World Academy of Science, Engineering and Technology
International Journal of Bioengineering and Life Sciences
Vol:17, No:12, 2023

Electron Density Discrepancy Analysis of Energy Metabolism Coenzymes

Alan Luo, Hunter N. B. Moseley

***Abstract*—**Many macromolecular structure entries in the Protein Data Bank (PDB) have a range of regional (localized) quality issues, be it derived from X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, or other experimental approaches. However, most PDB entries are judged by global quality metrics like R-factor, R-free, and resolution for X-ray crystallography or backbone phi-psi distribution statistics and average restraint violations for NMR. Regional quality is often ignored when PDB entries are re-used for a variety of structurally based analyses. The binding of ligands, especially ligands involved in energy metabolism, is of particular interest in many structurally focused protein studies. Using a regional quality metric that provides chemically interpretable information from electron density maps, a significant number of outliers in regional structural quality was detected across X-ray crystallographic PDB entries for proteins bound to biochemically critical ligands. In this study, a series of analyses was performed to evaluate both specific and general potential factors that could promote these outliers. In particular, these potential factors were the minimum distance to a metal ion, the minimum distance to a crystal contact, and the isotropic atomic b-factor. To evaluate these potential factors, Fisher's exact tests were performed, using regional quality criteria of outlier (top 1%, 2.5%, 5%, or 10%) versus non-outlier compared to a potential factor metric above versus below a certain outlier cutoff. The results revealed a consistent general effect from region-specific normalized b-factors but no specific effect from metal ion contact distances and only a very weak effect from crystal contact distance as compared to the b-factor results. These findings indicate that no single specific potential factor explains a majority of the outlier ligand-bound regions, implying that human error is likely as important as these other factors. Thus, all factors, including human error, should be considered when regions of low structural quality are detected. Also, the downstream re-use of protein structures for studying ligand-bound conformations should screen the regional quality of the binding sites. Doing so prevents misinterpretation due to the presence of structural uncertainty or flaws in regions of interest.

***Keywords*—**Biomacromolecular structure, coenzyme, electron density discrepancy analysis, X-ray crystallography.

I. Introduction

MACROMOLECULAR structure determination (i.e., 3D modeling from experimental data) provides in-depth molecular representations for interpreting structure-function relationships within biological and biomedical research contexts. However, in practice, there are issues with the regional (local) quality of the entity being modeled, which can hinder accurate interpretation. These are evident regardless of the structure determination method being used, such as X-ray crystallography or NMR spectroscopy [1]. Additionally, these issues are known to differ between X-ray crystallography and NMR spectroscopy, suggesting that the modeling approach should be well-considered in advance [2].

The approach considered in this study is X-ray crystallography due primarily to its significance to the history of biological research and its availability in the PDB [3], [4]. The protein structure data used in this study were drawn directly from the PDB. At a basic level, X-ray crystallography works by first acquiring a crystallized form of the molecule in question. The resulting crystals are then illuminated with beams from an X-ray source, creating diffraction patterns from the beam passing through the crystals. These diffraction patterns are then transformed into electron density maps in a process that simultaneously generates a representative molecular model. This process depends on many experimental factors, including the quality and properties of the crystal, the quality of the resulting electron density maps, and necessary expert human interaction in the process. Therefore, global and regional (localized) issues in the resulting molecular structures often arise from this whole X-ray crystallographic structure determination process. These issues can hinder accurate interpretation of structure-function relationships, especially in molecular docking and virtual screening studies [5]-[11]. Also, advanced artificial intelligence tools such as *DeepMind's* AlphaFold2 failed to give reliable accuracy for protein structure modeling in high-throughput docking scenarios [12]. So currently, accurate experimentally-derived structures are required for these types of studies, especially at the site of binding. Therefore, regional structure quality needs to be assessed at these sites.

