crystals

MDPI

Review

B-Factor Rescaling for Protein Crystal Structure Analyses

Georg Mlynek [1], Kristina Djinović-Carugo [1,2,3] and Oliviero Carugo [4,*]

[1] Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, 1010 Vienna, Austria; georg.mlynek@univie.ac.at (G.M.); krstina.djinovic@embl.fr (K.D.-C.)
[2] Department of Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, 1000 Ljubljana, Slovenia
[3] European Molecular Biology Laboratory (EMBL) Grenoble, 38000 Grenoble, France
[4] Department of Chemistry, University of Pavia, 27100 Pavia, Italy
* Correspondence: olivieroitalo.carugo@unipv.it

Abstract: The B-factor, also known as the atomic displacement parameter, is a fundamental metric in crystallography for quantifying the positional flexibility of atoms within crystal lattices. In structural biology, various developments have expanded the use of B-factors beyond conventional crystallographic analysis, allowing for a deeper understanding of protein flexibility, enzyme manipulation, and an improved understanding of molecular dynamics. However, the interpretation of B-factors is complicated by their sensitivity to various experimental and computational factors, necessitating rigorous rescaling methods to ensure meaningful comparisons across different structures. This article provides an in-depth description of rescaling approaches used for B-factors. It includes an examination of several methods for managing conformational disorder and selecting the atom types required for the analysis.

Keywords: anisotropy; atomic displacement parameter; B-factor; B-factor rescaling; conformational disorder; protein crystal structure; Protein Data Bank

check for updates

Citation: Mlynek, G.; Djinović-Carugo, K.; Carugo, O. B-Factor Rescaling for Protein Crystal Structure Analyses. *Crystals* **2024**, *14*, 443. https://doi.org/10.3390/cryst14050443

Academic Editor: Jesús Sanmartín-Matalobos

Received: 16 April 2024
Revised: 1 May 2024
Accepted: 1 May 2024
Published: 7 May 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. The Utility of B-Factors

The B-factor, named the "atomic displacement parameter" by the International Union of Crystallography [1] and also known as the thermal or temperature factor, is a key quantity used to describe the inherent thermal vibrations of atoms in a crystal lattice. It is essential for understanding the dynamic properties of individual atoms in a crystal structure, providing significant information on the strength and flexibility of the material being studied [2].

Mathematically, the B-factor B is expressed as:

$$B = 8\pi^2 \left\langle u^2 \right\rangle \quad (1)$$

where $\langle u^2 \rangle$ is the mean square displacement of the atom from its equilibrium position. This parameter is traditionally measured in units of square angstroms (Å^2). The B-factor serves as a quantitative indicator of the extent to which individual atoms exhibit deviations from their average or equilibrium positions, primarily due to thermal excitations. In essence, it encapsulates the probabilistic nature of atomic positions within the crystal lattice, manifested as spatial dispersion.

The B-factor assumes a pivotal role in crystallography, exerting a profound influence on the accuracy of electron density maps generated through X-ray or neutron diffraction experiments. Higher B-factors are indicative of increased thermal motion, suggesting greater atomic disorder, while lower B-factors signify relative stability and a more orderly disposition of atoms within the crystal.

In the scientific literature, the concept of B-factors finds its roots in foundational works such as the Debye–Waller factor, elucidated by Peter Debye and Arthur L. Mackay in the

early 20th century [3,4]. Subsequently, the refinement of crystal structures employing B-factors became an integral part of crystallographic analysis. Researchers regularly employ the B-factor to refine crystal structures, unravel material properties, and discern the ramifications of atomic vibrations on the macroscopic behavior of crystalline substances.

Few recent reviews were published on the usage of B-factors in structural biology [5–7]; only some very recent findings are commented on below.

Pearce and Gros have recently proposed a method to break down B-factors into a hierarchical series of contributions, spanning from the overall mobility of the whole protein to the mobility of secondary structural elements, residues, and individual atoms [8].

Several studies were dedicated to protein engineering for biotechnological applications. The enzymatic production of L-tryptophan in *Escherichia coli* tryptophan synthase was enhanced by the application of rational molecular engineering, which included analyzing B-factors and designing in silico mutants [9]. Four variants of *Candida antarctica* lipase B, exhibiting a stereoselectivity of over 90% towards substrates with two stereo-centers, were produced; the B-factor analysis revealed that all of these variants included a loop that displayed much more flexibility compared to the wild-type enzyme [10]. Tang and coworkers were able to enhance the thermostability of subtilisin E-S7 peptidase by identifying unstable residues and regions by means of a B-factor analysis [11].

However, a number of additional, perhaps unexpected, applications of B-factors were reported. Blum and coworkers used B-factors as a benchmark for assigning electrical charges to zinc cations in insulin and thermolysin [12]. Espinosa and colleagues reported molecular dynamics simulations in a heart fatty acid binding protein crystal by tuning restraints based on B-factors [13]. Sánchez Rodríguez and colleagues examined the importance of B-factors in molecular replacement computations for solving the phase problem in protein X-ray crystallography [14]. Bae and colleagues analyzed the structural flexibility in two loops and in the C-terminal region of *Rhodothermus marinus* substrate binding protein to reach a deeper understanding of the binding mechanism [15]. Johnson and coworkers related B-factors to drug potency—rizotinib and lorlatinib—against C-ros oncogene 1 kinase and anaplastic lymphoma kinase [16].

The versatility of B-factors in delivering information about protein structural features is apparent. Nevertheless, as shown in the following chapter, B-factors possess a fundamental drawback, akin to a biblical original sin, that renders them challenging to use in their unprocessed state since their values are altered by multiple variables, some of which are unrelated to molecular mobility and flexibility.

2. B-Factor Non-Transferability

It is well recognized that B-factors may vary amongst structures of the same molecule due to many factors that are not directly connected to their molecular interpretation [17]. Recently, Ramos et al. determined 20 times the 1.6 Å resolution hen egg lysozyme crystal structure by using two dissimilar X-ray diffraction setups and by using a single automatic refinement pipeline and observed a variation in the B-factor for each independent crystal [18]. Similar observations were reported for human insulin and sperm whale myoglobin structures deposited in the PDB [18].

