

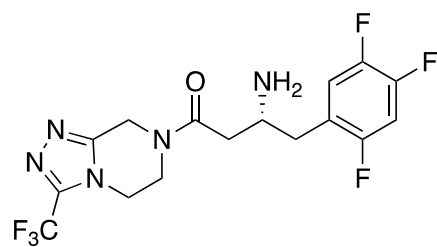
## Supporting Information

# Computer-assisted multistep chemoenzymatic retrosynthesis using a chemical synthesis planner

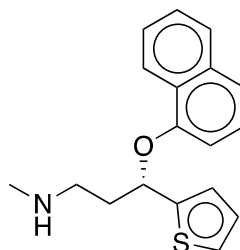
Karthik Sankaranarayanan and Klavs F. Jensen\*

Department of Chemical Engineering, Massachusetts Institute of Technology; 77  
Massachusetts Avenue, Cambridge, Massachusetts 02139, United States.

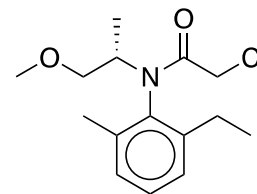
\*kfjensen@mit.edu



Sitagliptin



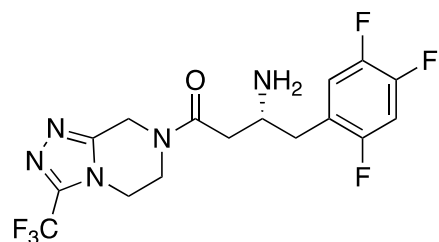
Duloxetine



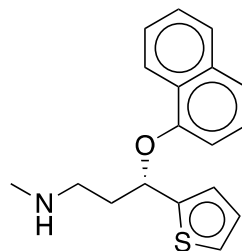
S-Metolochlor

	Sitagliptin		Duloxetine		S-Metolochlor	
Method Employed	Levin <i>et al.</i>	Coley <i>et al.</i>	Levin <i>et al.</i>	Coley <i>et al.</i>	Levin <i>et al.</i>	Coley <i>et al.</i>
Total # of chemo-enzymatic pathways	1	N/A	6	N/A	11	N/A
Total # of chemical pathways	2	36	2	241	6	500
Chemo-enzymatic synthesis precedent	1,2	N/A	3	N/A	4	N/A
Is literature precedent captured	No	N/A	No	N/A	No	N/A

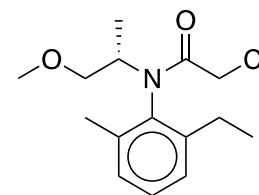
**Figure S1:** Demonstration of the current state of the art technique for chemoenzymatic synthesis planning on model compounds that are chemically dissimilar from natural products.<sup>5</sup> As a control, we use the chemical retrosynthesis planner developed by Coley *et al.* under similar settings.<sup>6</sup> 'N/A' stands for Not Applicable.



Sitagliptin



Duloxetine



S-Metalochlor

Method Employed	Sitagliptin		Duloxetine		S-Metalochlor	
	Levin <i>et al.</i>	This study	Levin <i>et al.</i>	This study	Levin <i>et al.</i>	This study
Total # of chemo-enzymatic pathways	1	49	6	195	11	31
Total # of chemical pathways	2	8	2	65	6	469
Chemo-enzymatic synthesis precedent	1,2		3		4	
Is literature precedent captured	No	Yes	No	Yes	No	Yes

**Figure S2: A comparison of our newly developed method with the current state of art technique developed by Levin *et al.*<sup>5</sup> Three model compounds Sitagliptin, Duloxetine, and (S)-Metalochlor were selected because they are man-made molecules with low chemical similarity to natural products.**

### Intermediates removed from buyables database

We removed all intermediates in the literature pathways presented in Figure 4 from our buyable database to prevent premature termination at expensive starting materials. The following list of SMILES strings were removed from the buyables database:

1. N/C(=C\C(=O)N1CCn2c(nnc2C(F)(F)F)C1)Cc1cc(F)c(F)cc1F
2. O=C(CC(=O)N1CCn2c(nnc2C(F)(F)F)C1)Cc1cc(F)c(F)cc1F
3. CC1(C)OC(=O)C(=C(O)Cc2cc(F)c(F)cc2F)C(=O)O1
4. CCN(C)C(=O)Oc1cccc([C@@H](C)OS(C)(=O)=O)c1
5. CCN(C)C(=O)Oc1cccc([C@@H](C)O)c1
6. N[C@@H](Cc1ccccc1Br)C(=O)O
7. CC(=O)[C@H](O)c1ccccc1
8. N#C[C@H](O)c1ccccc1Cl
9. N[C@@H]1CCc2ccccc21
10. O[C@H]1CCc2ccccc21
11. O[C@H]1CCc2ccccc21
12. N[C@@H]1CCc2ccccc21
13. CN(C)CC[C@H](Oc1cccc2ccccc12)c1cccs1
14. CN(C)CC[C@H](O)c1cccs1
15. CN(C)CCC(=O)c1cccs1

## References

- 1F. Ye, Z. Zhang, W. Zhao, J. Ding, Y. Wang and X. Dang, *RSC Advances*, 2021, **11**, 4805–4809.
- 2C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman and G. J. Hughes, *Science*, 2010, **329**, 305–309.
- 3X. Chen, Z.-Q. Liu, C.-P. Lin and Y.-G. Zheng, *Bioorganic Chemistry*, 2016, **65**, 82–89.
- 4D. A. Jackson, in *Biocatalysis for Green Chemistry and Chemical Process Development*, John Wiley & Sons, Ltd, 2011, pp. 255–276.
- 5I. Levin, M. Liu, C. A. Voigt and C. W. Coley, *Nat Commun*, 2022, **13**, 7747.
- 6C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison and K. F. Jensen, *Science*, 2019, **365**, eaax1566.