RICE

Wiess School of Natural Sciences
K.C. Nicolaou, Ph.D.
Department of Chemistry
Harry C. and Olga K. Wiess Professor of Chemistry

October 12, 2018

Re: Doctoral Thesis of Sara Szymkuć

Dear Committee Members:

It gives me great pleasure to write this brief evaluation of **Sara Szymkuć's** doctorate thesis under the supervision of Prof. Bartosz A. Grzybowski at The Institute of Organic Chemistry, Polish Academy of Sciences. Indeed, this is an extraordinarily strong thesis as reflected in the publications of the candidate based on her work included in it and the publicity they received.

Synthetic organic chemists practicing the art and science of organic synthesis in general, and total synthesis in particular, have been designing synthetic strategies for the synthesis of designed and naturally occurring molecules ever since the birth of this discipline as marked by the serendipitous of synthesis of urea by Frederick Wöhler in 1828. The design of such strategies evolved from serendipity to intuition, and then to logic. E.J. Corey's pioneering work in the 1960s turned the art of organic synthesis to science through retrosynthetic analysis. Despite Corey's efforts which included computer programs to generate retrosynthetic trees, the design of synthetic strategies remained almost exclusively the task of the synthetic chemist due to the lack of appropriate computer programs and sufficient computer power, until now. Admirably, this thesis contains the essence of a "quantum jump" in sophistication of such computer programs as encapsulated in its title, **"Teaching the Computer reactivity rules and strategies of automated retrosynthetic planning"** and the name ***Chematica*** platform for computer-assisted synthetic planning.

After a short introduction and historical background of the prior art, Sara proceeds to discuss the various problems and shortcomings of the previous programs and explains how she went about solving them achieving significant improvements that eventually led to Chematica. She first redefined chemical rules by introducing systematic machine-readable and controllable formats for taking into account all known synthetic reactions, reactivity conflicts, protection requirements, filters and speedy processing employing operations on strings rather than matrices. This required enormous efforts on her part for she had to learn, understand and creativity manipulate and assimilate so much and from so many disciplines in order to be able to form the web needed to produce translation of data bases to logical and optimal synthetic pathways toward targeted molecules. Taking into account feasibility issues she introduced scoring functions [Chemical Scoring Function (CSF) and Reaction Scoring Function (RSF)] for the various parameters involved in synthetic routes such as reactivity, availability, or ease of construction of starting materials and other factors that are usually considered by chemists. In other words, she systematically and thoroughly organized in her computer format what the human mind would normally think about, but not in such an organized and broad manner as her program.

The thesis also contains a number of other important elements derived from collaborative work with other scientists such as mathematicians, theorists and organic chemists leading to estimation of yields, strategy designs, handling of cyclization reactions, and other useful parameters. Most importantly, the candidate was able to derive from Chematica a number (eight) of synthetic strategies to pharmaceutically important molecules that were carried out and impressively validated in the laboratories of Sigma-Aldrich, Polish Academy and Northwestern University. Thus, resulting in practical syntheses these projects demonstrated cost savings in terms of time and money. These exemplary synthetic routes are included in the thesis with reagents, conditions and yields inspiring confidence and promising further improvements and uses of the platform. The string designations and branching of the routes as well as the color coding and pricing of intermediates in Figures 8 and 9 are

E-mail: kcn@rice.edu | Office: 713-348-8860 | Fax: 713-348-8865 | Rice University Department of Chemistry-MS 602 | BioScience Research Collaborative | Ste. 363 | 6100 Main Street | Houston, Texas 77005-1827 | www.chem.rice.edu

RICE

Wiess School of Natural Sciences
K.C. Nicolaou, Ph.D.
Department of Chemistry
Harry C. and Olga K. Wiess Professor of Chemistry

very useful in following the flow of steps leading to the targeted products from the defined starting materials and offer perspectives in terms of cost and convenience.

The main part of her thesis is supported with reprints of her publications and other addendums resulting in an impressive document of high-quality scientific work. Furthermore, Sara has a vision as to where she wants to go and apply her gained knowledge and expertise in the future, a sign of a great career ahead of her.

In summary, Sara is to be congratulated and commended for her all-around contribution to the emergence of Chematica, a powerful tool that is bound to assist synthetic organic chemists in designing road maps to their target molecules. The precision and accuracy of Chematica should, of course, be continuously updated and improved in the age of increasingly more powerful computers, algorithms and big data.

My congratulations to Sara for an outstanding thesis and the wonderful way in which she wrote it in perfect English, and with admirable generosity in giving credit to her mentor and collaborators.

With my very best wishes,

Cordially,

K.C. Nicolaou, Ph.D.

E-mail: kcn@rice.edu | Office: 713-348-8860 | Fax: 713-348-8865 | Rice University Department of Chemistry-MS 602 | BioScience Research Collaborative | Ste. 363 | 6100 Main Street | Houston, Texas 77005-1827 | www.chem.rice.edu