Peer Review Information

Journal: Nature Computational Science
Manuscript Title: Directional Multiobjective Optimization of Metal Complexes at the Billion Scale
Corresponding author name(s): Professor David Balcells

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:

Date:	30th November 23 15:47:50
Last Sent:	30th November 23 15:47:50
Triggered By:	Kaitlin McCardle
From:	kaitlin.mccardle@us.nature.com
To:	david.balcells@kjemi.uio.no
BCC:	kaitlin.mccardle@us.nature.com
Subject:	Decision on Nature Computational Science manuscript NATCOMPUTSCI-23-0697A
Message:	** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors. **

Dear Professor Balcells,

Your manuscript "Directional Multiobjective Optimization of Metal Complexes at the Billion-Scale with the tmQMg-L Dataset and PL-MOGA Algorithm" has now been seen by 3 referees, whose comments are appended below. You will see that while they find your work of interest, they have raised points that need to be addressed before we can make a decision on publication.

The referees' reports seem to be quite clear. Naturally, we will need you to address *all* of the points raised.

While we ask you to address all of the points raised, the following points need to be substantially worked on:

- Please be sure to clarify the novelty of your method to address concerns raised by

Reviewer #1.

Please use the following link to submit your revised manuscript and a point-by-point response to the referees' comments (which should be in a separate document to any cover letter):

[REDACTED]

** This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first. **

To aid in the review process, we would appreciate it if you could also provide a copy of your manuscript files that indicates your revisions by making use of Track Changes or similar mark-up tools. Please also ensure that all correspondence is marked with your Nature Computational Science reference number in the subject line.

In addition, please make sure to upload a Word Document or LaTeX version of your text, to assist us in the editorial stage.

To improve transparency in authorship, we request that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We hope to receive your revised paper within three weeks. If you cannot send it within this time, please let us know.

We look forward to hearing from you soon.

Best regards,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

Reviewers comments:

Reviewer #1 (Remarks to the Author):

The article "Directional Multiobjective Optimization of Metal Complexes at the Billion-Scale with the tmQMg-L Dataset and PL-MOGA Algorithm" describes the generation of a systematic dataset of ligands that are based on the Cambridge Structural Database and the use of this dataset to assemble metal complexes in a systematic manner which enables the implementation of this workflow together with a multiobjective

genetic algorithm for inverse molecular design. The methods are described in a clear manner and the corresponding data and code is also provided making the work readily usable by interested readers, provided they are sufficiently skilled in python. However, I think that the manuscript currently does not provide proper benchmarking of their algorithm allowing the reader to assess the performance of their approach. In addition, I think the current version of the manuscript does not have a proper discussion. Finally, it is not clear to me how relevant the demonstrated multiobjective optimization task is for real-world applications. Overall, I think the manuscript is potentially publishable, but only when these major issues are addressed. Please find my detailed comments below:

Major issues:
My major issue with the manuscript is that it does not provide proper benchmarking of the PL-MOGA algorithm against alternative approaches making it entirely unclear whether the implemented method provides any improvement over alternative algorithms.

It seems to me that the main novelty of the multiobjective genetic algorithm is the pareto-lighthouse component. While I see the utility of this approach, I would say it is a rather minor advancement over existing approaches. I also think it would be useful to highlight inverse molecular design scenarios in which this particular implementation will be superior to alternative approaches.

Another limitation of the current manuscript is the focus on the property pair polarizability and HOMO-LUMO gap. While these properties are somewhat relevant for some applications of metal complexes, they seem to me more like a toy task to demonstrate the capabilities of the multiobjective optimization algorithm. This is meaningful from the perspective of algorithm development as they allow to characterize the implemented functions, from the point of chemistry, these properties are not immediately descriptive of a real-world design task.

In the current manuscript, I think that the discussion is largely a summary of the findings rather than a real discussion. I would recommend the authors to rework this section and add a proper discussion to the text.

Page 6, line 85: The authors state the following: "tmQMg-L has high chemical diversity; e.g., the random twenty-eight samples shown in Figure S1 represent twelve different ligand categories." To substantiate their claim of diversity, I think the authors should use a more systematic metric than the number of categories present in a small random sample.

Page 8, line 103: The authors state the following: "The success rate in the assignment of the charges was 95%." I think the authors should state in the text how the success rate was determined.

Page 11, line 159: The authors state the following: "Given the interest in the use of square planar Pd(II) TMCs as chemotherapy drugs, as well as their applications in catalysis, we decided to tackle the optimization of alpha and epsilon along the Pareto front of this TMC space." I think the authors should comment on the relevance of this task for catalysis.

Minor issues:
Page 15 line 195: The authors state the following: "Previous implementations have used constant thresholds that need to be known a a priori, which can be challenging in some applications." I think one of the two "a" need to be removed.

Reviewer #1 (Remarks on code availability):

I also inspected the code. It is very nice to see all the code and the dataset provided. I think the code is in a good state, overall. However, I think the authors should consider the following for improvements. Right now, the PL-MOGA algorithm requires their users to modify python code to implement custom optimization runs. I think that this is a relatively high entrance barrier for somebody with limited experience in python. I would suggest to at least allow for command line use, for instance via the argparse package. Similarly, I think both repositories would benefit immensely from documentation. Finally, as the code requires installation of a few additional packages, it would be highly beneficial for usability to add full installation instructions to the repository or its documentation.

Reviewer #2 (Remarks to the Author):

This article, by Kneiding, Nova, Balcells uses multiobjective optimization on a diverse transition metal complex data set to simultaneously optimize polarizability and HOMO-LUMO gap. The authors use a genetic algorithm to find diverse TMCs on the Pareto front and analyze their chemical diversity.

This article is very well written and is clearly reproducible. I commend the authors for their work and for making their code available. I believe the article is publishable as is, but would like to make some suggestions to improve the article:

Major suggestions:
1. The authors introduce a large dataset that contains a synthesizable library of TMC ligands with charges extracted. Some other authors: https://pubs.acs.org/doi/epdf/10.1021/acs.jctc.2c00468 created a iterative method for charge determination that does not require any calculations. Does the authors' algorithm for charge determination have any benefit over something like this? I ask because removing the use of NBO for charge determination may help accelerate screening.
2. How do the authors distinguish between compositional isomers of transition metal complexes with the same ligands arranged differently within the same coordination geometry? The example for the 4 coordinate cis- and trans- case is given in Figure 1B, but I am curious about octahedral complexes, where this problem can become more challenging. Clearly, the authors already have thought about this, as mentioned on page 5, but I would like to know how isomers are encoded in genes. Additionally, are HOMO-LUMO and polarizability sufficiently distinguished in two isomers of the same composition?
3. On page 13, when the Tanimoto coefficient is measured, does the ordering of the SMILES matter? It appears that there should be some sort of rule, since the ligand SMILES are concatenated, else you may have different concatenations for the same

complex.