Amongst the sources of variability, it is possible to divide those that are inherent to the X-ray diffraction experiments and those that depend on data computational processing. In the first category, it is necessary to mention incident beam alignment and optics, instability of the mechanics, systematic errors in measurements of diffraction intensities, primary or secondary extinction, varying density of solid-state defects and mosaicity, radiation damage, which may cause variations in the same protein between atoms that are strongly impacted by X-ray absorption and those that remain unaffected, and the variable content of amorphous solvent within the crystals. In the second category, it is possible to mention peak detection and integration, signal-to-noise cutoff, and background handling. Also, the use of stereochemical restraints, such as those on bond lengths and angles, can have a considerable impact on B-factors, especially with low-resolution data.

In this regard, it is crucial to bear in mind that the variability of the B-factors in a single protein crystal structure is influenced by the weights assigned to another type of refinement restraint. Indeed, the B-factors of atoms that are connected by a covalent bond are forced to be comparable in order to account for the rigidity of the covalent bonds [19,20]. This unquestionably has an impact on the distribution of B-factor values.

Moreover, it should not be forgotten that rigid body motions of molecular moieties within the crystal may occur and influence scattering and B-factors [21,22].

Beyond these experimental/computational potential artifacts, it is essential to remember the importance of crystallographic resolution, which can be either the cause or the consequence of B-factors' variability. In fact, given that resolution is defined as

$$resolution = \frac{\lambda}{2sin\theta} \tag{2}$$

where λ is the X-ray wavelength and θ is the diffraction angle, and given the dependence of the atomic form factor f on B described by

$$f = f_0 \cdot exp\left(-\frac{Bsin^2\theta}{\lambda^2}\right) \tag{3}$$

where f_0 is the atomic form factor at B = 0 Å^2, it results that

$$B = -4\left(\ln\frac{f}{f_0}\right)resolution^2 \tag{4}$$

where $f \leq f_0$.

Structures exhibiting significant disorder and/or thermal motion have lower resolution. This is evident due to the impact of thermal motion on the structure factors, and it was observed in the results deposited in the Protein Data Bank [23–25] that higher B-factors are associated with lower-resolution crystal structures [6].

3. Scaling

Given that B-factors may vary from one structure to another for reasons different from local mobility/flexibility, it is mandatory to rescale them when comparing different structures. In this section, we will enumerate various scaling procedures that have been used in structural biology.

There are two approaches: one that takes into account both the average B-factor of the structure and its standard deviations and another that just considers the average value.

A common rescaling is a Z-transformation to zero mean and unit variance. The B-factor B_i of the ith atom is modified according to

$$Br_i = \frac{B_i - B_{ave}}{B_{std}} \tag{5}$$

where B_{ave} and B_{std} are the average B-factor and its standard deviation computed on all the n atoms according to

$$B_{ave} = \frac{\sum_{i=1}^{n} B_i}{n} \tag{6}$$

$$B_{std} = \sqrt{\frac{\sum_{i=1}^{n}(B_i - B_{ave})^2}{n-1}} \tag{7}$$

Given that few protein atoms may be outliers, since their B-factors are considerably larger or smaller than others, they can be removed, and the rescaled B-factors are computed according to

$$Br_i = \frac{B_i - B_{ave,out}}{B_{std,out}} \tag{8}$$

where $B_{ave,out}$ and $B_{std,out}$ are the average B-factor and its standard deviation computed on all n atoms with the exclusion of the outliers. The definition of an outlier in statistical distributions is relatively nebulous [26]. The following strategy was suggested in the B-factor analysis [27]. The M_i values were computed for each B_i, and the latter one was considered an outlier if $M_i > 3.5$:

$$M_i = \frac{0.674 \cdot |B_i - B_{med}|}{MAD} \tag{9}$$

where B_{med} is the median of the B-factors, and MAD is computed according to

$$MAD = median\left[\sqrt{(B_i - B_{med})^2}\right] \tag{10}$$

A slightly different method has been proposed [28]. The rescaled B-factors are computed in two different ways, depending on the values of MAD, according to

$$Br_i = \frac{B_i - B_{med}}{\frac{1.235}{n}\sum_{i=1}^{n}(B_i - B_{med})^2} \quad \text{if } MAD = 0 \tag{11a}$$

$$Br_i = \frac{B_i - B_{med}}{1.486 \cdot MAD} \quad \text{if } MAD \neq 0 \tag{11b}$$

However, in the experience of the writer, MAD is usually very close to zero.

Additionally, alternative B-factor rescaling approaches that do not rely on Z-transformation have also been suggested. They consider only the average B-factor of the structure and not its standard deviation.

Actually, the first rescaling procedure, proposed by Karplus and Schulz back nearly four decades ago [29], was defined as

$$Br_i = \frac{B_i + P}{B_{ave} + P} \tag{12}$$

where P is an arbitrary parameter, the value of which is determined empirically—and iteratively—in such a way that the following quantity

$$\sqrt{\frac{\sum_{i=1}^{n}(Br_{ave} - Br_i)^2}{n}} \tag{13}$$

where Br_{ave} is the average value of the rescaled B-factors, is equal to 0.3.

Another B-factor rescaling that can be found in the scientific literature [18] is simply the following:

$$Br_i = \frac{B_i}{B_{ave}} \tag{14}$$

Interestingly, rescaled B-factors can assume negative or positive values with Equations (5), (8) and (11), while they assume values always larger than 0 with Equations (12) and (14).

Other rescaling procedures may be designed and used. For example, Schneider rescaled B-factors in the linear range from 1 to 100 [30], and it is possible that alternative methods have been used sporadically. In this manuscript, however, only the most common rescaling techniques are examined and discussed.

There are a few important points to note about the rescaling techniques discussed above.

First, the problem of conformational disorder must be addressed [31,32]. Frequently, particularly in protein crystal structures at high resolution, conformational disorder (alternative conformation) is observed in some atoms, which exhibit two or more equilibrium positions, generally in close proximity to each other. Each of these positions is associated

with an occupancy value larger than 0 and smaller than 1, and the sum of all the occupancy values of an atom is equal to 1—unless there is evidence that some atoms are partially absent, for example, because of radiation damage [33–35]. In the scientific literature dealing with B-factors analyses, the problem of conformational disorder is generally not described explicitly, and it is unclear if all the conformations are handled individually or if only one of them is considered, usually the first or the one with the highest occupancy.

