

Univerzita Karlova
Přírodovědecká fakulta

Studijní program Bioinformatika

Obor Bioinformatika

Kateřina Faltejsková

Identifikace transproteinových pórů
Identification of transprotein pores

Bakalářská práce

Vedoucí bakalářské práce: Mgr. Marian Novotný, Ph.D.

Praha, 2018

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 7. května 2018

Podpis autora

Děkuji svému školiteli, Mgr. Mariánu Novotnému, PhD., za odborné konzultace a trpělivou pomoc při psaní této práce.

Abstrakt

In silico analýza proteinových struktur je nástrojem prozkoumávání existujících proteinů. Využívá se i při designu proteinů a prozkoumávání variant mutageneze, protože struktura proteinu je určujícím faktorem pro jeho funkci.

Protein není kompaktní, má na svém povrchu vychlípeniny a uvnitř prázdné prostory. Mezi tyto prázdné prostory patří mimo jiné tunely a póry. Tunely a póry mají významnou funkční úlohu při transportu látek přes membránu nebo k aktivním místům v proteinu. V současnosti existuje velké množství programů k jejich detekci. Tyto programy jsou vystavěny na velmi různorodých principech. Velmi často z těchto principů vyplývají omezení a předpoklady na vlastnosti zkoumaného proteinu nebo póru. Tato práce se věnuje shrnutí používaných přístupů a porovnání programů na těchto principech založených.

Klíčová slova: pór, protein, MOLE, funkční analýza

Abstract

Protein structure examination *in silico* is an instrument used while analysing existing protein as well as protein designing and searching mutagenesis variants, because the structure of a protein is a determining factor of protein function.

Protein does not have a compact structure. There are pockets on the surface and void spaces inside the protein. Into the category of void spaces inside of the protein we classify tunnels and pores. The main function of tunnels and pores is in transport of material through a membrane or towards active sites buried inside the protein. At the moment there is plenty of tunnel or pore detection programs. These programs are built on various approaches. It is common for such a program to be limited by its approach or to have pre-requirements about the examined protein or pore. This work is meant to summarize the approaches and to compare programs built upon these approaches.

Key words: pore, protein, MOLE, functional analysis

Obsah

Úvod	4
1 Prázdné prostory v proteinové struktuře	5
1.1 Póry	5
1.2 Dutiny a kapsy	6
1.3 Tunely	6
2 Algoritmická detekce prázdných prostor v proteinové struktuře	7
2.1 Výpočet a analýza proteinového povrchu	7
2.1.1 Obepínající koule	7
2.1.2 Konvexní obal	8
2.1.3 Povrch přístupný rozpouštědлу	8
2.2 Analýza vnitřního prostoru proteinu	9
2.2.1 Algoritmy založené na vyplňování prostoru mezi atomy	9
2.2.2 Algoritmy využívající geometrické vlastnosti tunelů a pórů	9
2.2.3 Algoritmy založené na tvorbě Voroného diagramu proteinové struk- tury	10
2.2.4 Algoritmy využívající vlastností transmembránových proteinů	13
2.2.5 Algoritmy zahrnující molekulární dynamiku proteinů	13
3 Porovnání programů	15
3.1 Cílení a specializace programů	15
3.2 Dostupnost a vizualizace	15
3.3 Porovnání z hlediska povinných vstupů	15
3.4 Porovnání výstupů	16
3.4.1 Hledání tunelů	16
3.4.2 Hledání pórů	18
3.5 Výběr programu	21
4 Příklady využití programů	22
4.1 Použití při zkoumání proteinů	22
4.2 Použití při designu proteinů	22
4.3 ChannelsDB	23
Závěr	24
Seznam použité literatury	25

Úvod

Proteiny jsou významnou funkční a strukturní složkou živých organismů. Skládají se z řetězce či řetězců aminokyselin, které se sbalují do prostorových útvarů. Výsledný tvar proteinu je určen sekvencí řetězců aminokyselin, které protein tvoří, a podmínkami a modifikacemi, kterým byl protein vystaven během sbalování. Struktura proteinu není kompaktní, ale obsahuje množství prázdných prostor, které jsou často konzervované a hrají významnou roli. Jednou z významných funkcí, kterou nevyplněné prostory v proteinu mohou plnit, je selektivní a regulovatelný přenos nejrozumnějších nákladů zpravidla z jedné strany membrány na druhou. Proteiny plnící tuto funkci jsou často pevně zanořeny v membráně. (shrnutí v Alberts a kol. (2002))

Další možnou funkcí prázdných prostorů uvnitř nebo na povrchu molekuly enzymu je částečné oddělení aktivního místa od okolního prostoru a rozpouštědla. Některá aktivní místa (například u cytochromu P450 (Cojocar a kol., 2007)) jsou ukryta uprostřed proteinové struktury, kam je substrát z povrchu transportován úzkými přívody. Takto se prázdné prostory mohou podílet na selektivitě enzymu. Tunel o délce alespoň 15 Å je přítomen v 64,2% enzymů, tunel dlouhý 50 Å můžeme nalézt v 10% enzymů (Pravda a kol., 2014).

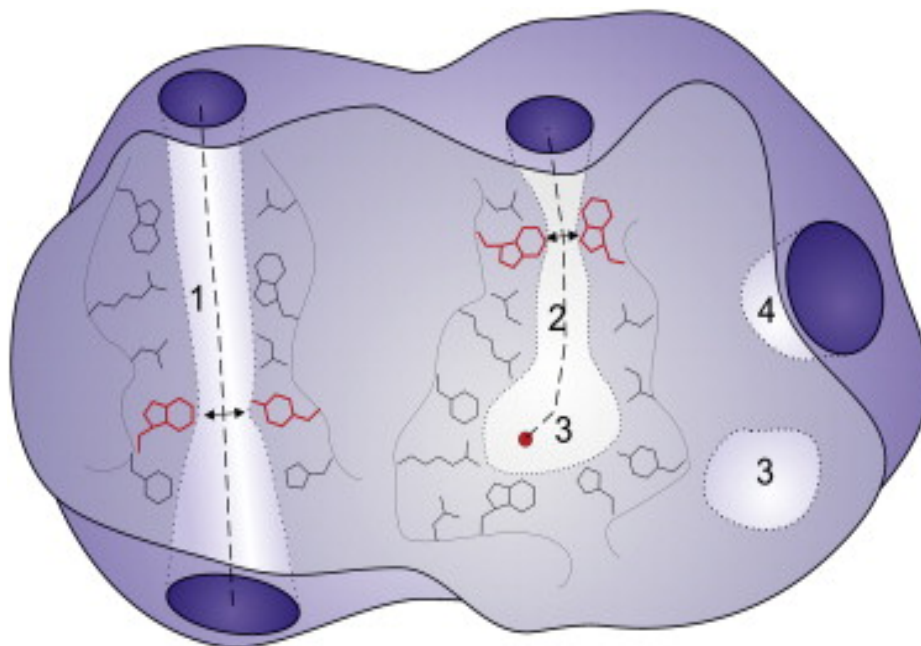
Znalost polohy, strukturních vlastností a funkce prázdných prostorů v molekulách proteinů s regulační i strukturní funkcí lze využít ke změně či modifikaci jejich funkce. Spolehlivá detekce pórů, tunelů a dalších prázdných prostor v proteinu *in silico* má potenciál při navrhování léčiv nebo navrhování proteinů s žádanými vlastnostmi.