Minor suggestions:
1. Since the authors are releasing a dataset, they may be interested in the Weisfeiler-Lehman graph hash (detailed here: 10.1021/acs.jpclett.3c01214). The data set is useful as it is. However, having the Weisfeiler-Lehman graph hashes (both with and without atom number attribution) can help users filter by connectivity and composition uniqueness. This will make the data set even more useful.

- Aditya Nandy

Reviewer #2 (Remarks on code availability):

The code and data are reproducible. I suggest the authors upload the dataset on to a repository like zenodo which has better version control for data sets.

Reviewer #3 (Remarks to the Author):

In my opinion the following issues should be addressed before the paper is suitable for publication

1. The authors write "In the 1.37M space, the MOGA located 130 TMC hits over the (σ, ϵ) Pareto front with high chemical diversity and in an interpretable manner." But it's hard to judge whether that is a good performance. How many TMCs are there on the Pareto front? What percentage of these did the algorithm find?

2. The authors write "All these TMCs are present in the CSD, and, therefore, all ligands in tmQMg-L exist in at least one TMC that has been characterized experimentally, which shall enforce synthesizability in generative models." Well, that ensures the synthesizability of each individual ligand, but not of the TMC. It is not a given that when you mix the ligands and the metals that they will form the TMC the authors predict. As noted in this paper (10.1126/science.abj0999): "the factors that dictate nuclearity (e.g., monomer versus dimer), favored oxidation state, and ligation state of a catalyst are all too frequently barely understood". This should be noted.

3. The Lighthouse approach is presented assuming the objectives have positive values. While it is straightforward to extend it to negative values, I think it would be a little tricky for objectives that can be both positive and negative (e.g. logP values). Of course one could simply discard negative values if one wants to maximize the values, but what if positive values are rare and the starting population only has negative values. Since the paper introduces the Lighthouse, it would be good to have a discussion of this as it goes the the general applicability of the method.

4. The authors write "MO with genetic algorithms (MOGAs) has been implemented with different methods,53–55 …". I suggest also citing this paper: 10.1021/acs.jcim.8b00839, which has been quite influential in the field.

Jan Jensen (I choose to review this paper non-anonymously)

Author Rebuttal to Initial comments

Reviewer 1

Comment: *The methods are described in a clear manner and the corresponding data and code is also provided making the work readily usable by interested readers, provided they are sufficiently skilled in python.*

Reply: We thank the reviewer for the positive comments on the methods, data, and code contents of the manuscript. We agree on the need to present the python code in a format that facilitates its use by scientists lacking advanced skills in programming. The revisions made in this regard are described in the reply to the last comment made by this reviewer.

Comment: *However, I think that the manuscript currently does not provide proper benchmarking of their algorithm allowing the reader to assess the performance of their approach. In addition, I think the current version of the manuscript does not have a proper discussion. Finally, it is not clear to me how relevant the demonstrated multiobjective optimization task is for real-world applications. Overall, I think the manuscript is potentially publishable, but only when these major issues are addressed.*

Reply: We thank the reviewer for noting these important points. We agree on improving the benchmarking of our method, as well as its discussion, including potential applications to chemistry problems. The replies below provide further details on the revisions made in response to these points.

Comment: *My major issue with the manuscript is that it does not provide proper benchmarking of the PL-MOGA algorithm against alternative approaches making it entirely unclear whether the implemented method provides any improvement over alternative algorithms. It seems to me that the main novelty of the multiobjective genetic algorithm is the pareto-lighthouse component. While I see the utility of this approach, I would say it is a rather minor advancement over existing approaches.*

Reply: We agree with the reviewer on the importance of doing benchmark studies. However, we could not identify an alternative approach for transition metal complexes that can be directly compared to our method and thus be used as a benchmarking reference. Our coordination geometry-adapted, full-ligand genetic operations are fundamentally different from the ligand fragment approaches reported in the literature. Another fundamental difference, and advantage, is that our method does not rely on SMILES strings, making it applicable to transition metal complexes. These are both relevant and novel components of our method.

Nonetheless, in order to quantify the performance of our method, we defined baselines. For the Pareto-Lighthouse algorithm, we determined the number of additional Pareto solutions found by a directed optimization relative to a baseline consisting of a non-directed one. For example, in the 1.37M space, whereas the non-directed calculation found only 53 of 130 solutions at the center of the Pareto front (i.e. 41%), more than double of this amount, 126 of 130 (i.e. 97%), were found in the same region with the center-directed calculation. Further, we think that the Pareto-Lighthouse algorithm is a major advancement because it enables fine control over the aim and scope of multiobjective optimizations. This is especially valuable in complex design tasks involving multiple properties within large metal-organic spaces. Also related to performance, we would like to note that the use of an xTB fitness in our

method is a novel approach in genetic algorithms applied to metal complexes (we could only find a precedent in this recent preprint: Strandgaard, Jensen et al. ChemRxiv 2023, DOI: 10.26434/chemrxiv-2023-t73mw). Taking the DFT fitness as baseline, we determined that our xTB-based approach accelerates the multiobjective optimization by a factor of 81. Further, we would like to note that the performance of our method is largely due to both the PL-MOGA algorithm and the tmQMg-L dataset. The combined use of these two elements delivers a significant advantage. To the best of our knowledge, there is no other method enabling the directional multiobjective optimization of Pareto front metal complexes within vast chemical spaces made of synthesizable and highly diverse metal ligands; a capability that is enabled by both the algorithm and the dataset.

Finally, and regarding the comments made by other reviewers also related to benchmarking, we would like to note that in these revisions 1) We determined that, in the 1.37M space, and after exploring only 1% of it, the PL-MOGA found 18 of the 30 dominating points defining the Pareto front (*i.e.* 60% success rate), and 2) We compared our method for ligand charge assignment against the one reported by the Kulik group (Duan, Ladera, Liu, Taylor, Ariyarathna, Kulik, *J. Chem. Theory Comput.* **2022**, *18*, 4836-4845), finding an agreement of 98.9%.