The second problem is the presence of heteroatoms. All protein crystal structures contain water molecules that hydrate the protein surface and also a few water molecules that are buried in the protein core [36,37]. In addition, the structures may also include other kinds of heteroatoms, such as metal cations, cofactors, or inhibitors that selectively interact with proteins and even organic or inorganic atoms and molecules, the occurrence of which is coincidental and may be attributed to their presence in the crystallization cocktails. It is, in general, unclear if the B-factors of these heteroatoms are considered in the computation of B_{ave}, B_{std}, and B_{med} and in the individuation of the outliers. In general, it seems that they are disregarded, though this approach is questionable since the presence and the crystallographic refinement of the heteroatoms clearly influence the B-factors of protein atoms.

4. Available Computational Resources

There is a shortage of freely available software to analyze B-factors. Nevertheless, there are a few interesting tools.

The Bandit server has recently made significant progress in enabling structural biologists to perform B-factor rescaling (https://bandit.uni-mainz.de; the source code is also freely downloadable; accessed on 26 April 2024) [38]. Equations (5), (6), (11) and (12) can be used to obtain the rescaled B-factor; however, Equation (14) is missing. Two protein crystal structures can be either uploaded or downloaded from the Protein Data Bank, and their raw or rescaled B-factors compared. A simple molecular graphics window allows one to easily identify interesting regions of the structure through a predefined color code.

This will undoubtedly assist structural biologists in optimizing the analyses of their structures.

Another program, B FITTER [39], calculates the average B-factor of each residue by considering all its non-hydrogen atoms and generates an output file with a ranked classification of the residues according to their average B-factor.

5. Case Studies

The paper does not aim to conduct a statistical survey on B-factor scaling methods in the Protein Data Bank. Thus, there is no need to extract representative subsets of protein structures. Rather, the focus is on selecting a few examples. We analyzed two medium-sized structures, one refined at high resolution and the other at low resolution. The crystal structure of human promyeloperoxidase (PDB 5mfa [40]) refined at high resolution (1.20 Å) in the C2 space group with one monomeric chain per asymmetric unit is shown in Figure 1-left. It contains 697 residues, the first 108 and the last missing from the electron density map, and 40 of them are observed in two different conformations. It also contains 766 water molecules and 244 heteroatoms different from water (heme, glucopyranose, and others), four of which are observed in two different conformations. The structure was refined using REFMAC [41,42] and phenix.refine [43]. In the deposited PDB file, phenix.refine is listed in the REMARK 3 record, indicating that at least in the last refinement round, phenix.refine was used. Because of the high resolution, the authors of this structure used an anisotropic B-factor model in the refinement [44]. For anisotropic refinement, six parameters per atom are needed and phenix.refine applies simple similarity restraints to enforce the physical correctness of the refined ADPs [44].

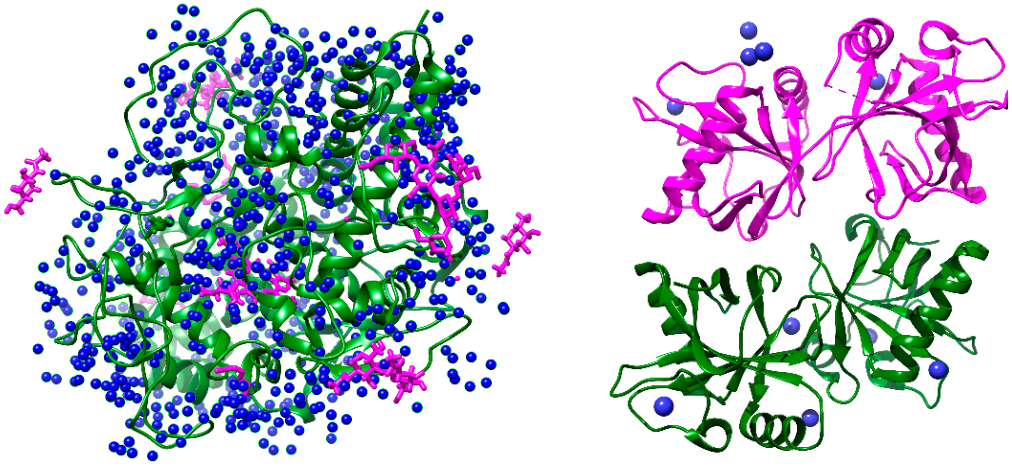


Figure 1. (**left**) Example of a protein crystal structure refined at high resolution (1.20 Å; PDB 5mfa) that contains both numerous water molecules (blue spheres) and numerous heteroatoms different from water (heme, glucopuranose, and others; magenta sticks). (**right**) Example of a protein crystal structure refined at low resolution (2.85 Å; PDB 6q5y) that contains two chains (A and B), few water molecules (blue spheres), and no other heteroatoms.

The crystal structure of the SPOC domain of human PHD finger protein 3 in complex with RNA polymerase II CTD diheptapeptide (PDB 6q5y [45]) was refined at low resolution (2.85 Å) in the $P3_2$ space group with two homodimers per asymmetric unit (Figure 1-right). It contains 676 residues, 62 of which are missing from the electron density map, only 10 water molecules, and no other heteroatoms. No evidence of conformational disorder was observed for any of the atoms, which is commonly anticipated at low resolution. The structure was again refined using REFMAC [41,42] and phenix.refine [43], whereas in this case, the structure used for deposition was refined with REFMAC using isotropic B-factors.

These two structures were used for three types of analyses: (i) the consequences of considering only the protein atoms or also the heteroatoms (solvent and eventual other atoms and molecules) in rescaling; (ii) the comparison between alternative scaling methods; and (iii) the impact of various approaches for handling conformational disorder.

6. Protein and Hetero-Atoms

It is first important to check how the choice of atoms used for the rescaling process (protein atoms, water atoms, and/or other heteroatoms) affects the rescaled B-factors.

In the high-resolution crystal structure, the average B-factor (B_{ave}, Equation (6)) increases from 17.7 Å^2 when only the protein atoms are considered to 19.7 Å^2 when both the protein atoms and the water atoms are considered and to 20.6 Å^2 when also the remaining heteroatoms are considered. The standard deviations of the B-factors (B_{std}, Equation (7)) also increase in the same order from 8.6 Å^2 to 10.4 Å^2 and to 11.8 Å^2—these values were computed by considering only the first conformer in the case of conformational disorder.

In contrast, the average B-factor and its standard deviation in the low-resolution structure remain largely unchanged regardless of whether the water molecules are included or excluded. This is because the low-resolution structure contains few water molecules.

