Prof. Matthias Bochtler
International Institute of Molecular and Cell Biology (IIMCB)
Institute of Biochemistry and Biophysics (IBB)
Warsaw, Poland

August 12th, 2016

Opinion on the PhD thesis by Małgorzata Szeląg

Małgorzata Szeląg has presented a PhD thesis entitled: "*In silico* comparative structural and functional analysis of STAT and IRF proteins to identify specific inhibitory compounds". The thesis was prepared under supervision of Prof. Hans Bluyssen and Dr. Anna Czerwoniec and is presented in the new form, containing an introduction to the topic, followed by the co-author statements of their contributions to the publications in the thesis, and finally the five publications representing the work summarized in the thesis.

1. Szelag, M., Sikorski, K., Czerwoniec, A., Szatkowska, K., Wesoly, J. and Bluyssen, H.A.R. (2013) In silico simulations of STAT1 and STAT3 inhibitors predict SH2 domain cross-binding specificity. *Eur J Pharmacol*, **720**, 38-48.
2. Szelag, M., Czerwoniec, A., Wesoly, J. and Bluyssen, H.A.R. (2014) Comparative screening and validation as a novel tool to identify STAT-specific inhibitors. *Eur J Pharmacol*, **740**, 417-420.
3. Czerwoniec, A., Szelag, M., Juszczak, K., Wesoly, J. and Bluyssen, H.A.R. (2015) CAVS-Novel in silico selection strategy of specific STAT inhibitory compounds. *J Comput Sci-Neth*, **10**, 186-194.
4. Szelag, M., Czerwoniec, A., Wesoly, J. and Bluyssen, H.A.R. (2015) Identification of STAT1 and STAT3 Specific Inhibitors Using Comparative Virtual Screening and Docking Validation. *Plos One*, **10**.
5. Szelag, M., Piaszyk-Borychowska, A., Plens-Galaska, M., Wesoly, J. and Bluyssen, H.A. (2016) Targeted inhibition of STATs and IRFs as a potential treatment strategy in cardiovascular disease. *Oncotarget*.

The first four papers deal with inhibition of JAK (Janus kinase) STAT (signal transducer and activator of transcription) signaling. They describe the (mostly computational) characterization of known STAT inhibitors, and the identification of potential new inhibitors. There is substantial (perhaps even large) overlap between the first four publications. The first publication deals with STAT specificity. The second and third publications describe the computational pipeline. The forth publication is primarily a description of a screen for new inhibitors (although some known compounds are discussed as well). Publication 5 focuses on inhibition of STATs and IRFs (interferon regulatory factors) specifically for cardiovascular disease and is (mostly) a review.

Prof. Matthias Bochtler
International Institute of Molecular and Cell Biology (IIMCB)
Institute of Biochemistry and Biophysics (IBB)
Warsaw, Poland

JAK-STAT signaling requires four main components: an extracellular ligand, a receptor, the JAK kinase, and the STAT transducer protein. JAK and STAT proteins are intracellular proteins. Upon binding of the ligand to the receptor, the JAK kinases, physically associated with the receptors irrespective of ligand binding, phosphorylate the intracellular domains of the receptors. This receptor phosphorylation recruits the STAT proteins, due to their SH2 domains, to the intracellular domains of the receptors, and hence brings them into proximity with the JAK kinases, leading to STAT phosphorylation. Phosphorylated STATs dimerize. STAT dimers translocate to the nucleus and act as transcriptional activators, eventually affecting processes including immunity, cell proliferation, and oncogenesis, but paradoxically also differentiation and apoptosis.

The JAK-STAT signaling pathway is highly modular. Three main pathways are distinguished: type I interferon pathways (interferons α and β) acting through STAT1 and STAT2, a type II interferon pathway (interferon γ), acting through STAT1, and a STAT3 dependent pathway stimulated by various ligands and involved in multiple developmental and proliferative decisions. Altogether, humans have four JAK kinases (JAK1, JAK2, JAK3, and TYK2), and seven STAT proteins. The STAT proteins can form homodimers, as well as heterodimers. The combinatorial complexity of JAK-STAT signaling creates opportunities and challenges in targeting JAK-STAT pathways.

