Perspectives and Expectations in Structural Bioinformatics of Metalloproteins

Expectations for metal coordination geometries

Keywords: metal coordination, canonical coordination geometry, compressed angles, data sharing, collaboration vs competition

Sen Yao[a,b,c,d,e], Robert M. Flight[c,d,e], Eric C. Rouchka[a,b], and Hunter N.B. Moseley[c,d,e]

Affiliations
[a]School of Interdisciplinary and Graduate Studies
[b]Department of Computer Engineering and Computer Science
University of Louisville, Louisville, KY 40292, USA

[c]Department of Molecular and Cellular Biochemistry
[d]Markey Cancer Center
[e]Center for Environmental and Systems Biochemistry
University of Kentucky, Lexington KY 40356, USA

Correspondence to: hunter.moseley@uky.edu
Software and results available at: http://software.cesb.uky.edu, https://dx.doi.org/10.6084/m9.figshare.4229297, and https://dx.doi.org/10.6084/m9.figshare.4229333

ABSTRACT

Recent papers highlight the presence of large numbers of compressed angles in metal ion coordination geometries for metalloprotein entries in the worldwide Protein Data Bank, due mainly to multidentate coordination. The prevalence of these compressed angles has raised the controversial idea that significantly populated aberrant or even novel coordination geometries may exist. Some of these papers have undergone severe criticism, apparently due to views held that only canonical coordination geometries exist in significant numbers. While criticism of controversial ideas is warranted and to be expected, we believe that a line was crossed where unfair criticism was put forth to discredit an inconvenient result that compressed angles exist in large numbers, which does not support the dogmatic canonical coordination geometry view. We present a review of the major controversial results and their criticisms, pointing out both good suggestions that have been incorporated in new analyses, but also unfair criticism that was put forth to support a particular view. We also suggest that better science is enabled through: i) a more collegial and collaborative approach in future critical reviews and ii) the requirement for a description of methods and data including source code and visualizations that enables full reproducibility of results.

INTRODUCTION

Recently, Raczynska, Wlodawer and Jaskolski published the research article “Prior knowledge or freedom of interpretation? A critical look at a recently published classification of “novel” Zn binding sites” (1), a “critique” of our previously published paper “A less-biased analysis of metalloproteins reveals novel zinc coordination geometries” (2). The authors of this critique feel very strongly that novel zinc coordination geometries do not exist and that all of the examples put forth in Yao et al. 2015, are either misrepresented or based on a “few bad apple” structure

entries in the wwPDB. These conclusions are drawn from their close reexamination of 8 PDB entries identified in Yao et al. 2015 as representative entries of aberrant/novel coordination geometry clusters. The authors put forth point-by-point arguments to justify these conclusions, indicating that they could only reexamine these 8 PDB entries because "No list of PDB entries corresponding to the clusters identified in Y2015 was provided, only a figure for one representative structure per cluster was shown (Figs. YS1 and YS2, the latter reprinted here as Fig. 2)." (1). While we agree with several issues raised and suggestions made by the authors, we feel that they have not accurately represented the primary results presented in Yao et al. 2015, mainly that a significant number (in the thousands) of zinc coordination geometries (CGs) in the wwPDB have compressed angles and that these angles cause serious deviations from canonical CGs. Moreover, there are too many of these aberrant CGs to occur by chance or from "a few bad apple" structure entries. Also, these aberrant CGs are functionally distinct from CGs that do not have compressed angles and significantly complicate classification of metal binding sites into canonical CG models. Therefore in the following sections, we will address each of the primary issues and criticisms raised by authors, acknowledging good suggestions and improvements which we have already incorporated into our current analyses, but also indicating, in our opinion, where unfair criticism has been made.

MATERIALS AND METHODS

Relevant materials and methods are described in Yao et al., 2015, 2016 and Raczynska et al., 2016.

RESULTS and DISCUSSION

Access to published results

Raczynska et al. 2016 indicate in two places that they did not have access to the underlying results published in Yao et al., 2015:

> 1 - "No list of PDB entries corresponding to the clusters identified in Y2015 was provided, only a figure for one representative structure per cluster was shown (Figs. YS1 and YS2, the latter reprinted here as Fig. 2)."
>
> 2 - "The authors do not present these intermediate bond statistics, only the final values obtained after outlier rejection, so it is not possible to repeat the calculations using exactly the same parameters."