Figure 2 shows the comparison between the protein atom's B-factors of the high-resolution structure rescaled by considering only the protein atoms (*Atom*), by considering both the protein atoms and the water atoms (*Atom + Water*), and by considering the other heteroatoms (*Atom + Water + Hetero*). Minimal differences exist, though protein-rescaled B-factors are slightly larger when water and heteroatoms are taken into consideration. Since the standard deviation of the B-factors is lower when calculated using solely protein atoms, it is anticipated that the *Br* values are dispersed throughout a somewhat wider range of values. However, all rescaled B-factors are perfectly correlated (Pearson correlation coefficient = 1.000), and mean absolute differences are very small (0.201 ± 0.002 when comparing *Atom* with *Atom + Water*,

0.258 ± 0.004 when comparing *Atom* with *Atom* + *Water* + *Hetero*, and 0.066 ± 0.001 when comparing *Atom* + *Water* with *Atom* + *Water* + *Hetero*).

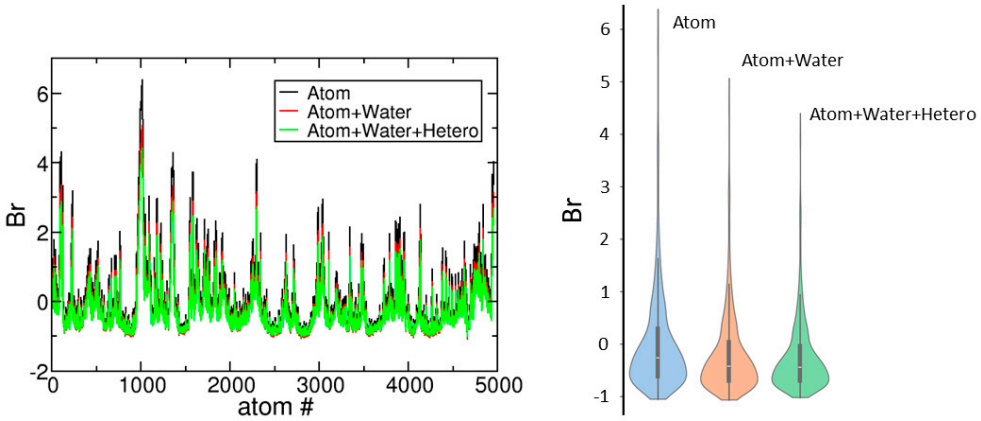


Figure 2. Comparison of the protein atomic B-factor of structure 5mfa rescaled by considering only the protein atoms (*Atom*), by considering both water atoms and protein atoms (*Atom* + *Water*), and by including the other heteroatoms (*Atom* + *Water* + *Hetero*).

In the low-resolution structures, which contain few water molecules, the mean absolute differences between rescaled B-factors are very small (0.00326 ± 0.00002), regardless of whether the water molecules are included or excluded.

7. Alternative Rescaling Methods

In addition to considering the inclusion or exclusion of atom types (protein atoms, water atoms, and/or other heteroatoms), it is important to evaluate the various scaling methods, referred herein after as *Z-scores* (Equation (5)), *Outliers* (Equation (8)), *Karplus* (Equation (12)), and *Ratio* (Equation (14)). Figure 3 shows the distributions of the rescaled B-factors of all protein atoms in both the high- and low-resolution structure (when rescaling is performed with all protein and heteroatoms and when only the first conformer is retained in the case of conformational disorder).

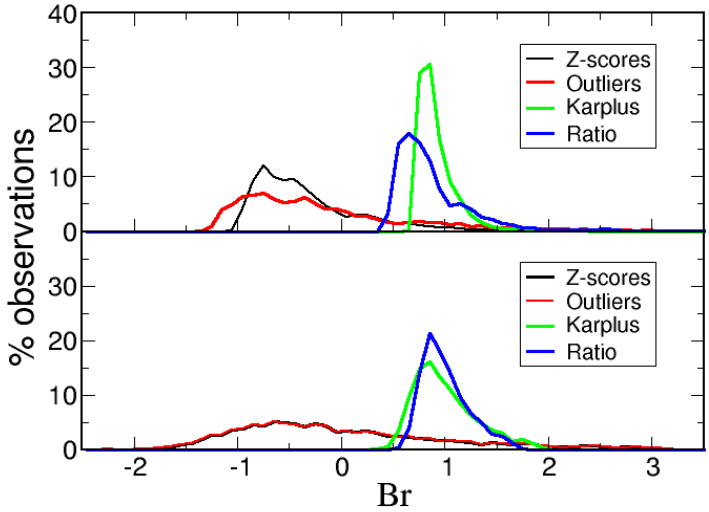


Figure 3. Distributions of the rescaled B-factors (Br) of protein atoms in structure 5mfa (**top**) and 6q5y (**bottom**) (rescaling performed with all protein and heteroatoms by considering only the first conformer in case of conformational disorder).

Clearly, there are two separate ways of rescaling. Rescaled B-factors produced using *Z-scores* or *Outliers* exhibit similarities to each other and distinctions from those produced using *Karplus* or *Ratio*, which, in turn, share similarities with each other.

As it is apparent from their mathematical definition, B-factors rescaled with *Z-scores* or *Outliers* may assume both negative and positive values, while those produced with *Karplus* or *Ratio* have only positive values. Furthermore, the variability of the B-factor, when rescaled using *Z-scores* or *Outliers*, is considerably higher than that observed when B-factors are rescaled using *Karplus* or *Ratio*.

Furthermore, there are differences between the high- and low-resolution structures. In the low-resolution structure (lower part of Figure 3), there is minimal disparity between B-factors rescaled with *Z-scores* and *Outliers*. This is dependent upon the fact that the proportion of B-factors classified as outliers and subsequently removed is quite small (0.4%). In contrast, the high-resolution structure (upper part of Figure 3) exhibits a significantly higher number of outliers (10%). As a consequence, the rescaled B-factors obtained using *Z-scores* and *Outliers* are slightly different.

Analogously, the B-factors rescaled with *Karplus* and *Ratio* exhibit similar distributions in the low-resolution structure (lower part of Figure 3), and there are more discrepancies in the high-resolution structure (upper half of Figure 3).

Although several rescaling approaches exhibit substantial discrepancies, they are all strongly correlated. Linear regressions performed between each pair of rescaling methods (Table 1) indicate that the Pearson correlation coefficient is systematically equal to one. Nevertheless, the intercepts deviate from zero, and the slopes deviate from one, except for the low-resolution structure B-factors that have been rescaled with *Z-scores* and *Outliers*, which are almost identical—this similarity is due to the limited number of outliers, as previously noticed.