The authors previously developed the pdb-eda software tool to enable regional structure quality assessment in terms of the number of electrons of regional discrepancy between the electron density and the associated molecular model [13], [14]. This region-specific, volume-based metric is calculated using the Fo-Fc electron density discrepancy map, which represents a comparison of the observed electron density (Fo) to the electron density calculated from the model (Fc), enabling a more precise understanding of regional structural quality in chemically meaningful units of electrons. However, this metric is limited to the resolution of the Fo-Fc electron density map itself.

The authors have used pdb_eda to systematically assess coenzyme-bound regional discrepancy data pulled from the PDB, discovering numerous outliers in the data. Initially, the authors focused on adenosine triphosphate (ATP)-bound regions, but expanded to the six most common coenzymes bound in PDB entries: ATP, adenosine diphosphate (ADP),

Alan Luo is with Carnegie Mellon University, USA (e-mail: alanluo@andrew.cmu.edu).

flavin adenine dinucleotide (FAD), guanosine-5'-diphosphate (GDP), guanosine-5'-triphosphate (GTP), and nicotinamide adenine dinucleotide (NAD). While the existence of outliers could simply be due to failings in the modeling software and/or human error, there are other potential factors. One possibility is the proximity of electron density distorting metal ions or nearby crystal contacts, i.e., contacts across crystal unit boundaries that can represent non-native intermolecular contacts [13], [15], [16]. Another possibility is the inherent structural variability of the molecule, which is indirectly quantified by the molecule's atomic isotropic b-factor, but can also represent "fudge factors" in the structure determination calculations [17]-[19]. It was also unknown at the time of study whether these outliers could exist in other coenzyme-bound data, not just ATP-bound data.

The overall goal of this study is to determine possible major causes of regional outliers that appear in electron discrepancy datasets within biochemical regions of interest. Each of these datasets is defined by what kind of biochemical ligand is bound to the regions of interest in the macromolecular structure, mostly proteins. For example, one dataset has ATP bound to each protein while another dataset has ADP bound. Each ligand was chosen for its biochemical significance as coenzymes or cosubstrates in many metabolic reactions (see Table I). For example, ATP and ADP are both essential coenzymes or cosubstrates in most biosynthetic metabolic pathways. In this study, two sets of analyses were performed to elucidate potential reasons for these regional electron discrepancy outliers. The first set was the b-factor analyses, which evaluated the effects of b-factors on the regional structural quality for the same ligand-bound proteins studied in the ligand-wide analyses. B-factors, also known as the Debye–Waller factor, temperature factor, and average atomic displacement parameter, are generated during the X-ray crystallographic model fitting process and are meant to represent the X-ray scattering caused by thermal motion, but often serves as a general optimizable parameter that also describes the blurring or washing out of electron density around a modeled atom. The second set was the contact distance analyses, which explored the effect of a specific type of contact distance (crystal contact or metal ion contact) on the local structural quality of ligand-bound proteins.

TABLE I
DESCRIPTION OF COENZYMES ANALYZED IN THIS STUDY

Ligand name	Number of relevant PDB entries	Number of binding site regions
ATP	1790	2551
ADP	2983	4425
GDP	1660	2407
GTP	1087	1546
FAD	2226	4433
NAD	1632	4234

II. MATERIALS AND METHODS

A. Data Resources

PDB entries containing the target coenzymes along with their associated electron density maps were downloaded from the PDB on July 22, 2022 for ADP and January 04, 2022 for coenzymes ATP, FAD, GDP, GTP, and NAD. These entries were then filtered to those derived from X-ray crystallographic data and having associated electron density maps available through PDBe [20] (see Supplemental Material).

B. Electron Density Discrepancy Analysis

Using the pdb_eda software [13], [14], an electron density discrepancy analysis (EDDA) was performed on each target coenzyme identified in each PDB entry within a 3.5 angstrom radius of the bound coenzyme region, but involving the volume created by the overlapping 3.5 angstrom radii of all atoms in the coenzyme. This analysis involves calculating the sum of absolute discrepancy of the voxels in the Fo-Fc electron density map within the 3.5 angstrom radius volume of the bound coenzyme region. The resulting sum is then converted into an absolute amount of discrepant electrons (referred to as abs-EDDA) for a better chemical interpretation of the magnitude of discrepancy that is comparable across PDB entries.