These statements are patently false, because while our previous article (2) was under peer review, we uploaded a 68 megabyte gzipped tar file (tarball) to our website containing all of the code used and the results from the paper as shown in Figure 1. The URL (http://software.cesb.uky.edu/, which redirects to http://bioinformatics.cesb.uky.edu/Main/SoftwareDevelopment) where this tarball can be downloaded, is listed on the title page of Yao et al. 2015 in both the PDF format and the web format (Figure 1A). At this website, a link to the tarball is clearly listed and is easily downloadable from our ftp site (Figure 1B and 1C). Within this tarball (Figure 1D), there is an output_manuscript subdirectory containing all of the published results (Figure 1E). As full disclosure, we updated the software and results tarball on December 7, 2015 (the webpage says December 15, but our timestamps indicate it was actually December 7) to remove improvements in the algorithm not reflected in the published material. These minor improvements had been accidentally added between the initial manuscript submission and its acceptance.

Regarding statement 1, the zinc site to cluster information is contained in the three files **normal_cluster_assg.RData**, **compressed_cluster_assg.RData** and **combined_cluster_assg.RData** in the **output_manuscript** subdirectory (Figure 1E). If Raczynska *et al* 2016 had examined this data, and needed help extracting the data from these files, we would have gladly generated a textual representation.

As for statement 2, all intermediate bond statistics are in files **stats.*.txt**, where the wildcard '*' are for different conditions. Also, the gzipped tarball contains **all** of the code used to run the calculations that generated the results presented in the manuscript. Therefore, anyone can use this code to generate the intermediate bond statistics, and examine them directly. Again, if Raczynska *et al* 2016 had tried to regenerate our results using this code, had difficulties, or contacted us, we would have gladly helped them generate the set of intermediate statistics for their examination.

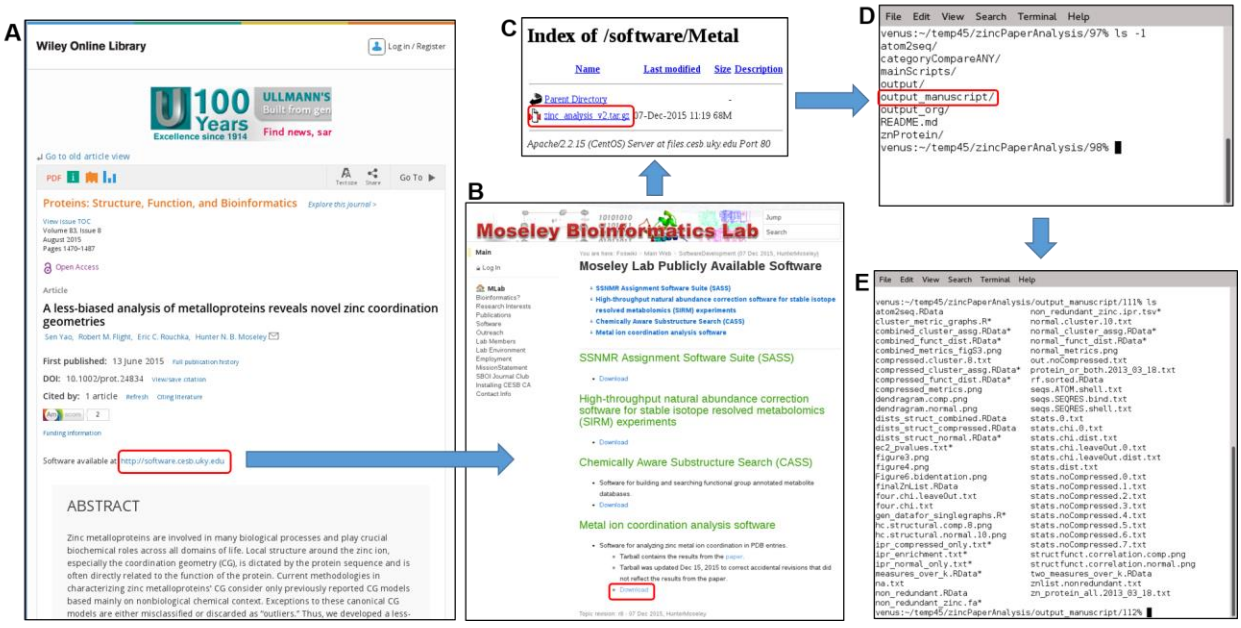

Fig 1. The web path providing full access to all results in Yao *et al*. 2015. (**A**) Web page of the online article. (**B**) Moseley lab wiki website. (C) Moseley lab ftp site. (**D**) Directory listing

of the tarball. (**E**) Directory listing of the output_manuscript subdirectory containing all of the published results.