1. Prázdné prostory v proteinové struktuře

Prázdné prostory ve struktuře proteinu mohou mít různé podoby a funkce. Jejich algoritmická detekce vyžaduje přesné rozdělení a vymezení z důvodů rozdílných geometrických podob, na které se jednotlivé programy mohou specializovat. Názvosloví bohužel mezi autory není ustálené, zde je uvedeno nejčastější.

1.1 Póry

Nejvýraznějším z proteinových prostor je pór, dutina vedoucí z povrchu proteinu skrz strukturu znovu na povrch, na obrázku 1.1 je znázorněn pod číslem 1. Bývá poměrně přímý. Často slouží jako spojení dvou stran membrány. (Brezovsky a kol., 2013; Pravda a kol., 2018; Sehnal a kol., 2013) Jeho funkcí může být přenos nízkomolekulárních látek přes membránu. Příkladem takových látek jsou aquaporiny, konkrétně AQP0 protein (Harries a kol., 2004), jehož prostřednictvím jsou do buňky přinášeny nízkomolekulární látky, hlavně voda.



Obrázek 1.1: Různé druhy prázdných prostorů v proteinové struktuře, převzato z Brezovsky a kol. (2013), upraveno. Útvar 1 představuje pór, útvar 2 tunel jako přístupovou cestu k dutině pod číslem 3. Druhý útvar pod číslem 3 je poddruh dutiny, "void", který není připojen k povrchu. Číslo 4 označuje kapsu.

Programy specializované pouze na detekci pórů často staví na vlastnostech transmembránových proteinů a jejich specifických vlastnostech nebo na tom, že směr (vektor) póru už je znám z předchozího zkoumání proteinu. Podrobné rozpracování tématiky je uvedeno v následující kapitole.

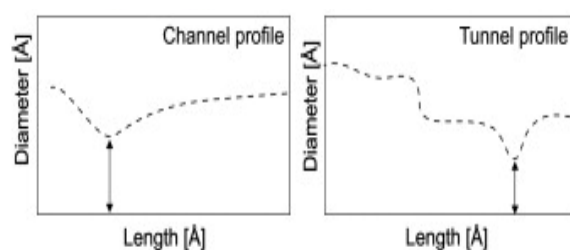
1.2 Dutiny a kapsy

Prázdný prostor uprostřed proteinové struktury nazýváme dutinou. K povrchu může být připojen tunelem. Dutina je obvykle širší než tunel, který ji spojuje s povrchem. Pokud není dutina připojena k povrchu, označuje se anglickým termínem „void“. Void i dutina jsou na obrázku 1.1 označeny číslem 3, dutina je připojena k povrchu, void není. (Brezovsky a kol., 2013; Pravda a kol., 2018; Sehnal a kol., 2013) Kapsa je prohlubeň na povrchu proteinu, na obrázku 1.1 pod číslem 4. Dutiny a kapsy (Brezovsky a kol., 2013) jsou často součástí aktivních míst enzymů. V dutině umístěné aktivní místo nalezneme například u cytochromu P450. (Cojocar a kol., 2007)

Detekce dutin a kapes se na rozdíl od detekce tunelů a pórů nepotýká s problémem nalezení druhého vstupního bodu. Nalézání dutin a kapes často vychází z popisu povrchu. Hledání dutin mnoho programů využívá jako přípravnou fázi při hledání tunelů a pórů.

1.3 Tunely

Tunelem rozumíme spojení povrchu proteinu s vnitřním prostorem (na obrázku 1.1 pod číslem 2). Mívá členitější profil než pór. Porovnání profilů póru a tunelu můžete vidět na obrázku 1.2, první graf profilu náleží póru, druhý tunelu. (Brezovsky a kol., 2013)



Obrázek 1.2: Porovnání profilu póru (první graf) a tunelu (druhý graf), převzato z Brezovsky a kol. (2013), upraveno. Povšimněte si výkyvů v profilu tunelu v kontrastu k póru, jehož průměr klesá či stoupá poměrně rovnoměrně.

Mohou sloužit jako přívody substrátu k aktivním místům hlouběji zanořeným v proteinové struktuře enzymu v kavitách, respektive k odvodu produktů. (Brezovsky a kol., 2013; Pravda a kol., 2018; Sehnal a kol., 2013), viz obrázek 1.1. Takto se podílejí na selektivitě enzymů. Příkladem tunelu, který odvádí produkt do cytosolu, je protein velké ribozomální podjednotky (Voss a kol., 2006) nebo tunely přivádějící substrát do aktivního místa výše zmíněného cytochromu P450. (Cojocar a kol., 2007)

2. Algoritmická detekce prázdných prostor v proteinové struktuře

Použití počítačové simulace na analýzu členitého trojrozměrného útvaru má obdoby i v dalších vědních oborech, například v geografické analýze zemského reliéfu nebo při 3D tisku. Při analýze proteinové struktury lze navíc kromě množiny atomů (reprezentovaných většinou jako koule v prostoru) použít k analýze fyzikální a chemické vlastnosti aminokyselin tvořících zkoumaný protein.

Detekce prázdných prostor se ve většině programů dá rozdělit do dvou fází. První fází je zpravidla detekce povrchu proteinu. Povrch je často členitý a tedy složitý na popsání, některé programy proto používají jeho různě přesnou aproximaci.

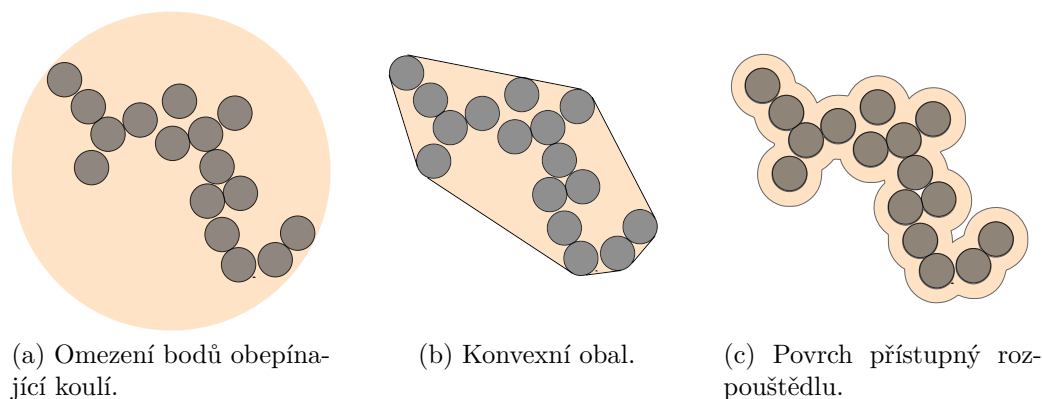
Dalším krokem detekce je prohledávání proteinové struktury jako celku, zejména její vnitřní části. Její princip se velmi odlišuje napříč různými přístupy. Některé principy analýzy nevyžadují znalost povrchu (PoreWalker, viz níže), nebo používají jenom detailní analýzu povrchu (HOLLOW, viz níže).