Action in the manuscript: The revised manuscript now includes the following elements: the gain in performance introduced by the Pareto-Lighthouse algorithm (lines 210-213) and the xTB fitness relative to baselines (lines 278-281), the performance in finding the dominating points of the Pareto front (lines 174-175), and the comparison to the Kulik's method for ligand charge determination (lines 300-301).

Comment: *I also think it would be useful to highlight inverse molecular design scenarios in which this particular implementation will be superior to alternative approaches.*

Reply: We agree on the value of making this highlight. For example, relative to the junction tree variational autoencoder (JT-VAE) method, our method has the advantage of enabling generative tasks with ligands that are charged. Further, and besides being superior, our method could be used to augment a JT-VAE model in the following ways: 1) extend the optimization of metal complexes made of generated ligands from homoleptic to heteroleptic, 2) extend the conditional generation of metal complexes from one to multiple properties, and 3) improve the synthesizability of the generated ligands.

Action in the manuscript: The revised discussion section of the manuscript highlights an inverse molecular design scenario in which our method can be superior to, for example, a variational autoencoder, where it can be used to augment the resulting models (lines 281-284).

Comment: *Another limitation of the current manuscript is the focus on the property pair polarizability and HOMO-LUMO gap. While these properties are somewhat relevant for some applications of metal complexes, they seem to me more like a toy task to demonstrate the capabilities of the multiobjective optimization algorithm. This is meaningful from the perspective of algorithm development as they allow to characterize the implemented functions, from the point of chemistry, these properties are not immediately descriptive of a real-world design task.*

Reply: The joint optimization of the HOMO-LUMO gap and the polarizability can be seen as a toy task to assess our method since the resulting fitness has a conveniently moderate computational cost. However, this task has been also suggested to be relevant in drug design where both stability and van der Waals interactions need to be maximized (Lilienfeld, Müller, Tkatchenko, *Nat. Rev. Chem.* **2020**, *4*, 347-358). In a materials discovery funnel aimed at a real-world problem, our method can be used as the first filter reducing the number of potential candidates by several orders of magnitude. In the next filter, the selection can be further constrained by, for example, first-

principles calculations, yielding a small collection of hits than can be finally verified experimentally. We are interested in developing these filters for the discovery of metallodrugs, but this is out of the scope of the present work. Further, the xTB fitness of our model gives access to all these other quantum properties: HOMO and LUMO orbital energies, dipole moment, heat capacity, enthalpy, entropy, and the electronic and free energies. These properties can be related to several topics of interest, including photochemistry, solubility, and thermodynamics. In this regard, expanding the Pareto-Lighthouse to more dimensions over larger populations should be both affordable and easy to implement with the ZEROMASK function operating on the xTB fitness.

Action in the manuscript: In the revised discussion section, we explain how our method can enable a discovery funnel in a real-world design task exploring massive chemical spaces (lines 274-278), as well as other applications related to the multiple and diverse properties that can be added to the fitness (lines 278-281). Further, in the revised "PL-MOGA" section, the statement "The interplay between these two molecular properties is relevant in drug discovery, since, ideally, a commercial active compound maximizes both alpha, enforcing weak interactions with biomolecules, and epsilon, enforcing stability against heat or light" has been complemented by citing the work of Lilienfeld, Müller, and Tkatchenko mentioned above (lines 157-159).

Comment: *In the current manuscript, I think that the discussion is largely a summary of the findings rather than a real discussion. I would recommend the authors to rework this section and add a proper discussion to the text.*

Reply: We agree with the reviewer on this important point, and we have re-written the entire Discussion section.

Action in the manuscript: We have changed the text of the discussion section (lines 262-288) to put a stronger focus on the meaning of the results and their implications to the research field. The revised discussion also includes new elements arising from the replies to the comments of this reviewer and all others. Conclusions were also briefly summarized since articles in this journal do not have a section dedicated to them.

Comment: *Page 6, line 85: The authors state the following: "tmQMg-L has high chemical diversity; e.g., the random twenty-eight samples shown in Figure S1 represent twelve different ligand categories." To substantiate their claim of diversity, I think the authors should use a more systematic metric than the number of categories present in a small random sample.*

Reply: We agree on this point. In order to provide a more systematic and yet intuitive metric for diversity, we computed the histogram of the SMARTS strings representing the chemical patterns of the metal-bound atoms in the tmQMg-L ligands.

Action in the manuscript: The Supporting Information provides the histogram of these SMARTS patterns (page S4), which is also referred to in the revised "tmQMg-L ligand dataset" section of the manuscript (lines 85-86).

Comment: *Page 8, line 103: The authors state the following: "The success rate in the assignment of the charges was 95%." I think the authors should state in the text how the success rate was determined.*

Reply: We also agree on this point. The success rate was manually determined by the authors on a random selection of 500 ligands. The authors calculated the charge taking the ligand geometry as reference, instead of the NBO Lewis structure, and using the ionic electron-counting scheme. Responding to the comment made by another reviewer, we also benchmarked our method against another one recently reported by the Kulik group, finding an agreement of 98.9%.

Action in the manuscript: This point has been clarified in the revised manuscript under the "Ligand charge assignment" section of the Methods (lines 298-301). The data associated with the benchmark against Kulik's method is mentioned in the Data and code availability section (lines 426-427) and is available from a .csv file at https://github.com/hkneiding/tmQMg-L/tree/main/benchmarks.

Comment: *Page 11, line 159: The authors state the following: "Given the interest in the use of square planar Pd(II) TMCs as chemotherapy drugs, as well as their applications in catalysis, we decided to tackle the optimization of alpha and epsilon along the Pareto front of this TMC space." I think the authors should comment on the relevance of this task for catalysis.*

Reply: Thanks to this comment, we realized that this statement was misleading, and we thus revised it.

Action in the manuscript: The revised manuscript now states "*Given the interest in the use of square planar Pd(II) TMCs as chemotherapy drugs, we decided to tackle the optimization of alpha and epsilon along the Pareto front of this TMC space. Other applications, including catalysis, would require the optimization of different properties*" (lines 159-162).

Comment: *Page 15 line 195: The authors state the following: "Previous implementations have used constant thresholds that need to be known a a priori, which can be challenging in some applications." I think one of the two "a" need to be removed.*

Reply: We thank the reviewer for noting this typo.

Action in the manuscript: This typo was corrected in the revised manuscript (line 196).

