this paper); [b]*p* is nondimensional, *R* and R_{free} are equivalent with percentages; [c]consistent *DPI*s is when the difference between *R*-based Cruickshank *DPI* values (*CR*) and the output of each *DPI* from the other equations is ≤0.05; Inconsistent *DPI*s is when the difference between *R*-based Cruickshank *DPI* values (*CR*) and the output of each *DPI* from the other equations is > 0.05 (see 3.2 Key variables for *DPI* consistency)

Based on the *DPI* consistency assessment, each variable required for the computation of *DPI*s was explored. According to the largest shifts between the distributions of the classes (see Figs. S-1–S-3 of the Supplementary material to this paper), the first 3 most important variables for each set: d_{min}, *p*, and *R* for Δ*BR*; d_{min}, *p* and R_{free} for Δ*CFR* and Δ*BFR*, respectively (Table III) were retrieved. For each variable, a range of values to find the threshold that best dis-

criminates between the classes, *i.e.*, the best trade-off between the retrieval of "Consistent *DPI*s" and "Inconsistent *DPI*s" class members (Figs. S-4–S-6 of the Supplementary material) was systematically explored.

The results pointed towards d_{min} to be the major factor contributing to the differences in *DPI* results. In fact, in 81.6 % of the cases, consistent *DPI* values between *CR* and *BR* results were obtained for d_{min} < 2.3 Å. For 76 % of the structures, d_{min} < 2 Å identified consistent *DPI*s in the Δ*CRF* and the Δ*BRF* sets. The another important variable is p. The threshold values of p < 10000, <13000 and <15000, retrieved were 73.1, 75.5 and 76 %, respectively, of the 'Consistent *DPI*' group in Δ*BR*, Δ*CRF* and Δ*BRF* sets, respectively.

Based on the current results, for most PDB files, consistent *DPI* values are expected to appear at d_{min} < 2 Å. The resolution of the PDB structure serves as input only for *CR* and *CRF* calculation and describes the level of detail present in the diffraction pattern. Although, in general, high-resolution structures (*e.g.*, d_{min} < 2 Å) ensure sufficient density to obtain good models, this is not always true.[5,16]

Failures in DPI *computations*

Out of 143070 structures, *DPI* computation failed for 3125 due to missing variables or inadequate component values, such as $p = n_{obs} - qN_i$ or $(1–qN_i/n_{obs})$ in Eqs. (1) and (3), respectively (see Table II). In the case of 41262 entries, *i.e.*, 30 % of the PDB, *DPI* computations by all 4 equations were not possible (Gurusaran *et al.*[20] filters were not applied here).

Two major variables that impaired the computation of *DPI*s, *i.e.*, p and qN_i/n_{obs} (Table IV) were identified. Values of $p \leq 0$ or $qN_i/n_{obs} \geq 1$ made *CR* and *BR* calculation impossible and showed a relatively high prevalence (~22 %, *i.e.*, ~32000 entries) in the PDB database. This explained the limited *R*-based *DPI* applicability compared to R_{free}-based *DPI*s. The later ones were predominantly affected by the availability of the R_{free} factor (in 6021 PDB entries).

TABLE IV. Percentages of Protein Data Base entries with missing/inadequate variable values (see Table I) involved in diffraction-component precision index computations (see Table II)

Variable	p	qN_i/n_{obs}	R_{free}	V	C	n_{obs}	V_M	N_{prot}	R	Other[a]
PDB entries, %	24.28	22.64	4.25	2.87	2.53	1.65	1.63	1.49	0.29	<0.1

[a]N_i, V_a and d_{min}

Errors in the calculation of the volume V (present in the Blow *DPI* equations eq 3 and eq 4) affected 2.87 % (4068) structures. In this particular case, V can be computed using the unit cell lengths and angles (see Table I), providing very similar values to the V_M-based volume (Pearson correlation of 0.858, Fig. S-7). Consequently, the corresponding *DPI* values indicated very strong correlations of 0.997 and 0.995 with *BR* and *BRF* *DPI*s, respectively. Moreover, it can be applied to 2788 more PDB compared to *BR* and 3310 more compared to *BRF*

(*DPI* cutoff < 1 Å). Larger values of *DPI*, *e.g.*, >2 Å, are encountered in cases with disproportionate n_{obs} relative to the volume or the number of atoms. In these cases, small n_{obs} can generate large *DPI* values.