2.1 Výpočet a analýza proteinového povrchu

Popis proteinového povrchu je úvodní částí detekce proteinových tunelů, pórů a kavit uvětšiny programů. Slouží jak k omezení prohledávaného prostoru, tak k samotné identifikaci kavit a startovních míst pórů a tunelů. Existují však i programy (např. PoreWalker, viz níže), které výpočet celého povrchu proteinu neprovádějí.

2.1.1 Obepínající koule

Nejjednodušší aproximací povrchu, která je zároveň nejméně přesná, je uzavření všech atomů (bodů v 3D prostoru) do nejmenší možné koule, která je pojme. Podobnou aproximaci využívá MolAxis (Yaffe a kol., 2008a). MolAxis dovoluje uživateli, aby kouli nastavil sám, a tím i určil prostor, kde bude program hledat prázdné prostory. Pokud uživatel chce hledat v celé proteinové struktuře, musí vypočítat obepínající kouli sám.



Obrázek 2.1: Různé možnosti aproximace povrchu.

2.1.2 Konvexní obal

Další možností, jak obsáhnout prostor vyplňovaný proteinem, je pomocí konvexního obalu (2.1b). Množinu M nazveme konvexní, pokud úsečka spojující libovolné dva body z množiny M sama leží v množině M . Konvexním obalem množiny M nazveme nejmenší konvexní množinu, která množinu M obsahuje. V důsledku toho lze pro výpočet zjednodušit proteinovou strukturu na množinu 3D bodů, kde bod označuje centrum atomu.

Na konstrukci konvexního obalu množiny se běžně používá algoritmus Quickhull (Barber a kol., 1996). Algoritmus postupně po dimenzích rozšiřuje množinu bodů konvexního obalu na základě zkoumání, zda je přidávaný bod na vnitřní nebo na vnější straně stěny. Quickhull má pro trojdimenzionální prostor složitost $O(n \log v)$, kde n je počet všech bodů množiny a v je počet bodů konvexního obalu.

Konvexní obal je přesnější vyjádření proteinového povrchu než obepínající koule. Pokud je vstupní množina bodů výrazně nekonvexní, konvexní obal musí prostor zaujímaný proteinem silně nadhodnotit. Existují techniky, které konvexní obal zpřesní. Například MOLE 2 (Sehnal a kol., 2013) před detekcí a aproximací molekulového povrchu rozdělí protein na tetrahedrony Voroného diagramu. Poté algoritmus vyprodukuje konvexní obal omezující množinu tetrahedronů. Do tetrahedronů v kontaktu s obalem jsou pokusně vkládány koule s uživatelem definovaným poloměrem *proberadius*. Tetrahedrony jsou z povrchu odstraněny, pokud kouli pojímou.

2.1.3 Povrch přístupný rozpouštědlu

Povrch přístupný rozpouštědlu (solvent accessible surface area, ASA) je velmi přesná metoda aproximace povrchu (2.1c). Vstupem k výpočtu je množina 3D koordinátů atomů popisující strukturu zkoumané molekuly a poloměr koule reprezentující molekulu rozpouštědla. Každý atom reprezentuje koule o Van der Waalově poloměru daného typu atomu se středem v bodě dané koordináty atomu. Atomy, k nimž lze přiložit koule rozpouštědla tak, že se povrchy koulí dotýkají a koule rozpouštědla neprotíná žádnou reprezentaci atomu zkoumané struktury, jsou definovány jako přístupné rozpouštědlu. (Lee a Richards, 1971). Povrch přístupný rozpouštědlu je definován jako množina všech středů, které může koule rozpouštědla zaujmout, aby byly splněny podmínky uvedené výše.

Většina dnešních programů pro výpočet ASA vychází ze Shrake-Rupleyho algoritmu (Shrake a Rupley, 1973). Shrake-Rupleyho algoritmus definuje rozpouštědlo jako molekulu vody a pro každý atom určí, kolik bodů z výběru bodů na povrchu koule generované Van der Waalovým poloměrem atomu je přístupných rozpouštědlu (viz výše). Tyto body pak generují mřížku proteinového povrchu.

Omezení proteinu určené touto metodou je výrazně přesnější než povrch aproximovaný pomocí konvexního obalu. Protože je však výpočet či aproximace proteinového povrchu v mnoha programech pouze pomocnou procedurou a protože výpočet povrchu přístupného rozpouštědlu má obecně složitost $O(n^2)$, kde n je počet atomů, používá se spíše konvexní obal, popřípadě jeho zpřesnění.

2.2 Analýza vnitřního prostoru proteinu

2.2.1 Algoritmy založené na vyplňování prostoru mezi atomy

První programy určené ke zkoumání proteinových pórů, tunelů a kavit vyžadovaly znalost polohy a směru póru předem. Nebyly určeny k detekci, ale k určení tvaru, šířky a aminokyselin lemujících pór. K těmto programům patří níže zmíněný HOLE.

Vyplňování prázdných prostorů imaginárními koulemi nemusí sloužit pouze k vyplnění póru nebo tunelu, ale lze tímto způsobem určit i povrch a jeho vychlípeniny. Tento přístup má nedostatek v případě uzavřených nebo zúžených tunelů a pórů. Proteiny jsou dynamické struktury, které jsou ovlivňovány i vazbou ligandů a interakcemi s dalšími proteiny. Tvar celého proteinu stejně jako tvar vychlípeniny není konstantní v čase. Průtok pórem může být regulován jeho uzavřením nebo ucpáním. Přístupy využívající pouhou analýzu povrchu nemohou rozlišit uzavřený vstup póru od kapsy nebo kavity. Tato skupina obsahuje například programy HOLE a HOLLOW.

HOLE

HOLE (Smart a kol., 1996) je jedním z nejstarších programů pro analýzu proteinových pórů. Vyžaduje zadání počátečního bodu někde v póru a koordinátů póru, výstupem programu je detailní popis tvaru a reliéfu póru. Popis je vyhledáván pomocí vkládání největší možné kapsule (s osou kolmou k vektoru póru, maximalizována je plocha kapsule na rovině normály vektoru póru), hledání středů probíhá pomocí Monte Carlo simulovaného žíhání. Výstupy poskytované programem HOLE často slouží jako benchmark kvality nově tvořených nástrojů. HOLE byl takto použit například při zkoumání kvality programů MOLE (Petřek a kol., 2007; Sehnal a kol., 2013) a MolAxis (Yaffe a kol., 2008a).