Comment: *I also inspected the code. It is very nice to see all the code and the dataset provided. I think the code is in a good state, overall. However, I think the authors should consider the following for improvements. Right now, the PL-MOGA algorithm requires their users to modify python code to implement custom optimization runs. I think that this is a relatively high entrance barrier for somebody with limited experience in python. I would suggest to at least allow for command line use, for instance via the argparse package. Similarly, I think both repositories would benefit immensely from documentation. Finally, as the code requires installation of a few additional packages, it would be highly beneficial for usability to add full installation instructions to the repository or its documentation.*

Reply: We thank the reviewer for making the effort of revising the code. We agree on all points mentioned, and we have thus implemented the command line use, revised the documentation (also including the command line use), and provided full installation instructions. The dataset documentation was also revised and it now refers to the Weisfeiler-Lehman graph hashes introduced upon the request made by another reviewer, as well as its availability from the Zenodo repository.

Action in the manuscript: The "Data and code availability" section has been re-written to provide a better description of the revised code repository, including the extensions resulting from the revisions (lines 425-433).

Reviewer 2

Comment: *This article, by Kneiding, Nova, Balcells uses multiobjective optimization on a diverse transition metal complex data set to simultaneously optimize polarizability and HOMO-LUMO gap. The authors use a genetic algorithm to find diverse TMCs on the Pareto front and analyze their chemical diversity. This article is very well written and is clearly reproducible. I commend the authors for their work and for making their code available. I believe the article is publishable as is, but would like to make some suggestions to improve the article.*

Reply: We thank the reviewer for the positive comments on our work.

Comment: *The authors introduce a large dataset that contains a synthesizable library of TMC ligands with charges extracted. Some other authors: https://pubs.acs.org/doi/epdf/10.1021/acs.jctc.2c00468 created a iterative method for charge determination that does not require any calculations. Does the authors' algorithm for charge determination have any benefit over something like this? I ask because removing the use of NBO for charge determination may help accelerate screening.*

Reply: We agree with the reviewer on the acceleration effect that removing the use of NBO would have on charge determination. However, we want to note that the NBO calculations also facilitated the definition of the metal coordination mode of the ligands. Further, we used the method mentioned by the reviewer to benchmark our approach, finding an agreement of 98.9% on the charges assigned to the ligands.

Action in the manuscript: The Methods section of the revised manuscript points to the method mentioned by the reviewer, including its agreement with our approach (lines 298-301). The data associated with this benchmark is mentioned in the Data and code availability section (lines 426-427) and is available from a .csv file at https://github.com/hkneiding/tmQMg-L/tree/main/benchmarks.

Comment: *How do the authors distinguish between compositional isomers of transition metal complexes with the same ligands arranged differently within the same coordination geometry? The example for the 4 coordinate cis- and trans- case is given in Figure 1B, but I am curious about octahedral complexes, where this problem can become more challenging. Clearly, the authors already have thought about this, as mentioned on page 5, but I would like to know how isomers are encoded in genes. Additionally, are HOMO-LUMO and polarizability sufficiently distinguished in two isomers of the same composition?*

Reply: Isomers were distinguished by the order of the ligands used to encode each TMC and considering the symmetries of the coordination geometry to ensure that all possible TMCs were generated and unique, excluding redundancies. We agree with the reviewer that encoding the octahedral coordination geometry and its isomers would be more involved though we think that it can also be implemented by adapting our approach to the square planar geometry. The property differences between isomers with the same composition can be small but also significant, depending on factors like the trans effect of the ligand and its connectivity.

Action in the manuscript: The encoding of the isomers is now mentioned in the Methods section of the revised manuscript (lines 311-312).

Comment: *On page 13, when the Tanimoto coefficient is measured, does the ordering of the SMILES matter? It appears that there should be some sort of rule, since the ligand SMILES are concatenated, else you may have different concatenations for the same complex.*

Reply: The ligands SMILES were concatenated as molecular fragments using the "." symbol. The resulting strings were then used to measure similarity leveraging the permutation invariance properties of both the Tanimoto coefficient and the Morgan fingerprints with which the former was computed.

Action in the manuscript: The permutation invariance underlying the computation of the Tanimoto coefficients, which was already mentioned in the SI, it is now also mentioned in the Methods section of the revised manuscript together with that of the Morgan fingerprints (399-401).

Comment: *Since the authors are releasing a dataset, they may be interested in the Weisfeiler-Lehman graph hash (detailed here: 10.1021/acs.jpclett.3c01214). The data set is useful as it is. However, having the Weisfeiler-Lehman graph hashes (both with and without atom number attribution) can help users filter by connectivity and composition uniqueness. This will make the data set even more useful.*

Reply: We thank the reviewer for noting the value of the Weisfeiler-Lehman graph hashes. We have now computed them, and they are included in the updated version of the tmQMg-L dataset provided with the revised manuscript.

Action in the manuscript: The inclusion of Weisfeiler-Lehman graph hashes in the tmQMg-L dataset is mentioned in the "Data and code availability" section of the revised manuscript, citing the reference mentioned by the reviewer (lines 429-431).

Comment: The code and data are reproducible. I suggest the authors upload the dataset on to a repository like Zenodo which has better version control for data sets.

Reply: We thank the reviewer for the positive comments on the code and data of our work and we agree on the appropriateness of using the Zenodo repository.

Action in the manuscript: The availability of the tmQMg-L dataset from Zenodo is mentioned in the "Data and code availability" section of the revised manuscript (lines 429-431).

Reviewer 3

Comment: *In my opinion the following issues should be addressed before the paper is suitable for publication. The authors write "In the 1.37M space, the MOGA located 130 TMC hits over the (α, ε) Pareto front with high chemical diversity and in an interpretable manner." But it's hard to judge whether that is a good performance. How many TMCs are there on the Pareto front? What percentage of these did the algorithm find?*

11

Reply: We thank the reviewer for noting this point, which was not presented with enough clarity in the original manuscript. The Pareto front of the 1.37M space has 30 dominating points, of which 18 were found by the MOGA thus giving a success rate of 60% and after exploring only 1% of the entire space. All other 112 TMCs in the last generation evolved by the algorithm were next to these dominating points and, therefore, within the Pareto front region.

Action in the manuscript: We revised the text of the "PL-MOGA algorithm" section to make this point clearer (lines 174-175).

Comment: The authors write "All these TMCs are present in the CSD, and, therefore, all ligands in tmQMg-L exist in at least one TMC that has been characterized experimentally, which shall enforce synthesizability in generative models." Well, that ensures the synthesizability of each individual ligand, but not of the TMC. It is not a given that when you mix the ligands and the metals that they will form the TMC the authors predict. As noted in this paper (10.1126/science.abj0999): "the factors that dictate nuclearity (e.g., monomer versus dimer), favored oxidation state, and ligation state of a catalyst are all too frequently barely understood". This should be noted.

