

Specifikace projektu Molpher

Základní informace:

Vedoucí: RNDr. David Hoksza, PhD – KSI MFF UK (hoksza@ksi.mff.cuni.cz)

Konzultant: Mgr. Daniel Svozil, PhD – LICH VŠCHT (daniel.svozil@gmail.com)

Počet studentů: 5 - 6

Motivace:

Mnoho oblastí chemické biologie, jako je např. výzkum léčiv, intenzivně využívá knihovny známých chemických sloučenin. Látky v těchto knihovnách jsou systematicky zkoumány za účelem nalezení sloučenin, které vykazují biologickou aktivitu - tj. buď daný biologický jev urychlují, nebo zpomalují. Biologicky aktivní molekuly se potom stávají klíčovými hráči v procesu identifikace nových léčiv.

Množinu všech možných chemických sloučenin nazýváme chemickým prostorem. Vzhledem k jeho enormní velikosti je nereálné testovat jejich biologické aktivity nejen experimentálně, ale i in silico. Proto je třeba vytipovat pouze podmnožinu prostoru, která by mohla být pro danou aplikaci zajímavá. Takovou podmnožinu prostoru nazýváme *cílenou knihovnou (focused library)* a generování těchto cílených knihoven je náplní projektu.

Představme si (zjednodušenou) situaci, kdy jsou k dispozici 2 chemické sloučeniny A a B (léčiva) působící na 2 různé cíle (tzv. receptory). Úkolem je pak např. navrhnout strukturu C , která bude vykazovat aktivitu vůči oběma receptorům zároveň. Dá se předpokládat, že výsledná sloučenina C bude v jisté míře obsahovat vlastnosti obou sloučenin A a B . Tj. v chemickém prostoru se bude C vyskytovat "mezi" A a B . Místo prohledávání celého chemického prostoru je tedy možné vygenerovat sloučeniny nacházející se v podprostoru, na který je možno nahlížet jako na určitý druh cesty, vymezeného molekulami A a B . Kroky na cestě mezi A a B lze definovat pomocí tzv. morfovacích operátorů, jak bylo ukázáno v [Hoksza2011]. Takto nalezené chemické struktury lze pak následně podrobit detailnějšímu testování na jejich biologickou aktivitu.

Popis projektu:

Cílem projektu je vytvoření aplikace umožňující cílené procházení chemického prostoru mezi startovací a cílovou sloučeninou a generování fokusovaných knihoven. Projekt by měl umožnit

- *Vizualizovat chemický prostor*, tj. uživateli bude umožněno vidět, kterou částí prostoru nalezená cesta prochází.

- *Preferovat jisté části prostoru při procházení.* Budou-li uživatele zajímat jisté části prostoru více než jiné, může definovat tzv. *decoy set* tj. množinu sloučenin ke kterým by se cesta měla přiblížit. Např. uživatel ví, že v dané části se nachází sloučeniny se specifickými vlastnostmi, které předpokládá u výsledné sloučeniny.
- *Logovat průchod prostorem.* Vzhledem k tomu, že dříve než bude možné experimentálně ověřit funkci vytipovaných sloučenin, bude potřeba množinu nalezených sloučenin podrobit dalšímu zkoumání (např. zjištění syntetizovatelnosti navržených sloučenin)výstup aplikace musí projít dalšími kroky (např. zjištění syntetizovatelnosti navržených sloučenin), než je možné prakticky vyzkoušet funkci dané sloučeniny, je třeba veškeré sloučeniny testované v průběhu průchodu prostorem logovat a dát uživateli efektivní nástroje pro filtrování těchto logů..
- *Paralelizovat průchod prostorem.* Algoritmy pro průchod velkými prostory jsou vysoce výpočetně náročné, a proto je třeba implementaci algoritmu paralelizovat, aby bylo umožněno v relativně krátkém čase vyzkoušet mnoho dvojic start-cíl.
- *Dávkové zpracování.* Předpokládá se nasazení projektu v prostředí, kde Molpher bude pouze jeden díl procesu a proto je třeba umožnit jeho efektivní využití za pomoci dávkového zpracování (týká se jak zadávání vstupních sloučenin, tak parametrizace výpočtu).
- *Interaktivita procesu procházení prostorem.* Uživatel by měl mít možnost jak aktivně vstupovat do procesu procházení prostoru, tak mít nástroje k modifikaci výsledku. Tzn., že uživatel bude moci online sledovat práci algoritmu s možností jeho pozastavení případně “nasměrování”. Dále musí mít uživatel možnost po skončení procesu rozšířit výslednou množinu o lokální okolí sloučenin na cestě. Když tedy některé vygenerované sloučeniny vypadají zajímavě, uživatel musí mít možnost prozkoumat i jejich lokálního okolí (podobné sloučeniny).

Složení týmu:

- Vizualizace – 1
- Algoritmizace (průchod prostorem) – 1-2
- Paralelizace (efektivita průchodu prostorem) – 1
- GUI, dávkové zpracování, logování – 1-2

Další informace:

- Jazyk aplikace není pevně dán, ale předpokládá se C++
- Projekt nevyžaduje specifické znalosti biologie nebo chemie
- Projekt bude vznikat ve spolupráci s VŠCHT a AVČR

[Hoksza2011] D. Hoksza, D. Svozil. Exploration of Chemical Space by Molecular Morphing. accepted to IEEE 11th International Conference on Bioinformatics & Bioengineering (BIBE 2011)