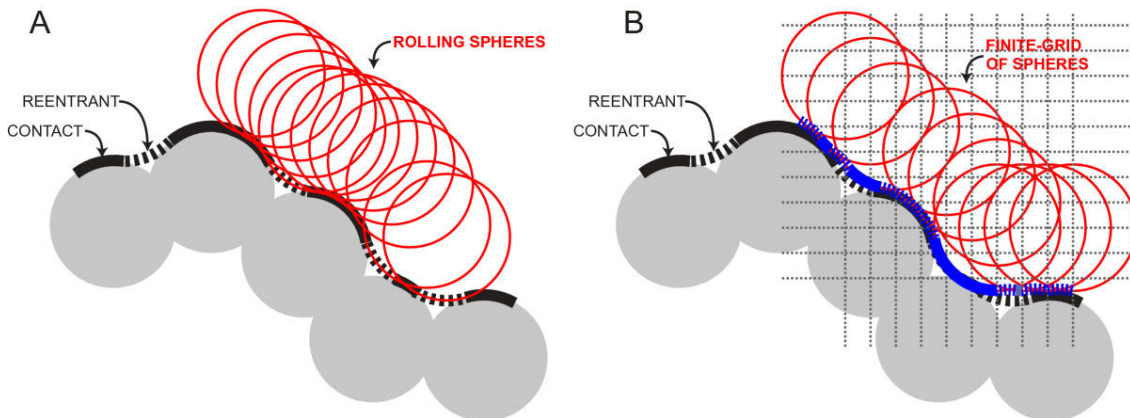
HOLLOW

HOLLOW detailně analyzuje vnitřní prázdné prostory tak, že je pokryje pomocnou mřížkou a uměle vyplní "dummy atoms"- koulemi se středem v uzlech mřížky a maximálním možným poloměrem takovým, aby neprotnul Van der Waalsovský poloměr žádného z atomů. Přesnost aproximace je ilustrována na obrázku 2.2, sekce B. Použití konečného počtu pomocných koulí se od ideálního případu, kdy je okolní prostor vyplněn nekonečně mnoha koulemi (obrázek 2.2, sekce A), odchyluje jen málo. Následně je definován vnější povrch a odstraněny "dummy atoms" nad povrchem. Výsledná mapa prázdných prostor v proteinu je pak analyzována. Pro hledání pórů program hledá dva výstupní body a cestu mezi nimi, která je zároveň obsažená ve vymezeném prázdném prostoru. (Ho a Gruswitz, 2008)

2.2.2 Algoritmy využívající geometrické vlastnosti tunelů a pórů

Na proteinovou strukturu lze nahlížet jako na komplexní prostorový objekt. Některé programy založily hledání prázdných prostorů v proteinu ryze na analýze geometrie proteinu. Níže popsané algoritmy založené na rozložení atomů proteinu do Voroného diagramu by se daly zařadit jako podskupina této.

Příkladem programu využívajícího tento přístup je CHUNNEL.



Obrázek 2.2: Ilustrace hledání proteinového povrchu pomocí vyplňujících koulí, porovnání povrchu generovaného nekonečným (A) a konečným (B) počtem koulí, převzato z Ho a Gruswitz (2008)

CHUNNEL

Prvním krokem vyhledávání programu CHUNNEL (Coleman a Sharp, 2009) je triangulace povrchu a výpočet konvexního obalu jako referenční obálky. Za předpokladu, že v proteinu je pór nebo tunel, alespoň část struktury vypadá jako torus nebo vícenásobný torus v případě, že tunelů (pórů) je ve struktuře více. Pomocí Eulerovy charakteristiky grafu lze určit, kolikanásobný torus struktura připomíná, tedy kolik tunelů (pórů) v proteinu je.

Eulerova charakteristika $\chi = V - E + F$, kde V je počet vrcholů grafu, E počet hran grafu a F počet stěn grafu, je pro N -torus rovna $\chi = 2 - 2N$.

Poté program postupně proteinovou strukturu "ořeže" tak, že z původní triangulace zbývá pouze povrch póru. Pokud je v proteinu více pórů, program umísťuje do již popsaného póru umělé zátky za účelem jejich izolování a prozkoumání. Tato vlastnost je pro program CHUNNEL specifická. Z programů navržených na hledání pórů (ne tunelů) je jediný, který v proteinu dokáže najít dva a více pórů /citepBrezovskyOverview.

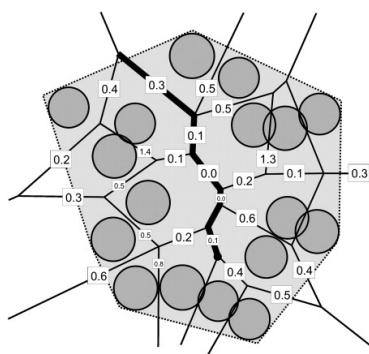
2.2.3 Algoritmy založené na tvorbě Voroného diagramu proteinové struktury

Voroného diagram $A(P)$ je definován jako rozklad množiny bodů P na n uzavřených oblastí $A(P) = \{A(p_1), \dots, A(p_n)\}$ takových, že každý bod a z $A(p_i)$ je blíže bodu p_i než k

jakémukoliv bodu p_j pro všechna $i \neq j$. Reprezentujeme-li atomy v proteinové struktuře jako koule o daném poloměru, jako výchozí množinu bodů můžeme použít středy těchto koulí.

Z Voroného diagramu algoritmy generují ohodnocený graf $G = (V, E)$, kde E , tedy množina hran, položí jako množinu rozhraní mezi jednotlivými podmnožinami rozkladu a V , množinu vrcholů grafu, jako střetnutí dvou a více hran. Hrany jsou ohodnoceny pomocí váhové funkce, která je specifická pro daný program, ale často závidí na délce hrany a její vzdálenosti k nejbližšímu atomu. V ohodnoceném grafu hledáme nejlevnější cestu. Na obrázku 2.3 je nejlevnější cesta ilustrována zvýrazněnou čarou.

Programy využívající tento princip patří mezi nejvyužívanější. Patří mezi ně MolAxis, MOLE a CAVER (a jejich verze).



Obrázek 2.3: Graf generovaný z dvourozměrného Voroného diagramu, převzato z Petřek a kol. (2007). Hrany jsou ohodnoceny váhovou funkcí programu MOLE. Kruhy ilustrují jednotlivé atomy.