Reply: We agree on this point, which is now noted in the revised manuscript, including the reference mentioned by the reviewer. In our work, the only way of controlling the synthesizability of the TMC is by tuning the degree to which the HOMO-LUMO gap is maximized, considering its correlation to stability and the fact that some TMC synthesis methods, like ligand exchange, are mostly under thermodynamic control.

Action in the manuscript: This point is now commented under the discussion section of the manuscript (lines 285-288).

Comment: The Lighthouse approach is presented assuming the objectives have positive values. While it is straightforward to extend it to negative values, I think it would be a little tricky for objectives that can be both positive and negative (e.g. logP values). Of course one could simply discard negative values if one wants to maximize the values, but what if positive values are rare and the starting population only has negative values. Since the paper introduces the Lighthouse, it would be good to have a discussion of this as it goes the general applicability of the method.

Reply: We think this point is both interesting and relevant, and we thus thank the reviewer for noting it. Our suggested solution is the following: Use an offset transforming the minimum negative fitness in any current population into a positive number and apply it to all its individuals. In this way, the relative distances in fitness space between the individuals will be conserved without altering the evolution behavior. In such framework, the offset must be recomputed and reapplied whenever new individuals enter the population. This implementation can be trivially simplified if there is prior knowledge on a lower bound of the fitness allowing to define a constant offset.

Action in the manuscript: We have added this point to the Methods section of the revised manuscript (lines 419-424) and it is also mentioned in the Discussion section which now discusses the general applicability of the method (lines 281-284).

<u>Comment</u>: *The authors write "MO with genetic algorithms (MOGAs) has been implemented with different methods,53–55 …". I suggest also citing this paper: 10.1021/acs.jcim.8b00839, which has been quite influential in the field.*

<u>Reply</u>: We thank the reviewer for noting this important reference.

<u>Action in the manuscript</u>: The reference mentioned by the reviewer has been added to the revised manuscript in the text introducing the use of genetic algorithms in molecular design (lines 52-54 citing reference 56).

Decision Letter, first revision:

Date:	16th January 24 13:59:20
Last Sent:	16th January 24 13:59:20
Triggered By:	Kaitlin McCardle
From:	kaitlin.mccardle@us.nature.com
To:	david.balcells@kjemi.uio.no
BCC:	kaitlin.mccardle@us.nature.com
Subject:	Decision on Nature Computational Science manuscript NATCOMPUTSCI-23-0697B
Message:	** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors. **

Dear Professor Balcells,

Your manuscript "Directional Multiobjective Optimization of Metal Complexes at the Billion-Scale with the tmQMg-L Dataset and PL-MOGA Algorithm" has now been seen by 3 referees, whose comments are appended below. You will see that while they find your work of interest, they have raised points that need to be addressed before we can make a decision on publication.

The referees' reports seem to be quite clear. Naturally, we will need you to address *all* of the points raised.

While we ask you to address all of the points raised, the following points need to be substantially worked on:

- Please add additional quantitative comparisons and experiments to address the points raised by Reviewer #1

In addition to these points, it would also be beneficial to address the following concerns:

- Please provide additional discussion to clarify the use of JT-VAE as a baseline. Additional experiments with other baselines are not strictly necessary.

You will also need to make some editorial changes so that it complies with our Guide

13

to Authors at https://www.nature.com/natcomputsci/for-authors .

In particular, I would like to highlight the following points of our style:

Nature Computational Science titles should give a sense of the main new findings of a manuscript, and should not contain punctuation. Please keep in mind that we strongly discourage active verbs in titles, and that they should ideally fit within 150 characters each (including spaces).

To improve the accessibility of your paper to readers from other research areas, please pay particular attention to the wording of the paper's abstract, which serves both as an introduction and as a brief, non-technical summary in about 150 words. It should include the background and context of the work, 'Here we show' or an equivalent phrase, and then the major results and conclusions of the paper. Because researchers from other sub-disciplines will be interested in your results and their implications, it is important to explain essential but specialised terms concisely. We suggest you show your summary paragraph to colleagues in other fields to uncover any problematic concepts.

We encourage you to archive the data reported in your manuscript in an accessible, persistent repository. If your data are archived prior to the acceptance of your manuscript, please provide us with the full citation as soon as you receive it so that a link to the data can be included in the publication. See http://www.nature.com/authors/policies/availability.html for more information.

If your paper is accepted for publication, we will edit your display items electronically so they conform to our house style and will reproduce clearly in print. If necessary, we will re-size figures to fit single or double column width. If your figures contain several parts, the parts should form a neat rectangle when assembled. Choosing the right electronic format at this stage will speed up the processing of your paper and give the best possible results in print. If you are in doubt about the correct format for your figures after reading our guidelines, please ask the art editors for advice computationalscience@nature.com.

Figure legends must provide a brief description of the figure and the symbols used, including definitions of any error bars employed in the figures.

As a guideline, Articles allow up to 50 references (excluding those cited exclusively in Methods).

Please include a statement before the Acknowledgements naming the author to whom correspondence and requests for materials should be addressed.

Finally, we require authors to include a statement of their individual contributions to the paper -- such as experimental work, project planning, data analysis, etc. -- immediately after the acknowledgements. The statement should be short, and refer to authors by their initials. For details please see the Authorship section of our joint Editorial policies at http://www.nature.com/authors/policies/authorship.html.

Please use the following link to submit your revised manuscript and a point-by-point response to the referees' comments (which should be in a separate document to any

cover letter):

[REDACTED]

** This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first. **

To aid in the review process, we would appreciate it if you could also provide a copy of your manuscript files that indicates your revisions by making use of Track Changes or similar mark-up tools. Please also ensure that all correspondence is marked with your Nature Computational Science reference number in the subject line.

In addition, please make sure to upload a Word Document or LaTeX version of your text, to assist us in the editorial stage.

To improve transparency in authorship, we request that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We hope to receive your revised paper within three weeks. If you cannot send it within this time, please let us know.

We look forward to hearing from you soon.

Best regards,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

Reviewers comments:

Reviewer #1 (Remarks to the Author):

The article "Directional Multiobjective Optimization of Metal Complexes at the Billion-Scale with the tmQMg-L Dataset and PL-MOGA Algorithm" is a revised version of a manuscript that I reviewed previously. I think the modified version is a significant improvement over the first submission. I would like to particularly highlight the efforts of the authors regarding the addition of a proper discussion and the improvements of the code base. However, in my opinion, the authors have not addressed several of my major concerns properly. Accordingly, I think that the manuscript still requires major revisions before it can be considered for publication. Please find my detailed comments below.