A systematic application of abs-EDDA to coenzyme bound regions across the PDB revealed regions with aberrantly high absolute electron discrepancy (see histograms in Fig. 1), i.e., outliers, which motivated the rest of our study. Multiple "outlier" definitions including top 1%, top 2.5%, top 5%, or top 10% were evaluated, as it was unclear which definition would provide the best signal-to-noise ratio for identifying explanatory relationships. Outlier definitions below 1% were not evaluated, due to the conservative nature of Fisher exact tests with highly unbalanced contingency tables, which can lead to a loss of statistical power. Likewise, chi-squared tests were avoided due to the approximation limitations caused by unbalanced contingency tables generated from these "outlier" definitions.

C. B-factor Analyses

Our first goal was to explore whether b-factors, which are atom-centric, could explain the outlier electron discrepancy of ligand-bound regions. However, b-factors values are notoriously specific to the PDB entry and require some form of normalization in order to be comparable across PDB entries [21]. Since only a relative ranking of b-factors was required to build 2x2 contingency tables for Fisher exact tests, a fractional ranking of b-factors in each PDB entry was used, i.e., the rank of b-factors within the PDB entry sorted in descending order. This ranking was divided by the total number of b-factors within the PDB entry. Next, the average fractional b-factor ranking was calculated across all atoms within 3.5 angstroms of the bound ligand. High versus low average fractional b-factor regions were defined as being 0.75 or higher versus below 0.75. Ligand-bound 3.5 angstrom regions were likewise defined as abs-EDDA outlier versus non-outlier to craft a 2x2 contingency table, which was evaluated using a Fisher exact test. The top 1%, top 2.5%, top 5%, or top 10% regional abs-EDDA outlier definitions were used in separate Fisher exact tests.

D. Distance Analyses

The next goal was to explore whether the distance observed from a region of interest to some phenomenon was related to that region's observed absolute local electron discrepancy

(discrepancy). In particular, crystal contacts -- intermolecular contacts due solely to crystal packing -- are known to affect protein-ligand binding [15], and metal ions are known to create regional distortions in electron density maps that affect regional structural quality [13], [22], [23]. Therefore, the distances evaluated were from the ligand atoms to crystal contacts and/or metal atoms.

1. Evaluating Relationship Between Metal Distance and abs-EDDA

The Bio.PDB package from the BioPython suite of packages [24] was used to parse each PDB entry containing a coenzyme residue. For each PDB entry, the distance between each metal atom and each coenzyme residue atom was calculated, and the minimum distance was used as the minimum metal contact distance (minMCD). PDB entries lacking metal atoms were excluded. The list of metal atoms detected are in Supplemental Material. Next, within 3.5 angstrom versus beyond 3.5 angstrom ligand-bound regions were defined as a minMCD less than or equal to 3.5 versus greater than 3.5. Ligand-bound 3.5 angstrom regions were likewise defined as abs-EDDA outlier versus non-outlier to craft a 2x2 contingency table, which was evaluated using a Fisher exact test. The top 1%, top 2.5%, top 5%, or top 10% regional abs-EDDA outlier definitions were used in separate Fisher exact tests.

Separate sets of Fisher exact tests were performed for each ligand individually: ATP, ADP, FAD, GDP, GTP, and NAD. In addition, regional abs-EDDA outliers were separated into highly positive, highly negative, and neutral discrepancy groups based on the EDDA being greater than 0.5 times the abs-EDDA, less than -0.5 times the abs-EDDA, or neither. These three separate groups labeled "positive", "negative", and "neutral" were likewise analyzed in separate Fisher exact tests versus minMCD.