"General" questions:

- The clinical success of JAK kinase inhibitors (Ruxolitinib/ Jakafi/Jakavi, against JAK1/JAK2 for psoriasis, myelofibrosis, and rheumatoid arthritis, Tofacitinib/Xeljanz/Jakvinus for myelofibrosis) provides some reason to believe that STAT inhibitors could be useful. However, the author provides no clear rationale for the benefit of *selective* STAT inhibitors. For which indications would selective STAT1 or STAT3 inhibitors be superior to a non-selective STAT inhibitor?
- SH2 domains are THE detector domains for phosphotyrosine residues. I suspect that there are SH2 domains (in proteins other than STATs) that bind phosphotyrosine residues followed by a hydrophobic residue. Are the SH2 domains of human STATs more similar to each other than to other human SH2 domains? In other words: for a STAT3 inhibitor, are the other STATs really the main anti-targets? Especially since the STATs seem to act rather synergistically than antagonistically to each other?
- The author concludes that STATTIC inhibits the STATs with comparable efficiency. At least for the most highly scored binding mode this conclusion is not convincing. In the PLOS ONE paper (Fig. 2) the sulfonate oxygen atoms (with partial negative charge) are placed in a negatively charged environment

Prof. Matthias Bochtler
International Institute of Molecular and Cell Biology (IIMCB)
Institute of Biochemistry and Biophysics (IBB)
Warsaw, Poland

in STAT1, and in a positively charged environment in STAT3. It is difficult to see why the two binding modes should be equally favorable. Do the authors have a qualitative explanation?

- The author validates comparable binding of STATTIC to STATs by monitoring their phosphorylation. The assay is based on the use of phospho-specific STAT antibodies. The rationale of this assay is not well explained in the publications or the thesis. Presumably blocking SH2 domains blocks recruitment of STATs to phosphorylated receptors, and hence to the JAKs, and thus ultimately phosphorylation of the STATs themselves. Is this the authors' interpretation of the data?
- Are the statements regarding the specificity of the antibodies only based on the manufacturer's statements, or have they been checked by independent own controls?
- Fludarabine is an adenosine analogue, which is likely phosphoylated *in vivo* (to the mono-, di-, and triphosphates, as considered by the authors). Does fludarabine acts as a JAK kinase inhibitor? Could the profile of inhibition of phosphorylation be explained by the specificity of JAK kinases? In my opinion, controls for the activity of JAK kinases are missing. Or have other authors already ruled out the possibility of JAK kinase inhibition by fludarabine?

Technical issues:

- At least one of known STAT inhibitors described in the work (curcumin) has been described as having a large number of targets other than STATs (see J. K. Lin, Adv Exp Med Biol. 2007;595:227-43, we happen to have checked other authors' claims that curcumin as a methyltransferase inhibitor). How many of the studied compounds (of disclosed identity) are specific STAT inhibitors and not just molecules that appear to be "sticky" and therefore target a wide variety of different proteins?
- The authors invest great care to model the entire STAT proteins. However, they then only investigate binding to the SH2 domain. Is the structure of the other domains at all relevant for the question specificity of the SH2 directed inhibitors?
- The authors create a "phylogenetic" tree of the STAT proteins based on manually curated alignments (Fig. 1 of the PLOS ONE publication). How has the tree been routed? Was an outgroup used? Bootstrap values have been calculated according to Methods, but I could not find them to validate the tree.
- Are global pairwise similarities between STATs really the best measure of pharmacological similarities of the SH2 pockets?
- The authors use rather high level theory (Gaussian, large orbital sets , DFT, vibrational frequencies to check for the ground state) to model small molecules, but then use rather crude docking (just three ligand poses in initial

Prof. Matthias Bochtler
International Institute of Molecular and Cell Biology (IIMCB)
Institute of Biochemistry and Biophysics (IBB)
Warsaw, Poland

screens) to search for potential binders. Is this curious combination really the best use of computational resources?

- The authors state affinities as -log(Kd). I presume they mean -log(Kd/M). If so, then the calculated affinities are in the range of 10 to 100 micromolar. Is this not much too low for pharmacologically useful compunds? Gleevec, famous for being a useful drug despite the bad Kd, is the only useful drug targeting a (mostly) a single protein with such weak affinity I know of. And unlike STAT inhibitors, Gleevec does not have to compete with another protein!

Policy issues:

- The authors are very generous in sharing their computational protocols and code (in a very professional way, via Git-hub), so that the work is definitely useful for computational chemists. However, in the PLOS ONE paper, the authors appear to withhold the identity of compounds that they have found in their computational screen. Instead of chemically meaningful compound names or structural formulas, only uninformative acronyms appear to be given. Is there a decoding table somewhere that I have missed? If not, this is in my eyes a major reservation against the work. If the real compound identities are indeed withheld, the work is essentially non-reproducible. I understand commercial interests, but these can be protected by patenting compounds prior to publication.

Despite some reservations expressed above, there is no doubt that work in five publications that have passed muster in peer review justifies the award of a PhD and therefore I recommend to proceed with the award of a doctoral degree.

Matthias Bochtler