Issues derived from missing ligands

We fully agree with Raczynska *et al* 2016 that some of the CGs examined in Yao et al 2015 are missing ligands, especially in the handful of examples pointed out by the authors. Part of the problem is due to the original approach we implemented for identifying a set of ligands together in a single chi-square statistical test that considers each ligand as a degree of freedom. This approach works well for picking the 4 "best" ligands based on expected bond-lengths, but is inherently flawed when the objective is to pick all reasonably plausible ligands. We have fixed this short-coming by using a single-ligand statistical test that leverages an additional set of algorithmic improvements including: i) ligand shell boundaries that minimize spurious ligand matching, ii) x-ray resolution-corrected bond-length standard deviations, and iii) a filter that prevents the selection of ligands where another atom is intervening between the potential ligand and the metal atom. Also in agreement with the authors, once these improvements in ligand selection were implemented, many of the low frequency events like the small numbers of metal-phosphorus ligations and hyper-compressed angles simply disappeared. We mentioned these low frequency events in Yao et al 2015 simply as full disclosure of our results at that time. This follows our basic scientific philosophy of showing and providing all of our results for others to evaluate as illustrated in Figure 1.

Issue with image chirality

Raczynska *et al* 2016 identified an issue in the chirality of the cysteines in wwPDB entry 4A48. This is a bona fide mistake on our part, where an image had been accidentally inverted for quick illustration of the type of perspective we had wanted for publication. We actually discovered a second image for the wwPDB 1RTQ entry with this same mirror image inversion issue within the same figure. These images, were never intended to be in the publication, but accidentally made it in when we forgot to recreate the correct version for publication. We have submitted an erratum to correct these images. Besides the unintentional mirror imaging of the actual structure, the wwPDB 4A48 entry illustrated a valid bidentation instance at the time of analysis. But in our subsequent analysis, the exact same zinc site (wwPDB entry 4A48) failed to pass our improved quality control filters. The author of the wwPDB entry (and the software they used) identified the structure as monomeric, but there was one potential ligand coming from a symmetry-related atom, which represents a possible artificial error or crystal packing. So rather than using the ligand set and structure suggested by Raczynska *et al* 2016, we removed structures like this to ensure a high quality of data for our current analyses. The zinc sites in the wwPDB 1RTQ entry, on the other hand, passed all new quality control filters and stayed in our current analyses.

Suggested improvements

Raczynska *et al* 2016 also suggested several improvements: i) including ligands with bond-lengths less than 1.3Å, ii) filtering out metal binding sites with low atom occupancy, and iii) including ligands from symmetry-related molecules. We have implemented and incorporated all of these improvements, but not necessarily in the manner they were originally suggested. First, we implemented a filter based on the first suggestion that removes metal binding sites from our analyses when they have atoms very close to the metal atom. Our rationale is that we should not

use sites with such bad steric clashes. However, this filter removed less than 0.8% of the metal binding sites, so their derived CGs had minimal impact on our aggregate results. Second, we added a 0.9 and higher atom occupancy filter, as suggested by the authors. This filter only removes about 5% of the metal binding sites. This also acts as a filter for not only low quality metal binding sites, but also removal of metal binding sites representing non-specific binding. The average structure-function correlations increased slightly than without the occupancy filter. Third, we implemented code that calculated and then included potential ligand atoms from symmetry models, both within the unit cell and in neighboring unit cells. However, we used CGs with non-biological unit ligands as a filter to remove metal binding sites that may represent artifacts from crystal packing and/or non-specific binding. As a result of all the filters we added and additional improvements implemented, we observed a significant increase in the structure-function Spearman correlation from 0.88 (p-value $< 2.2 * 10^{-16}$) to 0.90 (p-value $< 2.2 * 10^{-16}$) for the 4-ligand zinc sites. In addition, we see Spearman correlations for other metal analyses above 0.90 and even one above 0.95 (p-value $< 2.2 * 10^{-16}$).