2. Evaluating Relationship Between Crystal Distance and abs-EDDA

The pdb_eda's crystal contacts subcommand was used to systematically calculate the minimum crystal contact distance (minCCD) for each ligand-bound region of interest. The pdb_eda software uses the PyMOL package [25] to calculate neighboring crystal contact coordinates for a given PDB entry's asymmetric unit and then pdb_eda identifies the minCCD to a given list of atom coordinates, i.e., the atoms of a given coenzyme. Pragmatically, a 10-angstrom maximum distance was used for calculating neighboring crystal contact coordinates due to the high memory utilization of PyMOL in this task, which could exceed 1TB of RAM for some PDB entries. Even so, due to the computational expense of identifying minimum crystal contact distances out to 10 angstroms, there was no other choice but to limit the minCCD analyses to ATP-bound proteins only. Next, within 3.5 angstrom versus beyond 3.5 angstrom ligand-bound regions were defined as a minCCD less than or equal to 3.5 versus greater than 3.5. Ligand-bound 3.5 angstrom regions were also sorted into abs-EDDA outlier versus non-outlier to craft a 2x2 contingency table, which was evaluated using a Fisher exact test. The top 1%, top 2.5%, top 5%, or top 10% regional abs-EDDA outlier definitions were used in separate Fisher exact tests.

3. Evaluating Relationship Between Metal or Crystal Distance and abs-EDDA

The minMCD and minCCD distances were effectively combined by taking their minimum to create a min[MCD,CCD] distance. Since the minCCD analyses were limited to ATP-bound proteins only, these analyses were likewise limited. Next, within 3.5 angstrom versus beyond 3.5 angstrom ligand-bound regions were defined as a min[MCD,CCD] less than or equal to 3.5 versus greater than 3.5. Ligand-bound 3.5 angstrom regions were also sorted into abs-EDDA outlier versus non-outlier to craft a 2x2 contingency table, which was evaluated using a Fisher exact test. The top 1%, top 2.5%, top 5%, or top 10% regional abs-EDDA outlier definitions were used in separate Fisher exact tests.

4. Evaluating Linear Relationships

Another goal of the study was to determine whether a statistically significant linear relationship existed between regional electron discrepancy versus minMCD, minCCD, or min[MCD,CCD]. First, outlier ATP-bound regions were filtered into specific groups based on the minimum distance (minD). The "within_3.5" angstrom group, for example, includes ATP regions with minD within 3.5 angstroms of any ATP residue atom. The "within_10" group includes ATP regions with minD being within 3.5 angstroms of any ATP residue atom. The "beyond_10" group contains ATP-bound regions that have contact beyond 10 angstroms. The last "within_10+" group is the union of the "within_10" and "beyond_10" groups, setting the distance in the "beyond_10" group to the max minD across the dataset which was close to 10 angstroms for minD. Next, separate scatterplots were generated for the "within_3.5", "within_10", and "within_10+" angstrom groups limited to the 1%, 2.5%, 5%, and 10% outlier definitions as well as all ATP-bound regions. These graphs plotted the abs-EDDA against minD. Finally, the Pearson correlation coefficient and its corresponding p-value was then calculated to determine the strength of association (linear relationship effect size) between abs-EDDA and minD in all three groups.

III. Results

A. B-Factor Analyses

As shown in Table II, the b-factor analyses revealed a strong effect from the normalized b-factors. All of the aggregate datasets, which included binding sites for all six coenzymes, exhibited diminutive p-values, signifying high statistical significance in each. In other words, the association of outliers with high relative b-factors is extremely unlikely to be due to random chance (i.e., p-value = 1.43×10^{-20}). For individual coenzyme results, see Appendix Table V for the results corresponding to each individual coenzyme.

TABLE II
FISHER EXACT TEST RESULTS ON AVERAGE REGIONAL FRACTIONAL B-FACTORS VS. ABS-EDDA OUTLIERS

abs-EDDA outlier definition	p-value
1%	0.00967
2.5%	7.00e-05
5%	4.42e-10
10%	1.43e-20

B. Distance Analyses

In contrast to the b-factor analysis, the crystal contact and metal contact analyses revealed weak effect to no appreciable effect from nearby crystal contacts and metal contacts. For ATP-bound regions, a trend was shown for nearby crystal contacts as the outlier definition increased from 1% to 10%. Even so, the association between crystal contacts and abs-EDDA is weak.

