

Statistical analysis of macromolecular B values

R. Masmaliyeva^{1*}, G.N. Murshudov²

¹ *Institute of Molecular Biology and Biotechnologies ANAS, Baku, Azerbaijan*

² *MRC Laboratory of Molecular Biology, Cambridge, UK*

* *e-mail: rmasmaliyeva@gmail.com*

Key words: bayes statistics, validation, atom displacement parameters, refinement

Motivation and Aim: The parameters of macromolecular atomic models consist of positional, occupancy and temperature factors – B values for each atom. B values reflect the combined information about uncertainties of atoms, their mobility as well the resolution and noise in the experimental scattering data. There have been number of research articles directed towards the validation of positional parameters of atoms [1]. However, there are only few research papers covering the validation of atomic B values. Analysis of B value distributions as well as their differences for neighbouring atoms can shed light into the validity of atomic models and even may show locations of misinterpretation of the experimental data. Since B values are proportional to the variances of positional parameters, their distribution should obey the distribution of variances. In Bayesian statistics it is common to use the inverse gamma distribution as the conjugate prior for the variance parameters of normal distribution. Therefore we expect that the probability distribution of B values can be described as a shifted inverse gamma distribution. We add shifts to the inverse gamma distribution to account for possible mishandling of the experimental data by under- or over-blurring.

Methods and Algorithms: The inverse gamma distribution is a continuous univariate distribution. Parameters are: shape parameter (α) and scale parameter (β) and for shifted inverse gamma distribution there is an additional parameter – B_0 that may be equal to the minimum B value. We used Maximum likelihood estimation to evaluate the parameters of the inverse gamma distribution for each macromolecule from the PDB. For optimization of the likelihood function we used the modified Newton-Raphson method with Fisher information matrix.

Results: Our analyses confirmed that for most of the PDB [2] entries the inverse gamma distribution describes well the distribution of temperature factors. In this work analysis of differences between neighbouring B values in the PDB entries will also be presented.

Conclusion: Estimating distributions of temperature factors of protein molecules can be used as a validation tool for B values. We also plan to use the derived probability distribution of B values for designing new restraints for refinement thus increase reliability of derived parameters. Although for parameter estimation we use high-resolution macromolecular structures we expect that these restraints will be useful for medium and low-resolution structure refinement.

Acknowledgements: Supported by the Presidium of Azerbaijan National Academy of Sciences, grant of decree No. 5/9 dated on 15.03.2017.

References

1. Read R.J., Adams P.D., Arendall W.B. et al. (2011) A New Generation of Crystallographic Validation Tools for the Protein Data Bank. *Structure*. 19:1395-1412.
2. Berman H.M., Battistuz T., Bhat T.N. et al. (2002) The Protein Data Bank. *Acta Crystallogr*. 58:899-907.