Issues with incorrect modeling

As for two of the examples, PDB entries 3IFE and 1XTL, Raczynska *et al* 2016 had to download the original density file and remodel the structure to "disagree" with our coordination geometry model. Based on Raczynska's description of the improvements in *R/Rfree*, we agree that their manual remodeling appears to create better structural models for the two PDB entries; however, we cannot fully evaluate their models since they did not include them as supplemental material nor provide a link to them. But Raczynska et al 2016 interpretation of the bidentate aspartate ligand around 1XTL.A.1331 as a "superatom" relies on CheckMyMetal's pseudo-atom

interpretation of the bidentate coordination, which is heavily biased towards fitting the expected angles of canonical CGs (3). In addition, we do not see a clear justification for using a pseudo-atom representation for bidentate coordination. Either one trusts the wwPDB entry atom coordinates in a metal binding site especially if supported by the observed electron density map or one does not. Additionally, why would a pseudo-atom be a good guess at the correct ligand atom coordinates if one does not trust the original coordinates in the first place. A probabilistic rotamer search of sidechain conformations would have better justification.

However, in the larger picture, this suggested remodeling approach is not practical for a systematic analysis of all relevant entries in the PDB. Not to mention that most systematic analyses of public scientific repositories require a curation step to create a dataset of high enough quality to address the hypotheses being tested. Therefore, we have incorporated a new filter that detects systematically aberrant metal binding sites with respect to average normalized deviations in bond-lengths. We interpret that most of these systematically aberrant metal binding sites occur due to misassignment or incorrect modeling of the metal ion, causing a systematic extension in bond-lengths due to incorrect metal ion radius. A clear issue in Na and Mg versus other metal ions reflects what is generally known in the field to be the commonly misassigned metal ions. After incorporating the new filter, the bond length distributions improved dramatically, especially for Na and Mg. Also, the downstream structure-function correlation analyses improved. Given that proof-checking the modeling of each PDB file is unfeasible, our new quality-control filters can detect possibly misassigned metal ions to make sure real phenomena stand out from both background noise and the occasional systematic errors that slip through other filters, i.e. “a few bad apples”. A case in point, the 3IFE.A.411 zinc ion has an occupancy of only 0.7, which is screened out by our new filters. Also, Raczynska et al 2016

included A89Glu as a ligand to the zinc ion 1XTL.A.1331 in their remodeled structure. This ligand is 2.64 bond-length standard deviations away from the zinc ion in the original PDB entry and was filtered out by our 2.5 bond-length standard deviation cutoff.

CONCLUSIONS

We appreciate that Raczynska et al 2016 brought several issues to light in Yao et al 2015. Prior improvements in our methods, following publication, had highlighted some of the same issues. We also appreciate their suggested improvements, which we have now incorporated into our methods. These enhancements and others are illustrated in our companion publication that examines the coordination geometries across the five most prevalent metal ions in the wwPDB (Zn, Mg, Ca, Fe, and Na) [expected ref to Yao et al 201X]. Operating on the hypothesis that metal CGs are systematically analyzable across the wwPDB, this new paper fully demonstrates that all major results from Yao et al, 2015 still hold and actually are applicable to other protein-ligated metals: i) large numbers of compressed angles are present in metal CGs, ii) these compressed angles create aberrant CGs that complicate their classification into canonical CG models, and iii) metal sites in aberrant CGs exhibit distinct functional tendencies than those in canonical CGs. But we believe that a truly critical evaluation of the results in Yao et al, 2015 would lead most to accept the major results from a law of large numbers argument: there are too many compressed angles present to be due to the presence of a “few bad apple” PDB entries and these compressed angles lead to aberrant CGs that complicate their classification into canonical CG models. Also, these results demonstrate a high consistency (single bond length mode and low overall unimodal variation) of metal-ligand bond-lengths in metalloproteins reflecting expected strong dependency on physiochemical properties of metal ion coordination, while the large aberrations in metalloprotein CGs, especially with respect to bond angle variation (multimodality

of smallest angles) are necessary for the implementation of a diverse set of biochemical functions. In particular, the bond-length mode is very insensitive to errors and should be interpreted as the expected value of bond-length. On the other hand, the bond-length mean and standard deviation are more sensitive to errors due to minor skewness of unimodal bond length distributions, but these issues have been minimized after refinement via a variety of quality control filters. Likewise, bond angle modes are also insensitive to error. Moreover, the presence of distinctively compressed, significantly populated bond angle modes cannot be explained away as due to 'a few bad apples' or even a number of bad apples. Large numbers of erroneous metal binding sites across metals with similar deviations (systematic error) would be needed to produce the unexpected bond angle modes of the magnitudes we observe in the data. This pattern of systematic error is highly unlikely. A much better explanation is that the unexpected bond-length and bond-angle modes are a product of multidentation and specific functional group chemical properties for the most part and are required to implement a wide variety of biochemical functions. This view is supported by the enrichment of distinct biochemical and cellular function annotations for metal binding sites with compressed vs normal angles [expected ref to Yao et al 201X].