Jedním z algoritmů na hledání nejkratších (nejlevnějších) cest v grafu je Dijkstrův algoritmus. Dijkstrův algoritmus prohledává graf do šířky pomocí "vlny", která se rozšiřuje vždy do nejbližšího nenavštíveného vrcholu, tedy do vrcholu v takového, že součet hran na cestě z počátku do vrcholu v je nejmenší přes všechny dosud nenavštívené vrcholy. Algoritmu lze nastavit konečný vrchol cesty, po jehož dosažení skončí, nebo lze pomocí Dijkstrova algoritmu prozkoumat celý graf a zjistit délku nejkratší cesty z počátku do každého vrcholu (spojitého) grafu. (Dijkstra, 1959)

MOLE a jeho verze

Prvním krokem vyhledávání pomocí programu MOLE je nalezení konvexního obalu molekuly (na obrázku 2.3 naznačen čarou obepínající molekuly (prostor vybarvený světle šedě označuje aproximaci molekuly pomocí konvexního obalu) MOLE 2 konvexní obal dále zjemňuje nalezením kavit a odebráním prázdného prostoru kavit z reprezentace. Tyto kavity plní navíc úlohu automatických startovních bodů pro prohledávání grafu. Použití konvexního obalu tedy narozdíl od výpočtu ASA je důvodem, proč MOLE 2 dokáže automaticky určit startovní body tunelů a pórů - konvexní obal funguje jako referenční obálka pro identifikaci vchlípenin povrchu.

Je zkonstruován Voroného diagram pro celou proteinovou strukturu omezenou konvexním obalem, pomocí váhové funkce jsou ohodnoceny hrany grafu. V grafu pak program hledá nejkratší cesty pomocí Dijkstrova algoritmu.

MOLE cesty hledal sekvenčně. Na všech hranách nově nalezené cesty je následně zvýšeno ohodnocení, aby byla daná hrana nalezena v příštím spuštění Dijkstrova algoritmu nalezena s menší pravděpodobností. Tento přístup byl v MOLE 2 opuštěn, protože zkresloval výsledky (zvýšení ohodnocení hrany může skrýt biologicky významnější cestu, která nemá optimální cenu ve Voroného diagramu). MOLE 2 počítá cesty mezi všemi dvojicemi startovních bodů a následně je filtruje

Na vyhledávání pórů a tunelů pomocí programu MOLE 2 staví dále zmíněná ChannelsDB. (Pravda a kol., 2018)

CAVER a jeho verze

CAVER 1 vnitřní prostor proteinu aproximoval pomocí 3D mřížky omezené konvexním obalem proteinu, ohodnotil jejich uzly pomocí váhovou funkcí a hledal v takto vytvořeném grafu nejkratší a nejlevnější cestu. Váhová funkce závisela nepřímo úměrně na velikosti největší koule, která měla střed v uzlu a neprotínala Van der Waalsovský poloměr žádného atomu. Přičítalo se k ní malé číslo, aby nehrozilo dělení nulou. (Petřek a kol., 2006)

Další verze programu CAVER používají Voroného diagram. Nejnovější z nich je CAVER 3 (Chovancova a kol., 2012). Nejmenší atomy reprezentuje jako koule o Van der Waalsovském poloměru daného atomu r , atomy větší pak jako set koulí o poloměru r takových, že dohromady aproximují kouli o poloměru reprezentovaného atomu. To je jeden z hlavních rozdílů od programu MOLE, který poloměry atomů zanedbává.

Z takto reprezentovaných atomů vypočte Voroného diagram omezený konvexním obalem. Na Voroného grafu omezeném povrchem počítá pomocí Dijkstrova algoritmu nejkratší cesty v grafu z uživatelem daného bodu uvnitř póru nebo tunelu k tzv. povrchovým vrcholům (vrcholy, které jsou spojeny s vrcholy ASA). Cesty jsou poté prodlouženy k vrcholům ASA a filtrovány. (Chovancova a kol., 2012).

MolAxis

MolAxis atomy reprezentuje jako koule. Různé Van der Waalsovské poloměry atomů řeší stejně jako CAVER. Povrch nepočítá, uživatel zadává tzv. "bounding sphere", kouli, na kterou má program omezit své hledání. Uvnitř koule je vypočten Voroného diagram a ohodnoceny hrany. Poté je ze startovního bodu (ten může být zadán uživatelem, nebo určen automaticky jako největší komora v proteinu) spuštěn Dijkstrův algoritmus. Jeho pomocí je vytvořen co nejlehčí (nejlevnější cesta) strom (acyklický graf) popisující síť prázdných prostor v proteinu.

Strom je následně specificky vyhodnocován podle toho, zda uživatel očekává transmembránový protein (a transmembránový pór), nebo cytosolový protein (obsahující tunely).

Pro hledání tunelů je uživatelem zadán nebo automaticky vypočten startovní bod, od kterého program ve vypočteném stromě hledá cestu k povrchu podle váhy cest vzestupně (od nejnižší po nejvyšší). Cesty k povrchu jsou navíc tříděny podle toho, jak moc se od sebe liší (měřeno pomocí míry větvení cest).

Při hledání pórů MolAxis předpokládá, že se jedná o transmembránový protein a vyžaduje, aby uživatel zadal roviny membrány, kterou má pór vést. Vrcholy jsou rozděleny na ty, které leží nad rovinou membrány, a ty, které se nacházejí pod rovinou membrány. Z obou je sestaven strom. Mezi vrcholy obou skupiny pak program hledá spojující hranu mezi oběma stromy (každý vrchol je z jiné skupiny), která zároveň prochází přes rovinu membrány. Pór je vytvořen jako sjednocení cesty z povrchu ke spojující hraně (ve stromě "nad" rovinou membrány), spojující hrany a cesty od spojující hrany k povrchu (ve stromě

"pod" rovinou membrány). Výstupem programu je pór reprezentovaný cestou s nejmenším součtem vah na hranách (Yaffe a kol., 2008a).

2.2.4 Algoritmy využívající vlastností transmembránových proteinů

Transmembránové proteiny jsou zásadním funkčním prvkem biologických membrán - tonoplastu, membrány na chloroplastu, plazmatické membrány a dalších - pro transport látek přes membránu. Pór je hlavním a velmi výrazným prvkem těchto proteinů. Vlastnosti transmembránových proteinů jsou poměrně specifické. Sekundární struktury domén nalézajících se uvnitř membrány jsou typicky α -helixy nebo β -listy uzavřené do barelu (shrnutí v Alberts a kol. (2002)).

Příkladem programu hledajícího póry tímto způsobem je PoreWalker.

Porewalker

Porewalker (Pellegrini-Calace a kol., 2009) používá iterativní heuristický algoritmus vycházející z toho, že sekundární struktury proteinu směřují zhruba stejným směrem jako pór, který lemují. Vektor póru je inicializován jako průměr směrů sekundárních struktur, bod uvnitř póru je definován jako těžiště. Iterativně je nejprve optimalizován bod určující pór (program se snaží maximalizovat počet předpokládáných reziduí póru vzhledem k předběžné ose), poté je optimalizována osa pomocí "naplátkování" proteinu podle stávající osy. Po určení optimální osy program identifikuje residua lemující pór a vyplní volný prostor posloupaností válců a kuželových výsečí.