Table 1. Linear regressions between B-factors rescaled with different techniques. For each comparison between two techniques, the following information is shown: the Pearson correlation coefficient (pcc), the intercept, and the slope.

Rescaling Methods	5mfa			6q5y		
	Pcc	***Intercept***	***Slope***	***Pcc***	***Intercept***	***Slope***
Outliers versus *Z-scores*	1.000	0.48	1.76	1.000	0.01	1.02
Karplus versus *Z-scores*	1.000	1.00	0.30	1.000	1.00	0.30
Ratio versus *Z-scores*	1.000	1.05	0.60	1.000	1.00	0.23
Karplus versus *Outliers*	1.000	0.92	0.17	1.000	1.00	0.29
Ratio versus *Outliers*	1.000	0.89	0.34	1.000	1.00	0.22
Ratio versus *Karplus*	1.000	−0.94	1.99	1.000	0.25	0.76

8. Conformational Disorder

Ultimately, it is necessary to validate the influence of rescaled B-factors resulting from various approaches to handling conformational disorder. Three methodologies were evaluated: (i) using all conformations (*all conformers*), (ii) utilizing only the first conformer (*first conformer*), and (iii) multiplying the raw-B-factors by the occupancies (*weighted*). (Rescaling was performed by using Equation (5) with all proteins and heteroatoms).

Obviously, in the low-resolution structure, no differences are expected, given the absence of conformational disorder (lower half of Figure 4). However, a substantial fraction of atoms is conformationally disordered in the high-resolution structure. Nevertheless, only very minor discrepancies were observed amongst *all conformers*, *first conformer*, and the *weighted* methodologies of handling disorder (upper half of Figure 4).

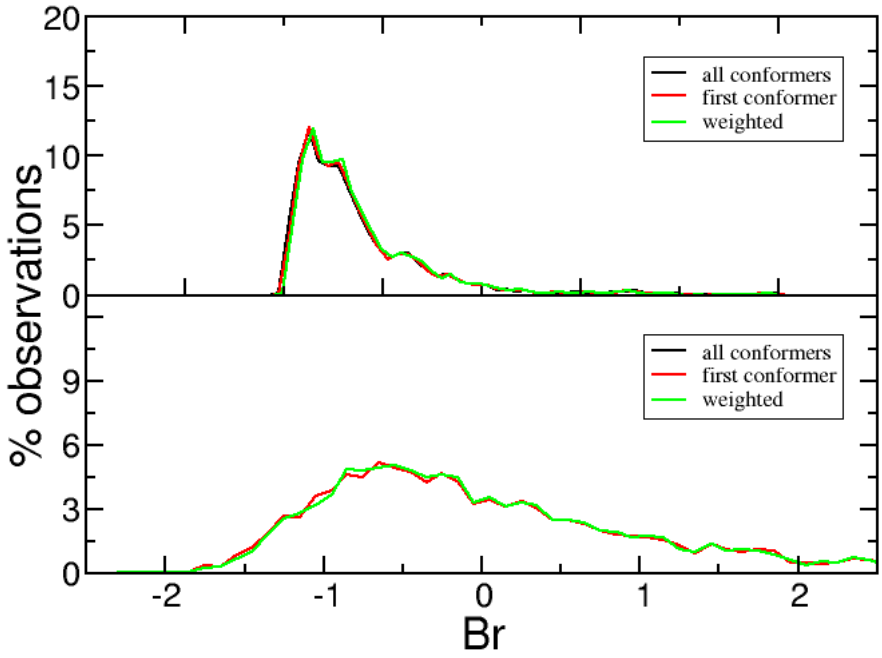


Figure 4. Distributions of the rescaled B-factors of protein atoms in structure 5mfa (**top**) and 6q5y (**down**) (rescaling performed with Equation (4) by using all protein and heteroatoms).

In both the high and low-resolution structures, all procedures for handling conformational disorder correlate very well with intercepts close to zero and slopes close to one (Table 2).

Table 2. Linear correlation between B-factors rescaled with Z-scores by using all proteins and heteroatoms. For each comparison between two methods of handling conformational disorder, the following information is shown: the Pearson correlation coefficient (pcc), the intercept, and the slope.

Pair of Variables	**5mfa**			**6q5y**		
	Pcc	***Intercept***	***Slope***	***Pcc***	***Intercept***	***Slope***
First conformer versus *All conformers*	1.000	0.00	0.98	1.000	0.00	1.00
Weighted versus *All conformers*	1.000	0.08	0.98	1.000	0.01	0.99
Weighted versus *First conformer*	1.000	0.08	1.00	1.000	0.01	0.99

9. Anisotropy

When the crystallographic resolution is sufficiently high, it is possible to perform anisotropic refinements [31,46,47], where the mean square displacement of the atom from its equilibrium position is defined as

$$\left\langle u^2 \right\rangle = \frac{U_{11} + U_{22} + U_{33}}{3} \tag{15}$$

where the three terms at the numerator are the diagonal elements of the anisotropic B-factor, which is the symmetrical tensor $\boldsymbol{U}$

$$\boldsymbol{U} = \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ U_{21} & U_{22} & U_{23} \\ U_{31} & U_{32} & U_{33} \end{bmatrix} \tag{16}$$

which is the variance–covariance matrix that describes the positional dispersion of the atom. Since there are six variables to be determined and refined (U_{11}, U_{22}, U_{33}, $U_{12} = U_{21}$, $U_{13} = U_{31}$, and $U_{23} = U_{32}$), it is necessary that the number of diffraction data is sufficiently

high—and this is why it is possible only at sufficiently high resolution. The Hamilton test, as defined by Merrit [48], provides an objective criterion for determining whether an isotropic or anisotropic B-factor model should be used in the refining process. If possible, the anisotropic refinement is preferable since it allows a better description of the atom's positional spread [44].

Using Equation (17), the diagonal elements of the symmetrical tensor U can be used to calculate an equivalent to the isotropic B-factor (B_{eq}), which can then be used for rescaling and allowing comparisons of structures refined isotropically and anisotropically.

$$B_{eq} = 8\pi^2 \frac{U_{11} + U_{22} + U_{33}}{3} \quad (17)$$

No studies have been devoted to analyzing in a systematic way the reproducibility of the anisotropic B-factors $\boldsymbol{U}$ and how they can vary if resolution or data collection conditions change (space group, ionic strength, pH, etc.) and with different refinement protocols (software, restraint weights, etc.).