A related issue is why it took so long to detect, characterize, and classify metal binding sites with compressed angles and aberrant CGs. An analysis in Yao et al. 2015 indicated that enough examples of aberrant Zn-ion CGs with compressed angles were present in the PDB since 2003 for reliable detection. Two years prior, Andreini et al. had indicated the presence of large numbers of outlier CGs (it was actually their largest class), but without much explanation for their existence (4). Also, the carboxylate shift phenomenon in Zn metal-ion coordination by proteins (5) had been documented in the literature since 1998 (6). One of the possible reasons is

that prior analyses for detecting ligand atoms were based primarily on matching possible ligand atoms to canonical CGs (3, 4, 6-9). The possibility of bidentation was often ignored or even misanalyzed as pseudo-atoms (3) (see previous discussion) based on a bias against anything that did not look like a canonical CG. In general, a biased search will not detect what was not looked for. Therefore, we believe that a canonical CG bias in prior analyses simply prevented the detection of aberrant CGs with compressed angles.

While we appreciate some of the issues raised by Raczynska et al 2016, we do think it would have been more productive had the authors contacted us prior to submitting their paper, as we could have discussed these issues and collaborated. We are also concerned with the general tone of their paper and a more recent follow-up viewpoint paper in *FEBS Journal* (10) that includes a range of criticisms from reasonable to false that appear centered on discrediting the existence of "novel" zinc coordination geometries. These papers follow a previously published paper (11) by some of the same authors that had a similar hypercritical tone regarding a group of other crystallographic papers (12). They are reminiscent of the recently published editorial in the New England Journal of Medicine indicating that "some front-line researchers" feel that those who make use of others datasets as "research parasites" (13, 14). These publications highlight the emergence of scientifically unproductive perspectives as a by-product of the establishment of large scientific repositories for data sharing (15, 16).

Through the use of the scientific method, research moves forward in a generally self-correcting fashion as scientists evaluate the claims, methods, and results of each other. We have publications that pointed out issues in published analyses (2), published software (17), and public databases (18). However, we have tried to raise these issues with an open and collegial tone, with due diligence in fully checking other authors' results, and, in certain cases, contacting the

authors in order to fully understand and check our own criticisms. Also, the examination of methods and results should be applied uniformly. Therefore, access to detailed descriptions of methods and results that allow reproducibility, which often now includes programming code, should be expected from everyone, including those well-established in a field. This expectation of examination creates a "trust but verify" philosophy that we follow ourselves. However, there are many examples of methods and results published without enough detail to reproduce the work. For example, MESPEUS, a database of metal coordination environments of metalloproteins, which was published without source code nor with enough detail to recreate the database (19, 20). The web interface to the database has been periodically unavailable (i.e. down for a few months and just recently back up) and the database itself would be rather hard for others to reproduce. As a possible counter example, CheckMyMetal (CMM) is a wonderful online resource for analyzing metal binding sites in protein structures (3). While no source code for CMM is available online (we checked the publication and both the hosting website and the corresponding author lab website), its methods are well-described in the literature and the authors indicated via email correspondence that they would provide some type of access to their code (see email correspondence in supplemental material). However, we found some minor differences in ligand selection between our methods and CMM in a few examples, but could not determine the exact reason for these differences, since we have had no access to the underlying code for CMM nor a more detailed description of its methods yet. Also, we would like to perform a systematic comparison of ligand selection results between our methods and CMM methods. So far, our attempts to contact the authors for additional details about CMM algorithms took some effort and time to receive an initial response (see the appendix for the five attempts over a month's period from three different email addresses). Finally, a very good

counter example of publication with adequate detail for evaluation and reproducibility is found in the FindGeo tool for determining coordination geometries of metal binding sites (9). The authors published a reasonable amount of description and made source code available, adequate for both evaluation and reproducibility. Some developers are posting their source code online for others to immediately use and evaluate, even before publication. One example is the LiteMol Viewer developed by David Sehnal (21) and used by the Protein Data Bank in Europe (PDBe) (22).