1. Evaluating Relationship Between Metal Distance and abs-EDDA

Most of the individual coenzyme datasets exhibited statistically insignificant p-values, i.e., p-values greater than 0.05. The exceptions were the FAD positive group with a p-value of 0.0140, and the GDP positive group with a p-value of 0.0379; however, after performing multiple testing correction for the number of statistical tests performed, these results are NOT to be trusted. See Appendix Table VI for each individual coenzyme's results.

2. Evaluating Relationship Between Crystal Distance and abs-EDDA

Each of the datasets actually showed low p-values, but only the 10% abs-EDDA outlier dataset had a p-value at an acceptable level of statistical significance, as shown in Table III. However, the trend is consistent with increasing outlier definition and supports a weak association between nearby crystal contacts and regional abs-EDDA across ATP-bound regions.

TABLE III
FISHER EXACT TEST RESULTS ON 3.5 ANGSTROM CRYSTAL CONTACT DISTANCE CUTOFF VS. ABS-EDDA OUTLIERS

abs-EDDA outlier definition	p-value
1%	0.102
2.5%	0.0861
5%	0.0777
10%	0.0251

3. Evaluating Relationship Between Metal or Crystal Distance and abs-EDDA

In this analysis, where both the metal and crystal contact distances were considered, the only statistically significant result came from the 10% dataset, as shown in Table IV. However, the trend is not consistent across an increasing outlier definition.

TABLE IV
FISHER EXACT TEST RESULTS ON 3.5 ANGSTROM CRYSTAL CONTACT AND METAL ATOM DISTANCE CUTOFF VS. ABS-EDDA OUTLIERS

abs-EDDA outlier definition	p-value
1%	0.301
2.5%	0.691
5%	0.104
10%	0.0258

4. Evaluating Linear Relationships

For most of the datasets, the linear relationship effect sizes (correlations) were weak at best and are likely due to random chance. Fig. 1 shows the scatterplots of abs-EDDA versus minMCD (minimum metal atom contact distance) limited to the top 1%, 2.5%, 5%, and 10% outliers. No linear relationship is visually obvious and all of the Pearson correlation coefficients are small and have statistically insignificant p-values.

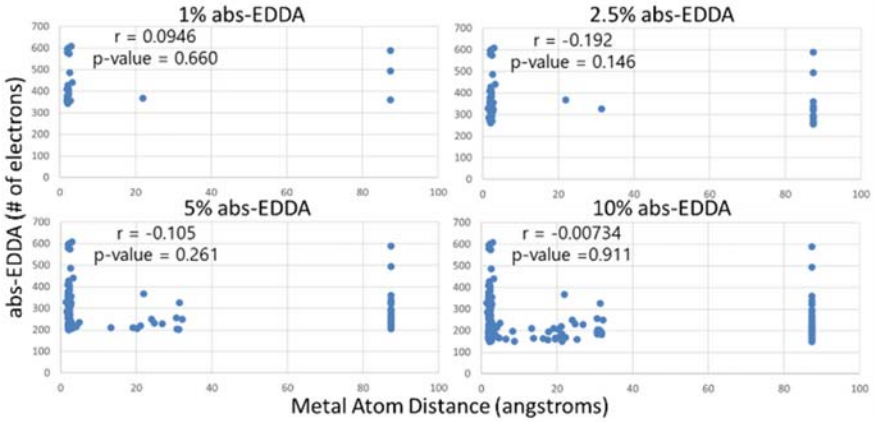

Fig. 1 Scatterplots of metal atom distance vs. abs-EDDA

Fig. 2 shows the scatterplots of abs-EDDA versus minCCD (minimum crystal contact distance) limited to the top 1%, 2.5%, 5%, and 10% outliers. Again, no linear relationship is visually obvious and all of the Pearson correlation coefficients are small and have statistically insignificant p-values.

