

Synthons as common ground: unifying SAR decomposition and building block search for libraries design at Lilly

Yuliana Zabolotna¹, Eduardo Harguindey¹, Liying Zhang², Alfredo Martin^{3*}

¹Eli Lilly, Small Molecule Drug Design, Chemoinformatics and Applied AI for Early Molecule Design, Madrid, Spain

²Eli Lilly, Small Molecule Drug Design, Chemoinformatics and Applied AI for Early Molecule Design, Boston, United States

³Eli Lilly, Small Molecule Drug Design, Computational Chemistry for Early Molecule Design, Madrid, Spain

Drug-design projects routinely require region-focused SAR decomposition of tested compounds and navigation of available building-block (BB) space under project-specific reaction and selection constraints. In current practice these tasks run on different representations: SAR analysis on enumerated final molecules, BB searches at the reagent level. We describe two deployments of a shared synthon-level framework that bridges this gap across advanced (AMD) and early (EMD) molecule design.

The first deployment, SAR-Gate, couples region-wise SAR decomposition with synthon-anchored BB-space navigation:

- an additive-contribution engine fragments project molecules at the synthetically accessible bonds and then interpolates activity/properties within explored chemical space of each region;
- a synthon generation engine decomposes BBs into reactive fragments directly comparable to those derived from tested compounds

The latter is conceptually related to Synth-On^[1] while built on the reaction knowledge base and code architecture of eDesigner^[2] and other unpublished related internal Lilly tools. Configured with project preferences and constraints, SAR-Gate is designed to generate routine synthesis proposals for both exploration of unexplored chemical matter (scaffold hopping, ring replacements) and exploitation of established series (synthon-based QSAR) which could be applied for both small molecule and non-classical modalities drug discovery (peptides, DEL, ionizable lipids for LNPs etc). Because the reaction knowledge base is HTE-compatible, SAR-Gate output is directly actionable in automated synthesis and direct-to-biology screening, positioning both engines as potential components of agentic laboratory pipelines.

The second deployment, Lly-BBToolKit, operates on project-agnostic synthon libraries pre-generated by SAR-Gate, exposing readily available BBs organized by reaction compatibility to medicinal chemists through an agentic conversational layer with a controlled action space. The conversational layer includes an agent for diversity selection, making Lly-BBToolKit most valuable in early-stage projects with limited SAR, where diverse exploration enables hit expansion and generating the initial signal that SAR-Gate can later refine

Bibliography :

[1] Zabolotna, Y.; Volochnyuk, D.M.; Ryabukhin, S.V.; Gavrylenko, K.; Horvath, D.; Klimchuk, O.; Oksiuta, O.; Marcou, G.; Varnek, A. *J. Chem. Inf. Model.* 62 (2022) 2151–2163. DOI: 10.1021/acs.jcim.1c00754

[2] Martin, A.; Nicolaou, C.A.; Toledo, M.A. *Commun. Chem.* 3 (2020) 127, DOI: 10.1038/s42004-020-00374-1