Major aspects:
The authors responded the following to my concern about the absence of sufficient benchmarks: "We agree with the reviewer on the importance of doing benchmark studies. However, we could not identify an alternative approach for transition metal complexes that can be directly compared to our method and thus be used as a benchmarking reference. Our coordination geometry-adapted, full-ligand genetic operations are fundamentally different from the ligand fragment approaches reported in the literature." I think the authors misunderstood what type of benchmarks I was referring to. Fundamentally, the algorithm is a standard genetic algorithm operating on a discrete design space. While the specific design space selected as main subject of study in this work might not have been studied before, there is nothing special about this design space mathematically compared to others. Hence, the performance of the vanilla genetic algorithm can be expected to be similar to previous work in other domains. However, what is added on top of this vanilla genetic algorithm is the so-called Pareto-Lighthouse component. Hence, in my opinion, the authors should compare the performance of a plain vanilla MOGA to their PL-MOGA approach on a standard multiobjective problem, which is not necessarily taken from chemistry. This is a topic that has been studied extensively in the literature and relevant benchmarks are available. In my opinion, this is important to demonstrate the effect of the PL component on the optimization performance. It is not clear to me whether the PL component is fundamentally different from using a fitness function that assigns certain regions of the original Pareto front with a higher fitness, and thus it might merely augment the fitness landscape rather than providing an actual advantage in terms of optimization performance. This is why I am not convinced that the metric the authors added regarding the number of identified Pareto points is meaningful. While the application of GFN-xTB-based methods might not be well explored yet for metal complexes, the use of this family of methods with genetic algorithms for organic molecules has been shown several times already and, thus, I do not consider this a very large conceptual advance. The author also state the following: " Further, we would like to note that the performance of our method is largely due to both the PL-MOGA algorithm and the tmQMg-L dataset. The combined use of these two elements delivers a significant advantage." How did the authors determine this advantage and how does it manifest in optimization performance? In my opinion, this also requires backing up in numerical benchmarks.

The authors state the following: "For example, relative to the junction tree variational autoencoder (JT-VAE) method, our method has the advantage of enabling generative tasks with ligands that are charged." I do not understand this statement. While the available pre-trained JT-VAE model might only generate uncharged molecules, there is nothing about the JT-VAE algorithm that makes generating charged ligands impossible. Additionally, I do not think that JT-VAE is an obvious baseline to select for discussing the advantages of the PL-MOGA algorithm. The more obvious baseline is comparison to a genetic algorithm that does not optimize over a well-defined ligand space but changes the constituent atoms of a ligand, for instance by operating on molecular graphs or molecular strings directly.

The authors state the following: "However, this task has been also suggested to be relevant in drug design where both stability and van der Waals interactions need to be maximized (Lilienfeld, Müller, Tkatchenko, Nat. Rev. Chem. 2020, 4, 347-358)." I do not think that this "hypothetical drug-design scenario" as the original authors

16

called it, is a very meaningful real-world design task. This is because the maximization of molecular polarizability is likely to lead to unselective binding which is typically not an effective mechanism for drug action. I might be wrong, however, it seems to me that it has not been demonstrated numerically that this is a meaningful design objective for real world molecular design. I do agree with the authors that it is a useful toy design objective, and it is excellent for benchmarking a particular inverse molecular design task, but I am not convinced that this is a design task relevant for realizing any functional molecules. It is true that the GFN-xTB simulations provides access to other properties, however, at least to my knowledge, none of them are immediately related to real-world design objectives either. Hence, I still think that this is a limitation of the current work.

Minor aspects:
The authors state: "We agree on this point. In order to provide a more systematic and yet intuitive metric for diversity, we computed the histogram of the SMARTS strings representing the chemical patterns of the metal-bound atoms in the tmQMg-L ligands." While I think that the histogram of the SMARTS strings is an improvement compared to the original version, I am still not convinced about its utility to quantify diversity. I think, in addition to the histogram, the authors should also compute a diversity metric based on Tanimoto similarity of the immediate coordination environment. In addition, and perhaps most importantly, I think the diversity should be compared to alternative datasets. Otherwise, it is hard to judge the diversity of the new ligand dataset compared to existing ones.

Reviewer #2 (Remarks to the Author):

The authors have addressed all of my concerns.

Reviewer #2 (Remarks on code availability):

The code is well structured and clear.

Reviewer #3 (Remarks to the Author):

The authors have addressed my concerns

Jan Jensen

Author Rebuttal, first revision:

Reference: NATCOMPUTSCI-23-0697B

Reviewer 1

Comment 1: *The article "Directional Multiobjective Optimization of Metal Complexes at the Billion-Scale with the*

tmQMg-L Dataset and PL-MOGA Algorithm" is a revised version of a manuscript that I reviewed previously. I think the modified version is a significant improvement over the first submission. I would like to particularly highlight the efforts of the authors regarding the addition of a proper discussion and the improvements of the code base. However, in my opinion, the authors have not addressed several of my major concerns properly. Accordingly, I think that the manuscript still requires major revisions before it can be considered for publication. Please find my detailed comments below.

Reply 1: We thank the reviewer for the effort of making these revisions and for noting the significant improvements already made. We think that this second set of revisions has also helped us to improve the manuscript further.

Comment 2: *The authors responded the following to my concern about the absence of sufficient benchmarks: "We agree with the reviewer on the importance of doing benchmark studies. However, we could not identify an alternative approach for transition metal complexes that can be directly compared to our method and thus be used as a benchmarking reference. Our coordination geometry-adapted, full-ligand genetic operations are fundamentally different from the ligand fragment approaches reported in the literature." I think the authors misunderstood what type of benchmarks I was referring to. Fundamentally, the algorithm is a standard genetic algorithm operating on a discrete design space. While the specific design space selected as main subject of study in this work might not have been studied before, there is nothing special about this design space mathematically compared to others. Hence, the performance of the vanilla genetic algorithm can be expected to be similar to previous work in other domains. However, what is added on top of this vanilla genetic algorithm is the so-called Pareto-Lighthouse component.*

Reply 2: We apologize for this misunderstanding. In this regard, we would like to clarify that if the vanilla genetic algorithm referred by the reviewer is our MOGA without the Pareto-Lighthouse component, the comparison between this algorithm and the PL-MOGA was already done and reported in the previous revision of the manuscript (lines 209-212 in the current version attached to this second revision).