CONCLUSION

In drug discovery studies crystallographic protein structures are used extensively for protein-ligand docking and molecular dynamics. Using a reliable measure to prioritize the fittest structures (if multiple is available) is of significant importance for the outcome of the experiments as recently shown by Halip *et al.*[22] *DPI* is one potentially valuable indicator, but difficulties in computation might have decreased its application in medicinal chemistry and cheminformatics.

Our comprehensive exploration of four *DPI* variants, showed that R_{free}-based *DPI* computations (*CRF* and *BRF*) can be determined for the vast majority of the currently available PDB entries and provide similar outputs to *R*-based *DPI* types (*BR* and *CR*), especially in structures with small *DPI* values. Zero or negative values for p or $1 - qN_i/n_{obs}$ are the major sources of computational failures in *R*-based computations. The differences in *DPI* results are strongly influenced by the resolution (d_{min}) and the number of parameters (p). In general, *DPI* computations in structures with resolutions <2 Å will result in small differences between the outputs of the four approaches

NOMENCLATURE

DPI: Diffraction-component precision index
CR: *DPI* computed with R-based *DPI* Cruickshank formula,
CRF: *DPI* computed with Rfree-based *DPI* Cruickshank formula,
BR: *DPI* computed with R-based *DPI* Blow formula
BRF: *DPI* computed with Rfree-based *DPI* Blow formula

SUPPLEMENTARY MATERIAL

Additional data and information are available electronically at the pages of journal website:https://www.shd-pub.org.rs/index.php/JSCS/article/view/10775, or from the corresponding author on request.

Acknowledgements. We wish to thank Dr. Ana Borota for her constructive comments and suggestions regarding the manuscript. This work was supported by "Coriolan Drăgulescu" Institute of Chemistry, Timisoara, Romanian Academy, Romania, project number 1.1.4/2019/2020.

ИЗВОД

ОБИМНО ПОРЕЂЕЊЕ ИНДЕКСА ПРЕЦИЗНОСТИ ДИФРАКЦИОНИХ КОМПОНЕНТИ ДАЈЕ ПРЕДНОСТ CRUICKSHANK R_{free} ФУНКЦИЈИ

SORIN AVRAM и CRISTIAN NEANU

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

Ова студија тежи да обезбеди прво обимно поређење различитих једначина индекса прецизности дифракционих компоненти (*DPI*), процени применљивост параметра, и направи препоруке за израчунавање *DPI*. *DPI* процењује просечну поузданост атомских

координата добијених финим-структурним подешавањем података дифракције за протеине, са применом у кристалографији и хеминформатици. Мада су Cruickshank и Blow предложили *DPI* једначине засноване на *R* и R_{free} за израчунавање *DPI* вредности, оне су нису погодне за коришћење за процену квалитета фајлова протеинске базе података (PDB), због нејасних протокола за издвајање података (да би им се приписале варијабле), сложености једначина, недостатка екстензивних студија применљивости и ограниченог приступа аутоматизованим израчунавањима. Да би превазишли ове недостатке целокупна RCSB PDB база података је проверена користећи Cruickshank и Blow *R* и R_{free} *DPI* варијације. Израчунавања на 143070 кристалних структура указују на то да су R_{free}-засноване *DPI* једначине применљиве на 30 % више протеинских структура у поређењу са *R*-заснованим *DPI* једначинама, са Cruickshank R_{free}-заснованим *DPI* (*CRF*) израчунавањима која превазилазе број успешних Blow R_{free}-заснованих *DPI* (*BRF*) израчунавања. Иако приказани резултати указују да, опште узевши, резолуција структуре <2 Å обезбеђују конзистентност различитих *DPI* израчунавања (разлике су < 0,05 Å), препоручује се да се користи *CRF DPI* због његове шире применљивости.

(Примљено 18. маја, ревидирано 31. августа, прихваћено 25. септембра 2021)

