

Supplementary Information

Explaining and avoiding failure modes in goal-directed generation of small molecules

Maxime Langevin^{1,2}
, Rodolphe Vuilleumier²
and Marc Bianciotto^{1*}

*Correspondence:

marc.bianciotto@sanofi.com

¹Molecular Design Sciences -
Integrated Drug Discovery, Sanofi
RD, 94400 Vitry-sur-Seine, France
Full list of author information is
available at the end of the article

Samples from the DRD2 dataset

Table S1 shows several examples of questionable active compounds retrieved from the DRD2 dataset. Those samples suggest that the detected activities on the assay (radioactive ligand binding competition assay) might come from different mechanisms. This heterogeneity, coupled with the small size of the dataset, could explain the observed difference between optimization and control scores.

Samples generated on the ALDH1 dataset

Figure S1, Figure S2 and Figure S3 display final samples generated by the three goal-directed algorithms. Even when optimization and control scores do not diverge, issues concerning the quality of molecules generated remain.

Samples generated on the JAK2 dataset with modified predictive model

Figure S4, Figure S5 and Figure S6 display final samples generated by the three goal-directed algorithms. Even when optimization and control scores do not diverge, issues concerning the quality of molecules generated remain.

Model-control score and optimization score comparison.

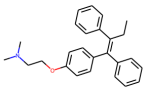
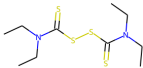
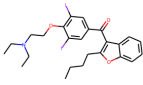
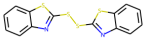
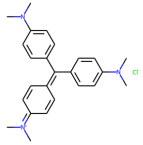
Author details

¹Molecular Design Sciences - Integrated Drug Discovery, Sanofi RD, 94400 Vitry-sur-Seine, France. ²PASTEUR, Département de chimie, École Normale Supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France.

References

1. Tredici, A.L.D., Ma, J.-N., Piu, F., Burstein, E.S.: Identification of the antiarrhythmic drugs amiodarone and lorainide as potent h3 histamine receptor inverse agonists. *Journal of Pharmacology and Experimental Therapeutics* **348**(1), 116–124 (2013). doi:[10.1124/jpet.113.208892](https://doi.org/10.1124/jpet.113.208892)

Table S1 Selection of questionable active compounds from the DRD2 dataset

Name	Structure	Comment
Tamoxifen		Selective estrogen receptor modulator
Disulfiram		Covalent inhibitor of the acetaldehyde deshydrogenase, reactive disulfide bond
Amiodarone		Anti-arrhythmic drug, found active on H3 receptor but not to D2 [1]
Benzothiazyl Disulfide		Contact allergen, features a reactive disulfide bond
Gentian Violet		Dye, known toxic to the CHO cells used in the assay