Fig. 3 shows the scatterplots of abs-EDDA versus min[minMCD,minCCD] (minimum metal atom or crystal contact distance) limited to the top 1%, 2.5%, 5%, and 10%

outliers. Like the previous figures, no linear relationship is visually obvious and all of the Pearson correlation coefficients are small and have statistically insignificant p-values.

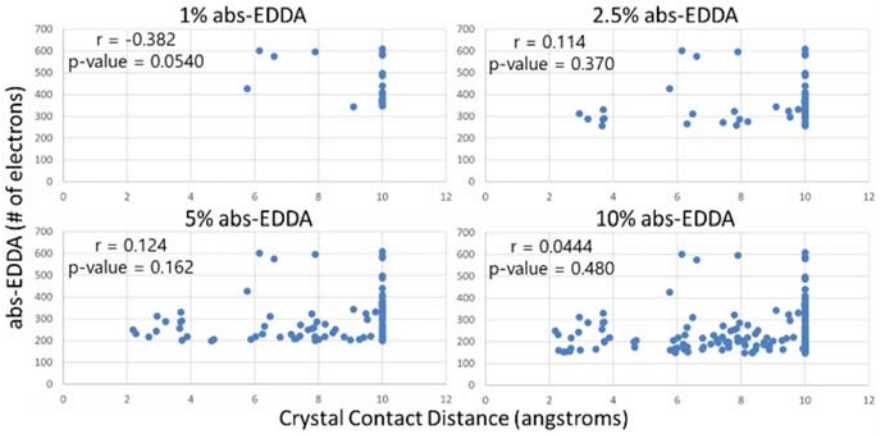

Fig. 2 Scatterplots of crystal contact distance vs. abs-EDDA; these results are limited to only ATP-bound regions

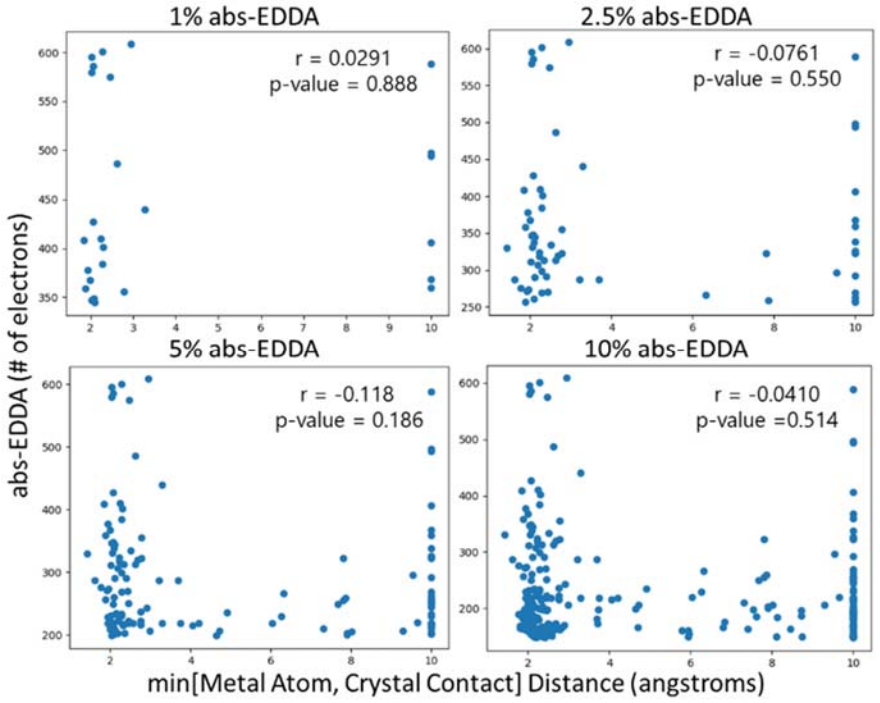

Fig. 3 Scatterplots of minimum of metal atom and crystal contact distance vs. abs-EDDA. These results are limited to only ATP-bound regions and to 10 angstroms