2.2.5 Algoritmy zahrnující molekulární dynamiku proteinů

Struktura proteinu není v čase konstatní. Regulace funkce je úzce spojena s regulací struktury. Konkrétně transportní funkce tunelů a pórů může být regulována jejich zúžením. Tato a podobné změny struktury představují pro programatickou detekci prázdných prostor problém. Příkladem kompenzace je přístup autorů programu CAVER 3 (Chovancova a kol., 2012), který uživateli dovoluje analyzovat více modelů proteinu a následně prozkoumat všechny snímky a nalezené prázdné prostory porovnávat za účelem získání ucelenějšího obrazu. V zatím nejnovější verzi CAVER 3.02 je podpora zkoumání tunelů z různých snímků proteinu během simulace molekulární dynamiky dále rozvíjena (Pavelka a kol., 2016).

Existují i pokusy inkorporovat molekulární dynamiku proteinů přímo do detekce. Dalším příkladem je program dxTuber.

dxTuber

Program dxTuber (Raunest a Kandt, 2011) nejprve pro protein spustí simulace jeho molekulární dynamiky. Délka simulace závisí na rozpouštědle, pro vodu autoři určili dobu simulace na 100 ps. Trajektorie atomů jsou pak převedeny do třírozměrného pole hustot atomů (hustota v buňce mřížky je položena buď jako průměrná hustota atomů v čase pro protein i pro rozpouštědlo nebo jako průměrná hustota atomů v čase pro rozpouštědlo a minimální hustota atomů v čase pro protein).

Mřížka je poté prohledávána třemi druhy vyhledávacích algoritmů pro nalezení různých druhů prázdných prostorů: rozšiřování do tří směrů pro nalezení kavit, rozšiřování do 2

směrů pro nalezení tunelů (uživatel určuje směr prohledávání, k dispozici jsou pouze směry tří hlavních os X, Y, Z) a pórů a rozšiřování do 1 směru pro hledání kapes a méně výrazných vychlípenin povrchu. Pro danou buňku C mřížky platí, že je možné rozšířit se přidáním C , pokud hustota buňky C nepřesáhla uživatelem nastavenou hustotu.

3. Porovnání programů

3.1 Cílení a specializace programů

Při porovnávání jednotlivých programů, které se na hledání proteinových pórů a tunelů specializují, je třeba vzít do úvahy určení jednotlivých programů. Obecně lze říct, že programy, které v proteinové struktuře mají najít tunel, bez obtíží naleznou i pór. Opačná implikace obecně pravdivá není. Algoritmy určené na hledání pórů, tj. HOLE, PoreWalker, často vycházejí ze specifických vlastností pórů - PoreWalker odvozuje umístění póru ze sekundárních struktur proteinu, HOLE explicitně hledá průchod molekulou.

Při volbě programu k analýze dané proteinové struktury je tedy třeba zvážit, jaký typ prázdného prostoru v ní očekáváme.

3.2 Dostupnost a vizualizace

Všechny zde popsané programy kromě PoreWalkeru jsou k dispozici ke stažení, buď jako konzolová nebo jako desktopová aplikace. Některé programy však nabízejí své služby na webovém serveru. V případě programu PoreWalker je webový server jediným možným přístupem. V tabulce 3.1 je uveden přehled, jaké programy webový server poskytují a na jaké adrese je možné server nalézt.

S dostupností programu se pojí i dostupnost zobrazovacích nástrojů pro výstupy programu. Jen některé z výše popsaných programů poskytují uživateli grafický interface, konkrétně se jedná o dxTuber (Raunest a Kandt, 2011). PoreWalker uživateli vizualizaci póru poskytuje ve formě webové stránky. Výstupy programů MOLE, CAVER a HOLLOW mohou být pro vizualizaci exportovány do programu PyMol, program MolAxis své výstupy exportuje ke zobrazení v programu Jmol.

Program HOLE je starší než všechny ostatní zde popisované programy. Při vizualizaci výsledků spoléhá na starší program molekulární grafiky Quanta. (Smart a kol., 1996)

3.3 Porovnání z hlediska povinných vstupů

Pro analýzu proteinové struktury, o níž nemáme z hlediska jejích pórů a tunelů žádné informace, je žádoucí použít program, který dokáže startovní místa či osy pórů a tunelů

Tabulka 3.1: Přehled dostupnosti programů pomocí webového serveru.

Program	Webový server	Reference
HOLE	neposkytuje	-
HOLLOW	neposkytuje	-
MOLE 2	mole.upol.cz	Berka a kol. (2012)
CAVER 3	neposkytuje	-
MolAxis	bioinfo3d.cs.tau.ac.il/MolAxis	Yaffe a kol. (2008b)
PoreWalker	www.ebi.ac.uk/thornton-srv/software/PoreWalker	Pellegrini-Calace a kol. (2009)
dxTuber	neposkytuje	-
CHUNNEL	neposkytuje	-

určit automaticky. Automatická detekce však může být zatížena chybami a nepřesnostmi v závislosti na použitém přístupu. Pokud však známe přibližnou pozici póru, dokážeme snadněji odhalit falešně pozitivní výsledky.

V tabulce 3.2 se nachází shrnutí, jaké informace výše popsané programy používají jako argumenty a zda programy poskytují možnost nutné informace zjistit automaticky.

3.4 Porovnání výstupů

3.4.1 Hledání tunelů

Z výše popsaných programů se k vyhledávání tunelů, jak byly definovány v kapitole 1, spíše nehodí Porewalker (vychází z orientace sekundárních struktur transmembránových proteinů, tunely nejsou omezeny jen na transmembránové proteiny), HOLE (vstupní informace o póru je možné odhadnout pomocí zobrazovacích programů, tunely jsou často méně opticky výrazné), problematické by mohlo být i použití programu CHUNNEL (geometrický přístup detekce by mohl zanedbat méně výrazné tunely). Zbývají programy MOLE, CAVER, HOLLOW a MolAxis.

MOLE se od programů CAVER a MolAxis liší svým přístupem k reprezentaci atomů. Všechny programy reprezentují atomy jako koule o stejném poloměru (kvůli výpočetním vlastnostem). MOLE zanedbává rozdíly ve Van der Waalově poloměru atomů, zatímco CAVER a MolAxis použijí poloměr nejmenšího atomu a větší atomy nahradí několika koulemi o tomto poloměru tak, aby sjednocení nahrazujících koulí mělo přibližně kulový tvar. (Brezovsky a kol., 2013) Na obrázku 3.1 jsou znázorněny výstupy programů výše zmíněných programů na téže struktuře. Bohužel, tunely na této struktuře nebyly podle databáze ChannelsDB (Pravda a kol., 2018) manuálně anotovány, k porovnání jsou k dispozici pouze predikce.