REFERENCES

1. P. Willett, *WIREs Comput. Mol. Sci.* **1** (2011) 10 (https://doi.org/10.1002/wcms.1)
2. G. Sliwoski, S. Kothiwale, J. Meiler, E.W. Lowe Jr., *Pharmacol. Rev*. **66** (2014) 334 (https://doi.org/10.1124/pr.112.007336)
3. D. M. Popović, S. D. Zarić, B. Rabenstein, E.-W. Knapp, *J. Am. Chem. Soc.* **25** (2001) 6040 (https://doi.org/10.1021/ja003878z)
4. S. Zarić, D. Popović, E.-W. Knapp, *Chem. Eur. J*. **6** (2000) 3935 (https://doi.org/10.1002/1521-3765(20001103)6:21%3C3935::AID-CHEM3935%3E3.0.CO;2-J)
5. G. L. Warren, T. D. Do, B. P. Kelley, A. Nicholls, S. D. Warren, *Drug Discov. Today* **17** (2012) 1270 (https://doi.org/10.1016/j.drudis.2012.06.011)
6. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res*. **28** (2000) 235 (https://doi.org/10.1093/nar/28.1.235)
7. S. K. Burley, H. M. Berman, C. Bhikadiya, C. Bi, L. Chen, L. Di Costanzo, C. Christie, K. Dalenberg, J. M. Duarte, S. Dutta, Z. Feng, S. Ghosh, D. S. Goodsell, R. K. Green, V. Guranovic, D. Guzenko, B. P. Hudson, T. Kalro, Y. Liang, R. Lowe, H. Namkoong, E. Peisach, I. Periskova, A. Prlic, C. Randle, A. Rose, P. Rose, R. Sala, M. Sekharan, C. Shao, L. Tan, Y. P. Tao, Y. Valasatava, M. Voigt, J. Westbrook, J. Woo, H. Yang, J. Young, M. Zhuravleva, C. Zardecki, *Nucleic Acids Res*. **47** (2019) D464 (https://doi.org/10.1093/nar/gky1004)
8. H. Berman, K. Henrick, H. Nakamura, *Nat. Struct. Biol*. **10** (2003) 980 (https://doi.org/10.1038/nsb1203-980)
9. H. Berman, K. Henrick, H. Nakamura, J. L. Markley, *Nucleic Acids Res*. **35** (2007) D301 (https://doi.org/10.1093/nar/gkl971)
10. wwPDB consortium, *Nucleic Acids Res*. **47** (2019) D520 (https://doi.org/10.1093/nar/gky949)
11. M. J. Domagalski, H. Zheng, M. D. Zimmerman, Z. Dauter, A. Wlodawer, W. Minor, *Structural Genomics. Methods in Molecular Biology, Vol. 1091,* Humana Press, Totowa, NJ, 2014, p. 297 (https://doi.org/10.1007/978-1-62703-691-7_21)
12. N. V. Zukang Feng, Luigi Di Costanzo, D. S. Goodsell, J. D. Westbrook, S. K. Burley, C. Zardecki, *Data Sci. J.* **19** (2020) 25 (http://doi.org/10.5334/dsj-2020-025)

13. A. T. Brunger, *Nature* **355** (1992) 472 (https://doi.org/10.1038/355472a0)
14. D .W. Cruickshank, *Acta Cryst., D* **55** (1999) 1108 (https://doi.org/10.1107/s0907444999004308)
15. D .M. Blow, *Acta Crystallogr., D* **58** (2002) 792 (https://doi.org/10.1107/s0907444902003931)
16. P. C. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson, M. T. Stahl, *J. Chem. Inf. Model.* **50** (2010) 572 (https://doi.org/10.1021/ci100031x)
17. P. C. Hawkins, G. L. Warren, A. G. Skillman, A. Nicholls, *J. Comput.-Aided Mol. Des.* **22** (2008) 179 (https://doi.org/10.1007/s10822-007-9166-3)
18. S. I. Avram, L. Crisan, L. M. Pacureanu, A. Bora, E. Seclaman, M. Balint, L. G. Kurunczi, *Med. Chem. Res.* **22** (2012) 3589 (https://doi.org/10.1007/s00044-012-0367-5)
19. K. S. D. Kumar, M. Gurusaran, S. N. Satheesh, P. Radha, S. Pavithra, K. P. S. Thulaa Tharshan, J. R. Helliwell, K. Sekar, *J. Appl. Crystallogr.* **48** (2015) 939 (https://doi.org/10.1107/s1600576715006287)
20. M. Gurusaran, M. Shankar, R. Nagarajan, J. R. Helliwell, K. Sekar, *IUCrJ* **1** (2014) 74 (https://doi.org/10.1107/S2052252513031485)
21. J. Goto, R. Kataoka, N. Hirayama, *J. Med. Chem*. **47** (2004) 6804 (https://doi.org/10.1021/jm0493818)
22. L. Halip, S. Avram, C. Neanu, *Struct. Chem.* **32** (2021) 1693 (https://doi.org/10.1007/s11224-021-01751-9).