A different type of anisotropic refinements is based on subdividing the protein into two or more moieties and refining each of them anisotropically as a rigid body [49,50]. This is known as TLS refinement since each moiety can vibrate along a straight line (T—translation), vibrate along an arc (L—libration), and vibrate along a helical path (S—screw). In this way, it is possible to determine and refine anisotropic B-factors for each atom.

To the best of our knowledge, no studies have been performed to ascertain the degree of reproducibility of TLS anisotropic B-factors. However, it is imperative that the anisotropic B-factors determined by TLS refinements are not compared to those obtained with non-TLS refinements because they obviously provide alternative descriptions of molecular flexibility. TLS refinements provide information on collective rigid-body motions for groups of atoms and not on individual atoms.

10. Conclusions

When B-factors are rescaled, the ensemble of atoms used for rescaling is often not explicitly described in the scientific literature. The ensemble may be limited to protein atoms only, include both protein atoms and the water atoms, or consider all atoms, including protein, water, and all other heteroatoms. However, the selection of atoms used for rescaling B-factors seems to be of marginal importance.

The number of water molecules observed on the surface of protein structures is directly proportional to the resolution of the structure and is also influenced to some extent by other structural features, including average B factor of the protein atoms, percentage of solvent in the crystal, R factor, grand average of hydropathy of the protein(s) in the asymmetric unit, number of heteroatoms that are not water molecules, and average solvent-accessible surface area of the amino acid residues [51]. A decrease in the number of water molecules is noticed as the resolution decreases. A loss in resolution by 0.5 Å results in an estimated 35% reduction in the number of water molecules [51].

Consequently, it is reasonable to foresee that at low resolution, B-factor rescaling might be little influenced by the selection of the atoms, the B-factors of which are used since there are few water molecules. However, even at very high resolution, like in the example of human promyeloperoxidase reported above (1.20 Å [40]), the selection of the type of atoms has a modest impact. In this structure, even if about 15% of the atoms are water atoms and about 5% of the atoms are heteroatoms belonging to non-water molecules, the rescaled B-factors are rather independent of the selection of the type of atoms used for rescaling.

Although this might appear to be quite surprising, one must consider that the flexibility of the water molecules at the protein surface is not comparable to that of liquid water, though their interactions with the protein surface have been observed to be not perfectly rigid [52]. Similar arguments are likely to be generalizable to other heteroatoms, different from water. B-factors of water and of other heteroatoms at the protein surface are thus

expected to be larger than those of the protein atoms but not extremely different from those of protein atoms.

It is likely that any rescaling strategy should anyway be based on all types of atoms (protein, water, and other hetero atoms) since the B-factor values of the protein atoms are influenced by the presence of the rest of the atoms.

The presence of conformational disorder has little impact on rescaled B-factors. This is absolutely predictable in low-resolution structures, where the limited amount of diffraction data prevents the characterization of conformational disorder. However, this is also observed at high resolution, where conformational disorder can be characterized for several atoms/residues.

Three alternative rescaling methods have been used. In one case, as is often done in structural bioinformatics, only one of the equilibrium positions was considered. In another case, all conformers were considered, and in the third case, B-factors were multiplied by the occupancy values. Nearly little variation in rescaled B-factors was observed. This is due to the fact that the B-factors of atoms that have two (or more) equilibrium positions are likely to be similar to the B-factors that are observed in only one equilibrium position. Consequently, all the rescaling methods produce nearly the same results.

On the contrary, different rescaling procedures can produce very different results. There are, basically, two types of rescaling. One is based on Z-transformations, using all B-factors [Equation (5)] or only those that are not considered outliers [Equation (7)], and the other is based on the ratio between the B-factor and the average B-factor [Equations (11), (12) and (14)]. Consequently, the B-factor rescaled according to these two methods are drastically different and cannot be directly compared. Both their values and distribution differ considerably, with the first being more widely scattered.

Rescaling according to Z-transformations seems preferable since it provides evaluations of the statistical significance of differences in *Br* values. However, these statistical significances require the normality of the *Br* distribution, and, in reality, deviations from normality can occur, and, as a consequence, it is possible to overinterpret the data.

The use of more than a single rescaling procedure is not a solution to this problem since B-factors rescaled with alternative methods are strongly correlated—and this is expected, given their mathematical definition.

Additionally, it is worth noting that this article exclusively examines two case studies, namely a high-resolution refined protein structure and a low-resolution refined protein structure. On large sets of well-controlled structures, however, not even a statistical comparison of alternative rescaling procedures is likely to yield sound conclusions regarding which method is systematically preferable.

It is thus just a matter of personal preference. One can adopt the preferred rescaling procedure, provided it is explicitly described in such a way as to ensure reproducibility.

It can be expected that in the future, it will become possible to improve the quality of B-factors, both by achieving higher crystallographic resolutions and by improving refinement methods—although with the risk that extremely user-friendly software may be associated with a deterioration in the theoretical knowledge of younger crystallographers. Consequently, B-factor rescaling will have to adapt to new experimental data. Of great importance, finally, is the analysis of new types of experimental data, particularly the equivalents of B-factors determined by means of Cryo-EM, which, to our knowledge, have not been analyzed in detail. These new experimental data may need new rescaling methods, and it may be necessary to develop methods to compare B-factors determined by different experimental techniques.

Author Contributions: Conceptualization, O.C.; methodology, G.M., K.D.-C. and O.C.; software, G.M. and O.C.; formal Analysis, G.M., K.D.-C. and O.C.; investigation, G.M., K.D.-C. and O.C.; resources, G.M., K.D.-C. and O.C.; data curation, G.M. and O.C.; writing—original draft preparation, O.C.; writing—review and editing, G.M., K.D.-C. and O.C.; visualization, O.C.; supervision, O.C.; project administration, O.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All data are available from the Protein Data Bank www.rcsb.org.