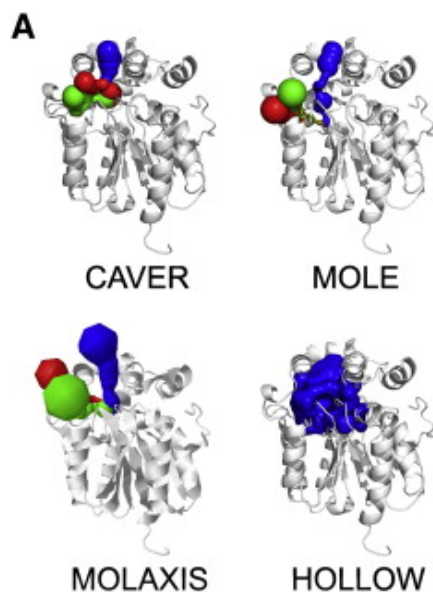
Z dostupných predikcí je vidět, že programy spoléhající na detekci pomocí Voroného diagramu (tj. MOLE, CAVER a MolAxis) se poměrně shodují. Červeně a zeleně vykreslené tunely na výstupech těchto programů víceméně splývají na obrázku 3.1 s modře znázorněným tunelem uloženým v ChannelsDB (obrázek 3.2) a modře znázorněný tunel na obrázku 3.1 se překrývá s fialovým tunelem na obrázku 3.2. Oproti tomu program HOLLOW nedokázal rozlišit jednotlivé tunely od sebe a poskytuje pouze bludiště prázdných prostor bez nápovědy.

Program dxTuber nebo jiné využití molekulární dynamiky by mohlo představovat velmi užitečný nástroj v této nebo podobné situaci. Pomocí vyhodnocení simulací molekulární dynamiky proteinu získáme informace o relevanci a stálosti tunelů v čase. Z takového výstupu by bylo možné usoudit, jak je predikce tunelů přesná i bez manuální anotace. Specifita programu dxTuber je důvodem, proč není v této práci zahrnut do porovnání samotného vyhledávání.

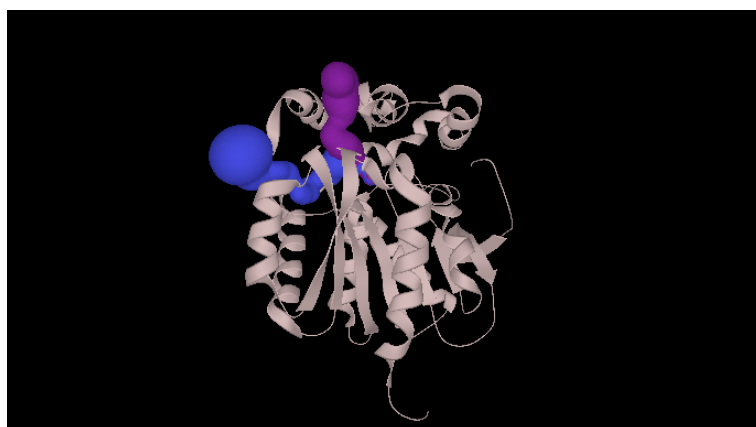
Protože funkcí mnoha tunelů je přívod substrátu k aktivním místům enzymů, jako počáteční bod při hledání tunelu lze použít právě aktivní místo enzymu. (Brezovsky a kol., 2013) Databáze aktivních míst enzymů Catalytic Site Atlas (Porter, 2004; Furnham a kol., 2014) může být zdrojem spolehlivých počátečních bodů pro hledání pomocí programu CAVER i pomocí programu MOLE.

Tabulka 3.2: Přehled informací požadovaných pro vyhledávání výše popsanými programy.

Program	Manuálně zadávané informace o póru/tunelu	Možnost spustit bez informací o prázdných prostorech	Automatický ekvivalent zjištění (pokud je poskytován)	Reference
HOLE	vektor osy póru, bod uvnitř póru	ne	neposkytuje	Smart a kol. (1996)
HOLLOW	nevyžaduje	ano	neposkytuje	Ho a Gruswitz (2008)
MOLE 2	počáteční body tunelu/póru	ano	střed nejhlubších tetraedronů v kavitách (viz výše)	Sehnal a kol. (2013)
CAVER 3	počáteční body nebo rezidua tunelu/póru	ne	neposkytuje	Chovancova a kol. (2012)
MolAxis	tunely – počáteční bod, póry – rovina membrány, tj. rovina přibližně kolmá ke směru póru	pro tunely ano, pro póry ne	tunely – střed největší kavity, pro póry není automatický ekvivalent poskytován	Yaffe a kol. (2008a)
PoreWalker	nevyžaduje	ano	neposkytuje	Pellegrini-Calace a kol. (2009)
dxTuber	určení osy k prohledávání (X, Y nebo Z)	ano (k dispozici grafický interface)	neposkytuje	Raunest a Kandt (2011)
CHUNNEL	nevyžaduje	ano	neposkytuje	Coleman a Sharp (2009)



Obrázek 3.1: Výstupy některých programů na hledání tunelů na stejné proteinové struktuře, převzato z Brezovsky a kol. (2013), upraveno. Zobrazeny výstupy při hledání tunelů na dehalogenáze DhaA (1cqW). Při porovnání jsou použity verze MOLE 1 a CAVER 1.

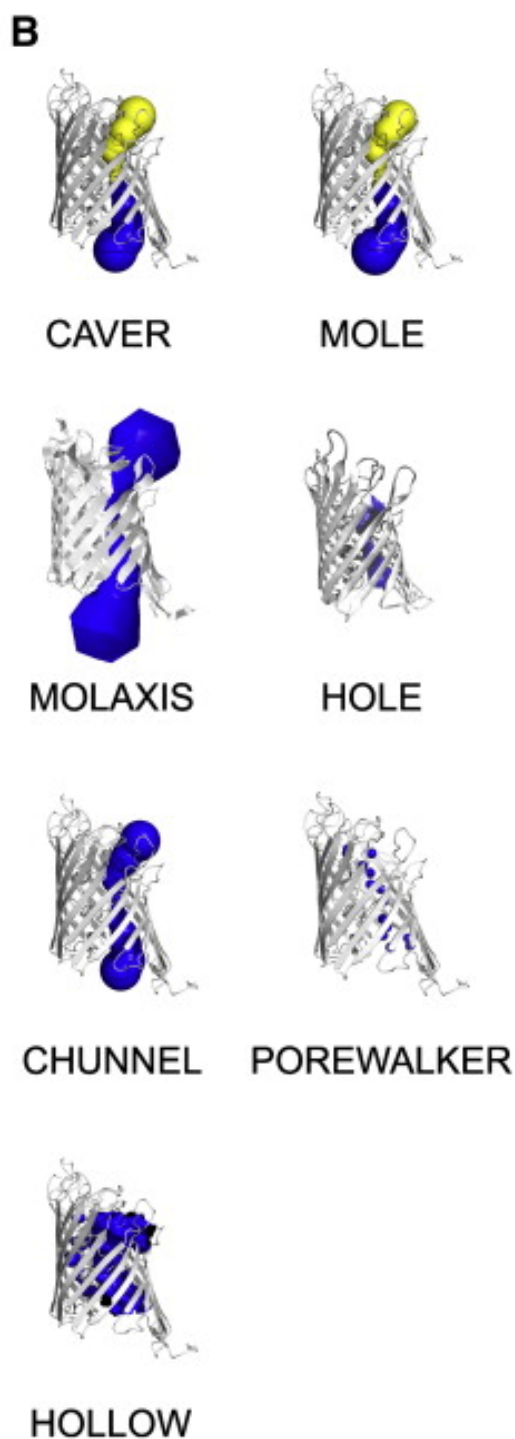


Obrázek 3.2: Automaticky vypočtené tunely ve struktuře dehalogenázy DhaA (1cqW), převzato z ChannelsDB. ChannelsDB používá k hledání tunelů a pórů program MOLE 2. V této struktuře bylo jako startovní bod tunelu použit aktivní místo anotované v Catalytic Site Atlas.

