Donnelly Centre
Cellular & Biomolecular Research
UNIVERSITY OF TORONTO

UNIVERSITY OF TORONTO

Hannes L. Röst
Assistant Professor
The Donnelly Centre
Department of Molecular Genetics
University of Toronto

September 14, 2022

Re: Evaluation of PhD thesis – Michal Aleksander Ciach

To whom it may concern,

It is my pleasure to provide an assessment of the PhD thesis of Michal Aleksander Ciach “Algorithms for computational mass spectrometry based on the optimal transport theory” supervised by Dr. Anna Gambin at the University of Warsaw. I have read the complete thesis and provide the following assessment.

Overall, the work is academically rigorous and demonstrates the candidates expertise in the field which is sufficient for obtaining a PhD. The thesis is both well written and well structured, easy to follow, and the figures are clear and well described. The author describes his work on a novel approach to compare mass spectra and applies to both nuclear resonance spectra and mass spectra. They develop an approach based on the Wasserstein distance (termed masserstein) and apply it to both simulated and experimental spectra. The thesis has a suitable introduction, followed by four research contributions, where the author first uses optimal transport theory to compare mass spectra. The student then uses this metric to solve the problem of regression of mass spectra to quantify signals measured as continous distributions but modelled as discrete peaks. The approach is then extended to noisy spectra and finally applied to segmentation in imaging mass spectra. The majority of the work has been published in peer-reviewed journals with the main article being published in *Rapid Communications in Mass Spectrometry*, the leading journal for technical and algorithmic work in mass spectrometry.

In Chapter 3, the author presents their original work on optimal transport theory, which is a clear contribution to original research and places the new metric in context with existing methods and establishes the metric as a feasible metric that can outperform other similarity metrics. In Chapter 4 the thesis describes how to solve the posed problem using linear programming. The novelty lies in the formulation of the problem and the application of the Wasserstein distance. This is conceptually different from previous approaches which bin data and run ordinary least squares as here no binning is required and the associated problems with binning are elegantly circumnavigated. The chapter also describes several simulation experiments where the accuracy of the method is explored and a theoretical sections containing proofs. In Chapter 5, the author presents an extension to the approach

for noisy spectra which allows them to deal with real experimental spectra. They apply their approach to a set of experimental spectra and compare the results to manual integration where the method performs well. In Chapter 6, the author applies the algorithm to experimental data derived from mass spectrometric imaging of lipids. The results clearly show an improvement on both simulated and experimental data.

The discussion at the end of the thesis is somewhat short and I found it to be lacking to explore the impact of the work on the field and point out future directions. Specifically, the thesis currently does not compare the approach with other methods based on least squares regression that it claims to supplant and this is clearly a future direction that should be discussed. Similarly a discussion on pre-processing both by vendor software and users (including centroid vs profile mode) and its effects on the algorithm in question would be an area for discussion and further study.

I judge that the thesis is **sufficient to grant a PhD in its current form**. The thesis is of exceptional quality both in its breadth and mastery in areas spanning mathematics, computer science, chemistry and molecular biology. I therefore judge the thesis **exceptionally good and recommend a consideration for the honorary distinction.** I would recommend the student perform minor modifications including typographical fixes (attached) and an expansion of the final discussion.

Sincerely,

Dr. Hannes Röst