



Barbora Kozlíková  
Assistant Professor  
Department of Computer Graphics and Design

Masaryk University, Faculty of Informatics  
Botanická 68a, 602 00 Brno, Czech Republic  
kozlikova@fi.muni.cz, decibel.fi.muni.cz

Brno, 12. 6. 2016

## Posudek oponenta bakalářské práce

Student: Tom Jankovec  
Vedoucí práce: Ing. Vojtěch Vonásek  
Oponent práce: RNDr. Barbora Kozlíková, Ph.D.

Předložená bakalářská práce studenta Toma Jankovce se zabývá problémem detekce tunelů v proteinech pomocí algoritmu hledání cest. V rámci práce pak navrhuje a implementuje rozšíření algoritmu Rapid-exploring Random Trees o možnost detekce více tunelů ve statické molekule. Práce je dále doplněna o výsledky srovnání řešení s existující aplikací CAVER 3.0.

Textová část práce je dle mého názoru velmi kvalitně zpracována. Oceňuji i volbu jazyka, který je na velmi dobré úrovni. Rovněž rozsahem je práce nadstandardní. Text je přehledně členěn do šesti kapitol. Po úvodním seznámení s problematikou a představení nejpoužívanějších stávajících přístupů se student zaměřuje na popis přístupu pomocí plánování cest a jeho vztah k detekci tunelů v proteinech. Ve třetí kapitole pak navrhuje vlastní řešení. Čtvrtá kapitola stručně seznamuje s možností rozšíření algoritmu pro hledání tunelů s ohledem na ligandy se složitějším tvarem. V páté kapitole student nejdříve zmiňuje několik implementačních detailů a poté prezentuje výsledky porovnání navrženého TOM-RRT algoritmu s nástrojem CAVER 3.0. Toto porovnání je provedeno na pěti proteinových strukturách. Po zhodnocení dosažených výsledků následuje závěrečné shrnutí.

K textu práce nemám z obsahového hlediska větších výtek. Oceňuji velké množství ilustračních obrázků, které dobře doplňují text. Často jsou však tyto obrázky uvedeny dříve, než jsou odkazovány v textu, což zhoršuje čitelnost. Velmi pozitivně cením zejména obrázek 1.6, který velmi přehledně seznamuje jak s celým přístupem, tak se strukturou stěžejní části práce.

Komentář si zaslouží definice pojmů aktivní místo a pór. Obrázek 1.2 ukazuje příklady těchto objektů. U obrázku a) bych zvolila spíše situaci, kdy je aktivní místo lokalizováno na konci tunelu. Tento případ je velmi běžný, uvedená ilustrace je spíše speciálním případem (alespoň v oblasti tunelů). Obrázek b) zobrazuje spíše tzv. kanál (channel) – jako pór se označují zejména cesty v transmembránových proteinech, které nejčastěji vykazují rovný a pravidelný tvar. Jedná se tedy o jakýsi speciální případ kanálů. V sekci 1.2 je uvedeno, že tunel směřuje z aktivního místa na povrch proteinu. Tunely slouží ale rovněž pro transport v opačném směru. Další komentář směřuje k vysvětlení gridového přístupu v sekci 1.3.1. Student zmiňuje, že hlavním problémem je skutečnost, že tunely nemusí být nalezeny. Toto je pravdou pro tunely, které mají velmi

úzká místa a takové tunely jsou často z biochemického hlediska nevýznamné. Mnohem větším problémem tohoto přístupu je závislost výsledků na rozlišení použité mřížky – s různým nastavením můžeme získat značně odlišnou informaci o tvaru a objemu spočtených tunelů.

V obrázku 1.10 student používá pojem ‘nerve’, ve skutečnosti se však jedná o Voroného hranu. V kapitole 1.3 je možné uvést rovněž metody detekce tunelů založené na molekulárním povrchu či využívající sondu pro nalezení cest.

U obrázku 1.1 by mělo být uveden nástroj, pomocí něhož byly screenshoty vytvořeny. Podobně u obrázku 1.5 si nejsem jistá, zda jej student vygeneroval sám nebo jej převzal. Ve druhém případě je třeba citovat zdroj. U obrázku 4.1 student tvrdí, že v proteinu 1AON je dobře vidět kavita. S tímto nemohu souhlasit – kdybych protein sama nestudovala dříve, bylo by pro mě obtížné toto z obrázku vyčíst. Podobně obrázek 4.2 není příliš přehledný.

V textu se objevuje několik jazykových chyb, nicméně vzhledem k rozsahu je nepovažuji za významné a nebudu je tedy ani uvádět. Pouze pro příští práce studenta zmíním, že ‘, e.g.’ je vždy uvozeno čárkami z obou stran a v odborném textu by se neměly používat zkrácené tvary (don’t, doesn’t apod.). Při citování zdrojů by se student měl příště vyvarovat odkazům na Wikipedii.

K praktické části práce nemám připomínek, program je plně funkční a zdrojové kódy dostatečně obsáhlé.

Na studenta mám následující otázky:

- Proč je v rovnici pro  $d(s_1, s_2)$  na straně 14 druhá část násobena číslem 10? Jak toto číslo bylo určeno?
- Proč algoritmy pro plánování cest používají vždy čtvercové okolí (viz obrázek 3.2)? Nebyla by vhodnější konvexní obálka? Podobně u detekce kolizí zmíněné v sekci 5.1.2 jsou uváděny OBB a AABB struktury pro urychlení hledání kolize. Proč se nevyužívají sofistikovanější metody?
- V rovnici na straně 30 jsou použity proměnné  $t_i$  a  $t_{i+1}$ , v následujícím vysvětlení však figurují proměnné  $q_i$  a  $q_j$ . Jedná se o chybu?
- U obrázků 5.5 a 5.9 ukazujících detail výsledků pro jednotlivé přístupy chybí zobrazení výsledku pro naivní přístup. Čím je to způsobeno?
- Jakým způsobem jste vybíral struktury pro porovnání s nástrojem CAVER 3.0?

Výsledky použití algoritmu TOM-RRT jsou z hlediska přesnosti výpočtu přesvědčivé, protože algoritmus dokázal spolehlivě detekovat nejvýznamnější tunely – významnost lze odvodit od číslování tunelů nástrojem CAVER 3.0, kdy pravděpodobnost relevance tunelů s nižším číslem je významně vyšší. Vytvořený algoritmus byl schopen tyto tunely odhalit. Problémem stávající implementace je však jeho rychlost. Algoritmus bude v budoucnu třeba optimalizovat, aby mohl konkurovat stávajícím používaným řešením. Pokud se v dalším kroku povede algoritmus rozšířit na simulace molekulových dynamik, může se stát významným nástrojem používaným vědci z biochemické komunity.

Závěrem konstatuji, že předložená práce svým rozsahem a kvalitou převyšuje nároky kladené na bakalářskou práci, proto ji hodnotím známkou **A (výborně)**.

Barbora Kozlíková