Review of: "NeuralTPL: a deep learning approach for efficient reaction space exploration"

Dragos Horvath[1]

1 University of Strasbourg

I am sorry, but after reading this text I have more questions than answers. You are speaking about "template space" and you define a "template" as (cite) 'A template T is defined as the reverse transformation from the product's reaction center set p to the reactant's reaction center set r." But... what is a "reaction center set"? That concept is never defined! For me, a reaction template should include the MINIMAL substructure containing ALL the dynamical bonds (which are made, broken or have their bond order altered). But... when I look at Figure 2, it's not what I see! Why do you need a SPECIAL template for oxalic acid N-acylating hydroxylamine, rather than writing THE general acylation template $C(=O)N \rightarrow C(=O)LG+N$? Of course, the immediate environment in which the template is embedded DOES play a key role (and sometimes the remote environment matters a lot too... if there is conjugation, like in p-nitroaniline!) But the rules by which these templates were extracted are obscure to me!

Now - my other key problem with this text is that I can't see its rationale: what do you need deep learning FOR, if the only question you ask is to find the templates concerning a given atom in a given product? A good old set of clever SMARTS would very much do the job: you do not need to be Nostradamus to recognize an amide bond and to predict that it can be obtained by acylation! Same, if you see a cyclohexene, you see Diels-Alder, etc. The real problem in computer-aided synthesis is not recognizing the possible breaking points of the product. RECAP rules are already there for this! The key point is not to realize that this amide bond COULD be targeted in synthesis, but WHETHER it SHOULD be targeted: is the N nucleophilic enough? Does the building block have other nucleophilic functional groups which may render acylation of the targeted N prone to side reaction (and if so, is there an intelligent protection/deprotection strategy to circumvent this?). What is the most cost-effective acylating agent to use? $C(=O)Cl$ will likely work - but I can't get my hands on it - should I use anhydride? Would acid and DCC or other water-scavenging agents do? Reaction planning is the difficult problem, and it's reaction condition-prone: thus multifactorial. There is no progress done with respect to the real hard problems here. I do not see how the approach presented here can be more useful than the standard use of Reaxys: draw my process, fish out similar published reactions based on the same "template" and get inspired by those...

Last but not least - please do NOT use the acronym CASP: it is already taken, for years, by protein folders: it stands for Critical Assessment of Structure Prediction and represents well-known global challenges in which protein folders compete to predict the structures of proteins and other (macro)molecules and

complexes with structures solved, but not yet publicly released. The community would be very angry to see you "pirating" their fetish acronym!