Comment 3: *Hence, in my opinion, the authors should compare the performance of a plain vanilla MOGA to their PL-MOGA approach on a standard multiobjective problem, which is not necessarily taken from chemistry. This is a topic that has been studied extensively in the literature and relevant benchmarks are available. In my opinion, this is important to demonstrate the effect of the PL component on the optimization performance. It is not clear to me whether the PL component is fundamentally different from using a fitness function that assigns certain regions of the original Pareto front with a higher fitness, and thus it might merely augment the fitness landscape rather than providing an actual advantage in terms of optimization performance. This is why I am not convinced that the metric the authors added regarding the number of identified Pareto points is meaningful.*

Reply 3: Based on this comment of the reviewer, we compared our method to a GA algorithm in which the fitness was expressed as the weighted sum of the two objectives, a standard baseline in this field. We did these numerical experiments for the same chemical problem consisting in the multiobjective optimization of the HOMO-LUMO gap and the polarizability since, in agreement with the reviewer (Comment 7), we think that this is an excellent toy objective for the benchmarking of inverse design tasks. The results showed that the weighted-sum GA has a lower performance. In the balanced optimization putting equal (1.0, 1.0) weights on both objectives, it found only 8 dominating points (18 with our method). Since the reviewer was not convinced with this metric, which was suggested by another reviewer, we also considered the average Tanimoto coefficient (TC) of the last generation.

18

This metric also showed a lower performance of the weighted-sum GA, yielding TC = 0.44 (0.33 with our method); i.e. the Pareto solutions were less diverse. Further, a major limitation of the weighted-sum approach relative to the PL-MOGA is that the scope of the optimization cannot be fine-tuned. For example: the (1.0, 1.0) optimization yielded exactly the same results as the (0.1, 0.1), and therefore, it was not possible to narrow a center run. When the optimization was directed to the extremes of the Pareto front using (0.2, 0.8) and (0.8, 0.2) weights, the performance of the two algorithms was more similar, and yet the PL-MOGA was slightly better in most cases, finding more dominating points and yielding smaller TC values. Besides these results, it is also important to note another intrinsic limitation of the weighted-sum GA: one needs to know the numerical limits of the objectives in order to normalize them before the weights can be used. Importantly, this prior information, which, normally, will not be available in a real-world design task, is not required by the PL-MOGA algorithm.

Action in the manuscript 3: In the manuscript, the Methods (lines 417-420), Data and code (line 431), and SI (lines 438-439) sections refer to this benchmark. In the SI, we added a 3-page full new section entitled "Weighted-sum benchmark" (page S21) to describe these results. The associated code has been added to the GitHub page (https://github.com/hkneiding/PL-MOGA/tree/main/benchmark_1M).

Comment 4: *While the application of GFN-xTB-based methods might not be well explored yet for metal complexes, the use of this family of methods with genetic algorithms for organic molecules has been shown several times already and, thus, I do not consider this a very large conceptual advance.*

Reply 4: We thank the reviewer for this comment. The first revision of the manuscript added a mention to the efficiency of using an xTB fitness relative to a DFT one (line 276 in the current version attached to this second revision), but it did not present it as a very large conceptual advance. We also want to note that xTB errors for organic and transition metal systems can be significantly different. Further, the literature shows that the use of xTB in genetic algorithms optimizing metal complexes is incipient. Nonetheless, we acknowledge this comment of the reviewer since it made us realize that previous work on the use of xTB methods in genetic algorithms for organic molecules could be better referred to in the manuscript.

Action in the manuscript 4: In the revised "PL-MOGA algorithm" section of the manuscript, we make an explicit mention to the use of xTB methods in genetic algorithms for organic molecules (lines 165-166).

Comment 5: *The author also state the following: "Further, we would like to note that the performance of our method is largely due to both the PL-MOGA algorithm and the tmQMg-L dataset. The combined use of these two elements delivers a significant advantage." How did the authors determine this advantage and how does it manifest in optimization performance? In my opinion, this also requires backing up in numerical benchmarks.*

Reply 5: We apologize for this misunderstanding: The significant advantage mentioned in the point-by-point-reply of the first revision (not in the manuscript) did not refer to optimization performance but to the nature of the chemical space explorations that can be done with it. To clarify this point: Meanwhile the tmQMg-L dataset maximized the diversity and size of the associated chemical spaces, the PL-MOGA algorithm enabled their efficient exploration, finding a large number of Pareto solutions after screening only a small portion of the overall space. These features of the method were already mentioned in the revised Discussion section of the manuscript, without referring to optimization performance (lines 262-270 in the current version attached to this second revision).

Comment 6: *The authors state the following: "For example, relative to the junction tree variational autoencoder (JT- VAE) method, our method has the advantage of enabling generative tasks with ligands that are charged." I do*

not understand this statement. While the available pre-trained JT-VAE model might only generate uncharged molecules, there is nothing about the JT-VAE algorithm that makes generating charged ligands impossible. Additionally, I do not think that JT-VAE is an obvious baseline to select for discussing the advantages of the PL-MOGA algorithm. The more obvious baseline is comparison to a genetic algorithm that does not optimize over a well-defined ligand space but changes the constituent atoms of a ligand, for instance by operating on molecular graphs or molecular strings directly.

<u>Reply 6</u>: We agree on this point with the reviewer. Following the suggestion made in this comment, we replaced the reference to a variational autoencoder by a genetic algorithm operating on molecular representations at the atom level. Relative to this baseline, our method can add the advantage of pre-optimizing a multi-ligand TMC before focusing on the further evolution of a single selected ligand.

<u>Action in the manuscript 6</u>: The revised Discussion section of the manuscript discusses the advantage added by the PL-MOGA relative to the baseline mentioned by the reviewer (lines 278-281).