Our own standard practice is to make both detailed descriptions and code available for our published research to facilitate full reproducibility (see Figure 1). We suggest that both code and detailed descriptions of algorithms should be the expected norm in peer-reviewed publication, with maybe rare exceptions for intellectual property and patent issues if concerns about evaluation can be properly addressed. However, long-term reproducibility and evaluation of published methods is difficult due to lack of persistence of source code and scientific results. For example, the FindGeo source code is available on a lab website. But what happens when this lab closes and the website goes away. The source code will likely be lost. The same fate is likely to happen to MESPEUS and CMM too. These issues are part of larger scientific data persistence, access, and citation issues at the heart of scientific reproducibility and reusability in general (23, 24). Luckily, new repository resources like FigShare (25), Zenodo (26), Dryad (27), and Dataverse (28) are providing facilities for data and source code artifact persistence, access, and citation. Also, GitHub (29) provides facilities for actively developing and maintaining software projects and, in collaboration with FigShare and Zenodo, provides dynamic persistence of active software projects designed for reuse (30). Therefore, data and source code persistence will likely become a new requirement for publication in the near future. Our previously available

tarball for Yao et al 2015 as well as the code and results for our new five metal analyses (31)[expected ref to Yao et al 201X] have been deposited to FigShare to ensure persistent availability long after our webserver is no longer accessible.

Also, we suggest that the visualization of the distribution of data and derived values should become the expected norm for the publication of computational, mathematics, and statistical methods in structural bioinformatics. With the right visualizations of data and derived statistics, a range of assumptions for the underlying methods employed in a given published analysis can be easily and adequately evaluated. In particular, histograms and probability distribution graphs are often very useful for visualizing the frequency of observations and determining modality, symmetry, and the specific type of distribution. For examples, look at the distribution figures in Yao et al., 2015 and 201X.

In conclusion, a lack of both openness and standards leads to results that cannot be verified. A lack of collaboration prevents synergy of expertise. And unbridled competitiveness is often single-minded, short-sighted, and error-prone. We fully believe that an open, collaborative, and collegially competitive approach produces better and faster results in science. So, the broader scientific community, spanning a range of expertise including inorganic chemistry, biochemistry, and structural bioinformatics, may be best at deliberating whether unexpected CG models in metalloproteins are just aberrant or truly novel. More important is determining how certain CG models enable specific biochemical functions.

ACKNOWLEDGEMENTS

The authors have no conflict of interest with this work. This work was supported in part by National Science Foundation NSF 1252893 (Hunter N.B. Moseley) and NIH 1U24DK097215-01A1 (Richard M. Higashi, Teresa W.-M. Fan, Andrew N. Lane, and Hunter N.B. Moseley). ECR is supported by grant P20GM103436 (Nigel Cooper, PI). The contents of this work are solely the responsibility of the authors and do not represent the official views of the NIH or the National Institute for General Medical Sciences (NIGMS).

REFERENCES

1. Raczynska J, Wlodawer A, Jaskolski M. Prior knowledge or freedom of interpretation? A critical look at a recently published classification of “novel” Zn binding sites. Proteins: Structure, Function, and Bioinformatics. 2016.
2. Yao S, Flight RM, Rouchka EC, Moseley HN. A less‐biased analysis of metalloproteins reveals novel zinc coordination geometries. Proteins: Structure, Function, and Bioinformatics. 2015;83(8):1470-87.
3. Zheng H, Chordia MD, Cooper DR, Chruszcz M, Müller P, Sheldrick GM, Minor W. Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. Nature protocols. 2014;9(1):156-70.
4. Andreini C, Cavallaro G, Lorenzini S, Rosato A. MetalPDB: a database of metal sites in biological macromolecular structures. Nucleic acids research. 2013;41(Database issue):D312-9. doi: 10.1093/nar/gks1063. PubMed PMID: 23155064; PMCID: 3531106.
5. Sousa SF, Fernandes PA, Ramos MJ. The carboxylate shift in zinc enzymes: a computational study. Journal of the American Chemical Society. 2007;129(5):1378-85. doi: 10.1021/ja067103n. PubMed PMID: 17263422.
6. Alberts IL, Nadassy K, Wodak SJ. Analysis of zinc binding sites in protein crystal structures. Protein science : a publication of the Protein Society. 1998;7(8):1700-16. doi: 10.1002/pro.5560070805. PubMed PMID: 10082367; PMCID: 2144076.
7. Patel K, Kumar A, Durani S. Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics. 2007;1774(10):1247-53.
8. Liu Z, Wang Y, Zhou C, Xue Y, Zhao W, Liu H. Computationally characterizing and comprehensive analysis of zinc-binding sites in proteins. Biochimica et biophysica acta. 2014;1844(1 Pt B):171-80. doi: 10.1016/j.bbapap.2013.03.001. PubMed PMID: 23499845.
9. Andreini C, Cavallaro G, Lorenzini S. FindGeo: a tool for determining metal coordination geometry. Bioinformatics. 2012;28(12):1658-60. doi: 10.1093/bioinformatics/bts246. PubMed PMID: 22556364.