IV. Discussion

The results presented here revealed a consistent general effect from region-specific normalized b-factors across co-enzyme binding sites, which is strongly supported by p-values ranging from 0.00967 to as low as 1.43×10^{-20} from 1% to 10% outlier definitions. This is further supported by five out of six individual coenzyme results at the 10% outlier in Appendix Table V. Only co-enzyme GDP did not have a statistically significant p-value of 0.502. In contrast, no specific effect was observed from metal ion contact distances and only a very weak effect from crystal contact distance as compared to the normalized b-factor results. This suggests that modeling software failures, including manual human intervention or lack thereof, are likely as important, or even more important, than these other factors. Thus, all possible factors, including human error, should be considered when regions of low structural quality are detected. Furthermore, regional abs-EDDA is useful for detecting low quality regions when Fo-Fc electron density maps are available. When such maps are not available, regional average normalized b-factors could be useful for detecting low quality regions.

V. Conclusion

From the results, neither metal atom contact distance nor

crystal contact distance explained abs-EDDA outliers co-enzyme binding sites across x-ray crystallographic entries in the PDB. However, high normalized b-factors within these binding site regions were associated with these outliers. To support scientific rigor, the regional quality of the ligand binding sites should be examined prior to any downstream analysis of ligand-bound protein conformations. This would increase the accuracy of such analyses by indicating any structural flaws in the protein structure.

APPENDIX

TABLE V
B-FACTOR ANALYSIS FOR EACH CO-ENZYME

Co-Enzyme	Outlier 1% p-value	Outlier 2.5% p-value	Outlier 5% p-value	Outlier 10% p-value
ADP	0.335	0.428	0.330	0.0492
ATP	0.564	0.0416	0.0662	0.0318
FAD	6.57e-03	4.87e-04	1.75e-07	9.48e-14
GDP	0.0197	0.268	0.270	0.502
GTP	0.837	0.0501	0.0344	0.0393
NAD	0.288	0.0267	1.55e-07	1.12e-17

TABLE VI
EVALUATING 2X2 CONTINGENCY TABLES OF OUTLIER ABS-EDDA VS. 3.5 ANGSTROM METAL ATOM DISTANCE CUTOFF FOR EACH COENZYME DATASET

Co-Enzyme	Pos	Neg	Neu
ADP	0.150	0.160	0.250
ATP	0.256	0.479	0.172
FAD	0.0140	0.931	0.567
GDP	0.0379	0.395	0.606
GTP	0.533	1.00	0.810
NAD	0.392	0.739	0.340

SUPPLEMENTAL MATERIAL

Supplemental material with all of the code, intermediate and final results, and graphs is available as a public Figshare item [26].