Acknowledgments: A. A. B. Stradella is gratefully acknowledged for constant support and helpful discussions. OC acknowledges support from the Ministero dell'Università e della Ricerca (MUR) and the University of Pavia through the program "Dipartimenti di Eccellenza 2023–2027". Michal Nagy is acknowledged for help in preparing figures.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Trueblood, K.N.; Bürgi, H.B.; Burzlaff, H.; Dunitz, J.D.; Gramaccioli, C.M.; Schulz, H.H.; Shmueli, U.; Abrahams, S.C. Atomic Dispacement Parameter Nomenclature. Report of a Subcommittee on Atomic Displacement Parameter Nomenclature. *Acta Crystallogr.* **1996**, *A52*, 770–781. [CrossRef]
2. Giacovazzo, C.; Monaco, H.L.; Viterbo, D.; Scordari, F.; Gilli, G.; Zanotti, G.; Catti, M.; Paufler, P. *Fundamentals of Crystallography*; Oxford University Press: New York, NY, USA, 2002.
3. Debye, P. Interferenz von Röntgenstrahlen und Wärmebewegung. *Ann. Phys.* **1913**, *348*, 49–92. [CrossRef]
4. Waller, I. Zur Frage der Einwirkung der Wärmebewegung auf die Interferenz von Röntgenstrahlen. *Z. Phys. A* **1923**, *17*, 398–408. [CrossRef]
5. Sun, Z.; Liu, Q.; Qu, G.; Feng, Y.; Reetz, M.T. Utility of B-factors in protein science: Interpreting rigidity, flexibility, and internal motion and engineering. *Chem. Rev.* **2019**, *119*, 1626–1665. [CrossRef] [PubMed]
6. Carugo, O. Uses and abuses of the atomic displacement parameters in structural biology. *Meth. Mol. Biol.* **2022**, *2449*, 281–298.
7. Carugo, O. Atomic displacement parameters in structural biology. *Amino Acids* **2018**, *50*, 775–786. [CrossRef] [PubMed]
8. Pearce, N.M.; Gros, P. A method for intuitively extracting macromolecular dynamics from structural disorder. *Nat. Commun.* **2021**, *12*, 5493. [CrossRef] [PubMed]
9. Xu, L.; Han, F.; Dong, Z.; Wei, Z. Engineering Improves Enzymatic Synthesis of L-Tryptophan by Tryptophan Synthase from *Escherichia coli*. *Microorganisms* **2020**, *8*, 519. [CrossRef]
10. Xu, J.; Cen, Y.; Singh, W.; Fan, J.; Wu, L.; Lin, X.; Zhou, J.; Huang, M.; Reetz, M.T.; Wu, Q. Stereodivergent Protein Engineering of a Lipase to Access All Possible Stereoisomers of Chiral Esters with Two Stereocenters. *J. Am. Chem. Soc.* **2019**, *141*, 7934–7945. [CrossRef]
11. Tang, H.; Shi, K.; Shi, C.; Aihara, H.; Zhang, J.; Du, G. Enhancing subtilisin thermostability through a modified normalized B-factor analysis and loop-grafting strategy. *J. Biol. Chem.* **2019**, *294*, 18398–18407. [CrossRef]
12. Blum, T.B.; Housset, D.; Clabbers, M.T.; van Genderen, E.; Bacia-Verloop, M.; Zander, U.; McCarthy, A.A.; Schoehn, G.; Ling, W.L.; Abrahams, J.P. Statistically correcting dynamical electron scattering improves the refinement of protein nanocrystals, including charge refinement of coordinated metals. *Acta Crystallogr. Sect. D Struct. Biol.* **2021**, *77*, 75–85. [CrossRef] [PubMed]
13. Espinosa, Y.R.; Alvarez, H.A.; Howard, E.I.; Carlevaro, C.M. Molecular dynamics simulation of the heart type fatty acid binding protein in a crystal environment. *J. Biomol. Struct. Dyn.* **2021**, *39*, 3459–3468. [CrossRef] [PubMed]
14. Rodríguez, F.S.; Simpkin, A.J.; Davies, O.R.; Keegan, R.M.; Rigden, D.J. Helical ensembles outperform ideal helices in molecular replacement. *Acta Crystallogr. Sect. D Struct. Biol.* **2020**, *76*, 962–970. [CrossRef]
15. Bae, J.-E.; Kim, I.J.; Xu, Y.; Nam, K.H. Structural Flexibility of Peripheral Loops and Extended C-terminal Domain of Short Length Substrate Binding Protein from *Rhodothermus marinus*. *Protein J.* **2021**, *40*, 184–191. [CrossRef] [PubMed]
16. Johnson, T.W.; Gallego, R.A.; Brooun, A.; Gehlhaar, D.; A McTigue, M. Reviving B-Factors: Retrospective Normalized B-Factor Analysis of c-ros Oncogene 1 Receptor Tyrosine Kinase and Anaplastic Lymphoma Kinase L1196M with Crizotinib and Lorlatinib. *ACS Med. Chem.* **2018**, *9*, 878–883. [CrossRef]
17. Borek, D.; Minor, W.; Otwinowski, Z. Measurement errors and their consequences in protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2003**, *59*, 2031–2038. [CrossRef]
18. Ramos, N.G.; Sarmanho, G.F.; Ribeiro, F.d.S.; de Souza, V.; Lima, L.M.T. The reproducible normality of the crystallographic B-factor. *Anal. Biochem.* **2022**, *645*, 114594. [CrossRef]
19. Tronrud, D.E. Knowledge-Based B-Factor Restraints for the Refinement of Proteins. *J. Appl. Crystallogr.* **1996**, *29*, 100–104. [CrossRef]
20. Thorn, A.; Dittrich, B.; Sheldrick, G.M. Enhanced rigid-bond restraints. *Acta Crystallogr.* **2012**, *A68*, 448–451. [CrossRef]
21. Wall, M.E.; Wolff, A.M.; Fraser, J.S. Bringing diffuse X-ray scattering into focus. *Curr. Opin. Struct. Biol.* **2018**, *50*, 109–116. [CrossRef]
22. de Klijn, T.; Schreurs, A.M.M.; Kroon-Batenburg, L.M.J. Rigid-body motion is the main source of diffuse scattering in protein crystallography. *IUCrJ* **2019**, *6*, 277–289. [CrossRef] [PubMed]

23. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542. [CrossRef] [PubMed]
24. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]
25. wwPDB Consortium. Protein Data Bank: The single global archive fro 3D macromolecular structural data. *Nucleic Acids Res.* **2019**, *47*, D520–D528. [CrossRef] [PubMed]
26. Iglewicz, B.; Hoaglin, D.C. *How to Detect and Handle Outliers*; ASQ Press: Milwaukee, WI, USA, 1993.
27. Smith, D.K.; Radivojac, P.; Obradovic, Z.; Dunker, A.K.; Zhu, G. Improved amino acid flexibility parameters. *Protein Sci.* **2003**, *12*, 1060–1072. [CrossRef] [PubMed]
28. Li, X.; Anderson, M.; Collin, D.; Muegge, I.; Wan, J.; Brennan, D.; Kugler, S.; Terenzio, D.; Kennedy, C.; Lin, S.; et al. Structural studies unravel the active conformation of apo RORγt nuclear receptor and a common inverse agonism of two diverse classes of RORγt inhibitors. *J. Biol. Chem.* **2017**, *292*, 11618–11630. [CrossRef] [PubMed]
29. Karplus, P.A.; Schulz, G.E. Preiction of chain flexibility in proteins. *Natuwissenschaften* **1985**, *72*, 212–213. [CrossRef]
30. Schneider, B.; Gelly, J.-C.; de Brevern, A.; Černý, J. Local dynamics of proteins and DNA evaluated from crystallographic B factors. *Acta Crystallogr.* **2014**, *D70*, 2413–2419. [CrossRef]
31. Dauter, Z.; Lamzin, V.S.; Wilson, K.S. The benefits of atomic resolution. *Curr. Opin. Struct. Biol.* **1997**, *7*, 681–688. [CrossRef]
32. Ševčík, J.; Lamzin, V.S.; Dauter, Z.; Wilson, K.S. Atomic resolution data reveal flexibility in the structure of RNase Sa. *Acta Crystallogr.* **2002**, *D58*, 1307–1313. [CrossRef]
33. Garman, E.F.; Owen, R.L. Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr.* **2006**, *D62*, 32–47. [CrossRef] [PubMed]
34. Shelley, K.L.; Dixon, T.P.E.; Brooks-Bartlett, J.C.; Garman, E.F. RABDAM: Quantifying specific radiation damage in individual protein crystal structures. *J. Appl. Crystallogr.* **2018**, *51*, 552–559. [CrossRef] [PubMed]
35. Carugo, O.; Djinovic-Carugo, K. When X-rays modify the protein structure: Radiation damage at work. *Trends Biochem. Sci.* **2005**, *30*, 213–219. [CrossRef] [PubMed]
36. Cerny, J.; Schneider, B.; Biedermannova, L. WatAA: Atlas of protein hydration. Exploring synergies between data mining and ab initio calculations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 17094. [CrossRef] [PubMed]
37. Carugo, O. Anisotropic waters in atomic resolution protein crystal structures. *Int. J. Biol. Macromol.* **2019**, *135*, 940–944. [CrossRef] [PubMed]
38. Barthels, F.; Schirmeister, T.; Kersten, C. BANΔIT: B′-Factor Analysis for Drug Design and Structural Biology. *Mol. Inform.* **2021**, *40*, e2000144. [CrossRef] [PubMed]
39. Reetz, M.T.; Carballeira, J.D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2007**, *2*, 891–903. [CrossRef] [PubMed]
40. Grishkovskaya, I.; Paumann-Page, M.; Tscheliessnig, R.; Stampler, J.; Hofbauer, S.; Soudi, M.; Sevcnikar, B.; Oostenbrink, C.; Furtmüller, P.G.; Djinović-Carugo, K.; et al. Structure of human promyeloperoxidase (proMPO) and the role of the propeptide in processing and maturation. *J. Biol. Chem.* **2017**, *292*, 8244–8261. [CrossRef] [PubMed]
41. Murshudov, G.N.; Vagin, A.A.; Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1997**, *53*, 240–255. [CrossRef]
42. Murshudov, G.N.; Skubák, P.; Lebedev, A.A.; Pannu, N.S.; Steiner, R.A.; Nicholls, R.A.; Winn, M.D.; Long, F.; Vagin, A.A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr.* **2011**, *D67*, 355–367. [CrossRef]
43. Liebschner, D.; Afonine, P.V.; Baker, M.L.; Bunkóczi, G.; Chen, V.B.; Croll, T.; Hintze, B.; Hung, L.W.; Jain, S.; McCoy, A.; et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr.* **2019**, *D75*, 861–877. [CrossRef] [PubMed]
44. Afonine, P.V.; Urzhumtsev, A.; Grosse-Kunstleve, R.W.; Adams, P.D. Atomic Displacement Parameters (ADPs), their parameterization and refinement in PHENIX. *Comput. Crystallogr. Newsl.* **2010**, *1*, 24–31.
45. Appel, L.-M.; Franke, V.; Bruno, M.; Grishkovskaya, I.; Kasiliauskaite, A.; Kaufmann, T.; Schoeberl, U.E.; Puchinger, M.G.; Kostrhon, S.; Ebenwaldner, C.; et al. PHF3 regulates neuronal gene expression through the Pol II CTD reader domain SPOC. *Nat. Commun.* **2021**, *12*, 6078. [CrossRef] [PubMed]
46. Schmidt, A.; Lamzin, V.S. Veni, vidi, vici—Atomic resolution unravelling the mysteries of protein function. *Curr. Opin. Struct. Biol.* **2002**, *12*, 698–703. [CrossRef] [PubMed]
47. Longhi, S.; Czjzek, M.; Lamzin, V.; Nicolas, A.; Cambillau, C. Atomic resolution (1.0 Å) crystal structure of Fusarium solani cutinase: Stereochemical analysis. *J. Mol. Biol.* **1997**, *8*, 730–737. [CrossRef] [PubMed]
48. Merritt, E.A. To B or not to B: A question of resolution? *Acta Crystallogr.* **2012**, *D68*, 468–477. [CrossRef] [PubMed]
49. Schomaker, V.; Trueblood, K.N. On the rigid-body motion of molecules in crystals. *Acta Crystallogr.* **1968**, *B24*, 63–76. [CrossRef]
50. Winn, M.D.; Isupov, M.N.; Murshudov, G.N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr.* **2001**, *D57*, 122–133. [CrossRef] [PubMed]

51. Gnesi, M.; Carugo, O. How many water molecules are detected in X-ray protein crystal structures? *J. Appl. Crystallogr.* **2017**, *50*, 96–101. [CrossRef]
52. Carugo, O. Mobility of water and of protein atoms at the protein-water interface, monitored by anisotropic atomic displacement parameters, are largely uncorrelated. *Amino Acids* **2020**, *52*, 435–443. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.