10. Rupp B, Wlodawer A, Minor W, Helliwell JR, Jaskolski M. Correcting the record of structural publications requires joint effort of the community and journal editors. The FEBS journal. 2016. doi: 10.1111/febs.13765.
11. Shabalin I, Dauter Z, Jaskolski M, Minor W, Wlodawer A. Crystallography and chemistry should always go together: a cautionary tale of protein complexes with cisplatin and carboplatin. Acta Crystallographica Section D: Biological Crystallography. 2015;71(9):1965-79.
12. Yeates TO. Responses to Crystallography and chemistry should always go together: a cautionary tale of protein complexes with cisplatin and carboplatin. Acta Crystallographica Section D: Biological Crystallography. 2015;71(9):1980-1.
13. Drazen JM. Data sharing and the journal. New England Journal of Medicine. 2016.
14. Longo DL, Drazen JM. Data Sharing. New England Journal of Medicine. 2016;374(3):276-7.
15. Berger B, Gaasterland T, Lengauer T, Orengo C, Gaeta B, Markel S, Valencia A. ISCB's Initial Reaction to The New England Journal of Medicine Editorial on Data Sharing. PLOS Comput Biol. 2016;12(3):e1004816.
16. Ohno-Machado L. A message to the next generation of biomedical informatics professionals. Journal of the American Medical Informatics Association. 2016;23(2):241-.
17. Flight RM, Wentzell PD. Potential bias in GO:: TermFinder. Briefings in bioinformatics. 2009;10(3):289-94.
18. Mitchell JM, Fan TW-M, Lane AN, Moseley HN. Development and in silico evaluation of large-scale metabolite identification methods using functional group detection for metabolomics. Comprehensive Systems Biomedicine. 2014:70.
19. Harding MM, Hsin K-Y. Mespeus—a database of metal interactions with proteins. Structural Genomics: General Applications. 2014:333-42.
20. Hsin K, Sheng Y, Harding M, Taylor P, and, Walkinshaw M. MESPEUS: a database of the geometry of metal sites in proteins. Journal of Applied Crystallography. 2008;41(5):963-8.
21. Sehnal D. LiteMol: Powerful and blazing-fast tool for handling 3D macromolecular data (not only) in the browser 2016. Available from: https://github.com/dsehnal/LiteMol.
22. Gutmanas A, Alhroub Y, Battle GM, Berrisford JM, Bochet E, Conroy MJ, Dana JM, Montecelo MAF, van Ginkel G, Gore SP. PDBe: protein data bank in Europe. Nucleic acids research. 2014;42(D1):D285-D91.
23. Starr J, Castro E, Crosas M, Dumontier M, Downs RR, Duerr R, Haak LL, Haendel M, Herman I, Hodson S. Achieving human and machine accessibility of cited data in scholarly publications. PeerJ Computer Science. 2015;1:e1.
24. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3.
25. FigShare. Available from: www.figshare.com.
26. Zenodo. Available from: https://zenodo.org/.
27. Dryad Digital Repository. Available from: http://datadryad.org/.
28. The Dataverse Project. Available from: http://dataverse.org/.
29. GitHub. Available from: https://github.com/.
30. Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, Leprevost F, Fufezan C, Ternent T, Eglen SJ, Katz DS, Pollard AK, Flight RM, Blin K, Vizcaino JA. Ten Simple Rules for Taking Advantage of git and GitHub. PLoS Comput Biol. 2016;12(7):e1004947.
31. Yao S, Flight RM, Rouchka E, Moseley HN. Source code, data, and results for five metalloprotein analyses.: FigShare; 2016. Available from: https://dx.doi.org/10.6084/m9.figshare.4229297.