REFERENCES

[1] Rosato, A., R. Tejero, and G.T. Montelione, *Quality assessment of protein NMR structures.* Current opinion in structural biology, 2013. 23(5): p. 715-724.
[2] Bhattacharya, A., R. Tejero, and G.T. Montelione, *Evaluating protein structures determined by structural genomics consortia.* Proteins: Structure, Function, and Bioinformatics, 2007. 66(4): p. 778-795.
[3] Berman, H., K. Henrick, and H. Nakamura, *Announcing the worldwide Protein Data Bank.* Nat Struct Biol, 2003. 10(12): p. 980.
[4] Burley, S.K., et al., *RCSB Protein Data Bank: Celebrating 50 years of the PDB with new tools for understanding and visualizing biological macromolecules in 3D.* Protein Science, 2022. 31(1): p. 187-208.
[5] Li, J., A. Fu, and L. Zhang, *An overview of scoring functions used for protein–ligand interactions in molecular docking.* Interdisciplinary Sciences: Computational Life Sciences, 2019. 11: p. 320-328.
[6] Stanzione, F., I. Giangreco, and J.C. Cole, *Use of molecular docking computational tools in drug discovery.* Progress in Medicinal Chemistry, 2021. 60: p. 273-343.
[7] Röhrig, U.F., et al., *Attracting Cavities 2.0: Improving the Flexibility and Robustness for Small-Molecule Docking.* Journal of Chemical Information and Modeling, 2023.
[8] Jacquemard, C., et al., *Binding mode information improves fragment docking.* Journal of Cheminformatics, 2019. 11(1): p. 1-15.
[9] Chachulski, L. and B.r. Windshügel, *LEADS-FRAG: a benchmark data set for assessment of fragment docking performance.* Journal of Chemical Information and Modeling, 2020. 60(12): p. 6544-6554.
[10] Chakraborti, S., K. Hatti, and N. Srinivasan, *'All That Glitters Is Not Gold': High-Resolution Crystal Structures of Ligand-Protein Complexes Need Not Always Represent Confident Binding Poses.* International Journal of Molecular Sciences, 2021. 22(13): p. 6830.
[11] Ding, K., et al., *Observing noncovalent interactions in experimental electron density for macromolecular systems: a novel perspective for protein–ligand interaction research.* Journal of Chemical Information and Modeling, 2022. 62(7): p. 1734-1743.
[12] Scardino, V., J.I. Di Filippo, and C.N. Cavasotto, *How good are AlphaFold models for docking-based virtual screening?* Iscience, 2023. 26(1).
[13] Yao, S. and H.N. Moseley, *Finding high-quality metal ion-centric regions across the worldwide Protein Data Bank.* Molecules, 2019. 24: p. 3179.
[14] Yao, S. and H.N. Moseley, *A chemical interpretation of protein electron density maps in the worldwide protein data bank.* PloS one, 2020. 15(8): p. e0236894.
[15] Søndergaard, C.R., et al., *Structural artifacts in protein− ligand X-ray structures: implications for the development of docking scoring functions.* Journal of medicinal chemistry, 2009. 52(18): p. 5673-5684.
[16] Cotelesage, J.J., et al., *Metalloprotein active site structure determination: synergy between X-ray absorption spectroscopy and X-ray crystallography.* Journal of inorganic biochemistry, 2012. 115: p. 127-137.
[17] Bramer, D. and G.-W. Wei, *Blind prediction of protein B-factor and flexibility.* The Journal of chemical physics, 2018. 149(13).
[18] Sun, Z., et al., *Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability.* Chemical reviews, 2019. 119(3): p. 1626-1665.
[19] Caldararu, O., et al., *Are crystallographic B-factors suitable for calculating protein conformational entropy?* Physical Chemistry Chemical Physics, 2019. 21(33): p. 18149-18160.
[20] Gutmanas, A., et al., *PDBe: protein data bank in Europe.* Nucleic acids research, 2014. 42(D1): p. D285-D291.
[21] Touw, W.G. and G. Vriend, *BDB: databank of PDB files with consistent B-factors.* Protein Engineering, Design & Selection, 2014. 27(11): p. 457-462.
[22] Warren, G.L., et al., *Essential considerations for using protein–ligand structures in drug discovery.* Drug Discovery Today, 2012. 17(23-24): p. 1270-1281.
[23] Yao, S., et al., *Aberrant coordination geometries discovered in the most abundant metalloproteins.* Proteins: Structure, Function, and Bioinformatics, 2017. 85(5): p. 885-907.
[24] Cock, P.J., et al., *Biopython: freely available Python tools for computational molecular biology and bioinformatics.* Bioinformatics, 2009. 25(11): p. 1422-1423.
[25] DeLano, W.L., *Pymol: An open-source molecular graphics tool.* CCP4 Newsl. Protein Crystallogr, 2002. 40(1): p. 82-92.
[26] Luo, A. and H.N.B. Moseley. *Supplemental Material Figshare*. 2023 September 15, 2023; Available from: https://doi.org/10.6084/m9.figshare.24144261.