<u>Comment 7</u>: *The authors state the following: "However, this task has been also suggested to be relevant in drug design where both stability and van der Waals interactions need to be maximized (Lilienfeld, Müller, Tkatchenko, Nat. Rev. Chem. 2020, 4, 347-358)." I do not think that this "hypothetical drug-design scenario" as the original authors called it, is a very meaningful real-world design task. This is because the maximization of molecular polarizability is likely to lead to unselective binding which is typically not an effective mechanism for drug action. I might be wrong, however, it seems to me that it has not been demonstrated numerically that this is a meaningful design objective for real world molecular design. I do agree with the authors that it is a useful toy design objective, and it is excellent for benchmarking a particular inverse molecular design task, but I am not convinced that this is a design task relevant for realizing any functional molecules. It is true that the GFN-xTB simulations provides access to other properties, however, at least to my knowledge, none of them are immediately related to real-world design objectives either. Hence, I still think that this is a limitation of the current work.*

<u>Reply 7</u>: We still think that the joint optimization of the HOMO-LUMO gap and the polarizability can be useful as a high-level filter in a drug discovery funnel, in line with the publication referred to in our first reply. Drug candidates can hardly excel in a binding study without having a significant polarizability, and, further, they will anyway have limited commercial value if a narrow HOMO-LUMO gap makes them unstable. We agree with the reviewer on that this may not connect directly to a real-world design task in drug discovery; *i.e.* additional, more specific filters (*e.g.* binding optimization) would be needed down the funnel. This is now acknowledged in the revised manuscript. In line with this, other xTB properties, like reaction energies and barriers, can be highly relevant to design tasks like the optimization of a chemical process, though likely not immediately if this refers to make it happen in the lab. However, in practice, this is true for any design task with a minimal degree of complexity, for which an inverse design approach will always require multiple steps.

<u>Action in the manuscript 7</u>: The paragraph mentioning the limitations of our approach under the Discussion section has been revised and it now mentions that real-world design tasks would require the undertaking of additional optimization tasks (lines 286-287).

<u>Comment 8</u>: *The authors state: "We agree on this point. In order to provide a more systematic and yet intuitive metric for diversity, we computed the histogram of the SMARTS strings representing the chemical patterns of the metal- bound atoms in the tmQMg-L ligands." While I think that the histogram of the SMARTS strings is an improvement compared to the original version, I am still not convinced about its utility to quantify diversity. I think, in addition to the histogram, the authors should also compute a diversity metric based on Tanimoto similarity of*

the immediate coordination environment. In addition, and perhaps most importantly, I think the diversity should be compared to alternative datasets. Otherwise, it is hard to judge the diversity of the new ligand dataset compared to existing ones.

Reply 8: We think that the SMARTS histogram of the tmQMg-L dataset is an informative representation of diversity as it shows both the coordination environments and their counts in the dataset, in a visual manner. However, we agree with the reviewer that additional measures, like the Tanimoto coefficient, which was already used in other parts of the original manuscript, will enrich the discussion on diversity. We thus computed a heatmap reflecting the distribution of the Tanimoto coefficients between the different coordination environments. Further following the suggestions made by the reviewer, we computed both the histogram and the heatmap for another ligand dataset, the OctLig from the Kulik group, for the sake of comparing it to our tmQMg-L dataset. Interestingly, whereas the Tanimoto heatmaps suggest that both datasets are similar in terms of diversity, the SMARTS histograms show that tmQMg-L has the advantage of being more evenly distributed, with a progressive decay of the number of instances of the most common coordination environments, which, in addition, is in most cases larger in our dataset.

Action in the manuscript 8: The revised SI provides a comparison between the SMARTS histograms and Tanimoto heatmaps of the tmQMg-L and OctLig ligand datasets (page S4).

Decision Letter, second revision:

Date: 5th February 24 15:38:27

Last Sent: 5th February 24 15:38:27

Triggered By: Kaitlin McCardle

From: kaitlin.mccardle@us.nature.com

To: david.balcells@kjemi.uio.no

CC: computationalscience@nature.com

BCC: kaitlin.mccardle@us.nature.com

Subject: AIP Decision on Manuscript NATCOMPUTSCI-23-0697C

Message: Our ref: NATCOMPUTSCI-23-0697C

5th February 2024

Dear Dr. Balcells,

Thank you for submitting your revised manuscript "Directional Multiobjective Optimization of Metal Complexes at the Billion Scale" (NATCOMPUTSCI-23-0697C). It has now been seen by one of the original referees and their comments are below. The reviewer finds that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Computational Science, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

TRANSPARENT PEER REVIEW
Nature Computational Science offers a transparent peer review option for original research manuscripts. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please remember to choose, using the manuscript system, whether or not you want to participate in transparent peer review.**
Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

Thank you again for your interest in Nature Computational Science. Please do not hesitate to contact me if you have any questions.

Sincerely,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

ORCID
IMPORTANT: Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance: https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

Reviewer #1 (Remarks to the Author):

The article "Directional Multiobjective Optimization of Metal Complexes at the Billion Scale" is a revised version of an article that I reviewed twice. I think that the authors made considerable additional efforts and addressed my concerns acceptably, in particular with the addition of an explicit comparison of their PL-MOGA approach to an MOGA making use of a weighted sum of two objectives. Hence, I recommend this article for publication in the current state without further revisions.

Reviewer #1 (Remarks on code availability):

See previous review reports.

nature research

Final Decision Letter:	

Date: 29th February 24 11:55:27

Last Sent: 29th February 24 11:55:27

Triggered By: Kaitlin McCardle

From: kaitlin.mccardle@us.nature.com

To: david.balcells@kjemi.uio.no

BCC: kaitlin.mccardle@us.nature.com,rjsproduction@springernature.com,computationalscience@nature.com,fernando.chirigati@us.nature.com

Subject: Decision on Nature Computational Science manuscript NATCOMPUTSCI-23-0697D

Message: Dear Professor Balcells,

We are pleased to inform you that your Article "Directional Multiobjective Optimization of Metal Complexes at the Billion Scale" has now been accepted for publication in Nature Computational Science.

Once your manuscript is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

Please note that *Nature Computational Science* is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see https://www.nature.com/natcomputsci/for-authors). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

23

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

Once your manuscript is typeset, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

If you have queries at any point during the production process then please contact the production team at rjsproduction@springernature.com.

You may wish to make your media relations office aware of your accepted publication, in case they consider it appropriate to organize some internal or external publicity. Once your paper has been scheduled you will receive an email confirming the publication details. This is normally 3-4 working days in advance of publication. If you need additional notice of the date and time of publication, please let the production team know when you receive the proof of your article to ensure there is sufficient time to coordinate. Further information on our embargo policies can be found here:
https://www.nature.com/authors/policies/embargo.html

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Computational Science as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). We also welcome suggestions for the Hero Image, which appears at the top of our home page; these should be 72 dpi at 1400 x 400 pixels in JPEG format. Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Computational Science logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

We look forward to publishing your paper.

Best regards,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

P.S. Click on the following link if you would like to recommend Nature Computational Science to your librarian: https://www.springernature.com/gp/librarians/recommend-to-your-library

** Visit the Springer Nature Editorial and Publishing website at www.springernature.com/editorial-and-publishing-jobs for more information about our career opportunities. If you have any questions please click here.**