3.4.2 Hledání pórů

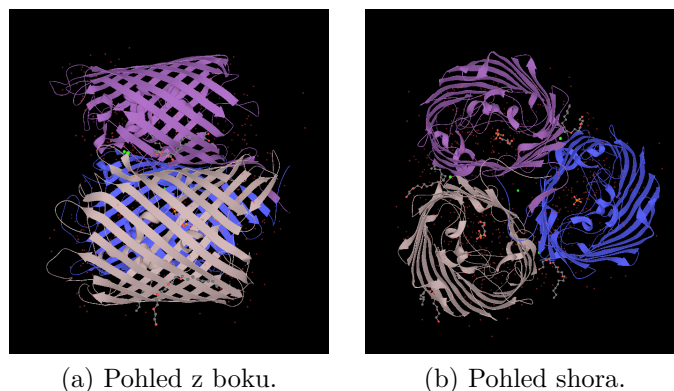
Programy, které dokáží v proteinové struktuře vyhledat tunely, zpravidla dokážou vyhledat i póry. V porovnání na obrázku 3.3 v sekci B můžeme vidět výstup programů CAVER a MOLE. Použitá verze programu MOLE neposkytovala uživateli nástroje k automatické detekci startovních bodů (Petřek a kol., 2007). Tato vlastnost je k dispozici v programu MOLE 2 (Sehnal a kol., 2013). CAVER ve svých vyšších verzích zcela změnil přístup vyhledávání. MolAxis rozlišuje, zda vyhledává póry nebo tunely (Yaffe a kol., 2008a), jeho použití při vyhledávání pórů se tedy úplně liší od použití při vyhledávání tunelů.

Všechny programy kromě HOLE a HOLLOW se dopouštějí nepřesností v popisu vnitř-



Obrázek 3.3: Výstupy některých programů na hledání pórů na stejné proteinové struktuře, převzato z Brezovsky a kol. (2013), upraveno. Zobrazeny výstupy na monomeru porinu OprP (2o4v). Při porovnání jsou použity verze MOLE 1 a CAVER 1.

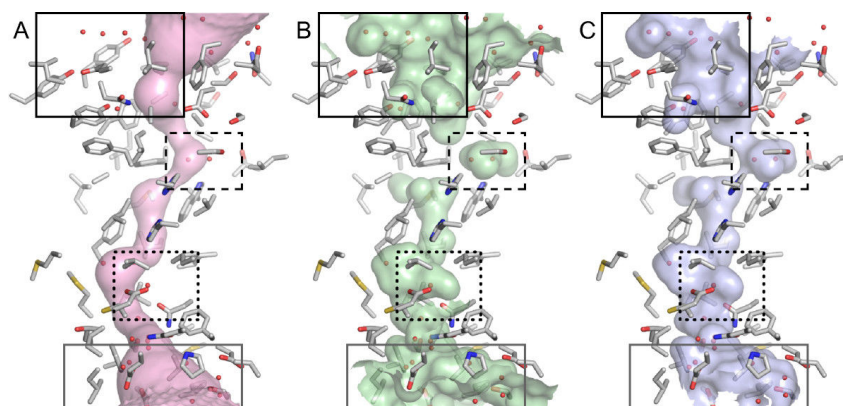
ního prostoru póru, pokud nemá na průřezu kruhový tvar, protože prostor vyplňují koulemi (v případě programu PoreWalker válci nebo kuželovými výsečemi (Pellegrini-Calace a kol., 2009)). Jako v případě zkoumané molekuly OprP 3.4, kruhový průřez nemají



Obrázek 3.4: Trimer porinu OprP, převzato z PDB

všechny póry. Průřez v případě OprP se spíše podobá elipse. V takovém případě se pomocí vložení útvaru s kruhovým průřezem pokryje pouze kruh o průměru rovném délce kratší ose elipsy. HOLE programu předchází používáním kapsulí namísto koulí při vyplňování póru. (Brezovsky a kol., 2013)

HOLLOW spoléhá na analýzu povrchu póru pomocí vyplňování koulemi výrazně menšími než průměr póru a získává mapu povrchu. Obrázek 3.5 znázorňuje, jak program, který vyplňuje pór koulemi, podhodnocuje velikost póru (zvláště v oblastech zvýrazněných obdélníky). Výsledný rozdíl je velmi dobře viditelný i na porovnání na obrázku 3.3.



Obrázek 3.5: Pór v proteinu Rh50 vypočtený programy MOLE 1 (A) a HOLLOW (C), spolu s referenčním povrchem póru (B), převzato z Ho a Gruswitz (2008). Výrazné rozdíly mezi jednotlivými oblastmi jsou zvýrazněny pomocí obdélníků.

Zjednodušení plynoucí z vyplňování póru tělesy s kruhovým průřezem (popřípadě s průřezem tvaru elipsy) nám poskytuje jednoduché a efektivní nástroje k měření šířky v nejúžších místech, která často působí jako regulační či hlavní výběrová. Při hledání šířky bottlenecku pomocí programu HOLLOW může být přesnost výpočtu povrchu póru i překážkou.

Dalším problematickým prvkem je hledání dvou blízkých pórů v jedné proteinové struktuře. Pouze program CHUNNEL je vybaven nástroji k jejich hledání, při jejich hledání pomocí dalších programů musí být hlavní pór zamaskován. (Brezovsky a kol., 2013)

Program HOLE je ze zde popsaných programů nejstarší, ale stále je používán a udržován. Při tvorbě nových programů se používá jako benchmark, oproti kterému by se program měl zlepšit, a při jeho publikování je často uveřejněn i příklad, kdy nový program skutečně lepší je, v roli benchmarku však dosud nebyl nahrazen. Jeho určitou nevýhodou je použití nedeterministického Monte Carlo algoritmu. Při opětovném spuštění se výsledky mohou lišit.

3.5 Výběr programu

Jak už bylo uvedeno, při výběru programu pro analýzu pórů a tunelů proteinu hraje zásadní roli, jaké prvky očekáváme a jaké informace máme o proteinu k dispozici.

V případě, že hledáme tunely a nemáme o nich žádnou informaci, v úvahu připadají programy MOLE, MolAxis a HOLLOW. Z výše uvedeného porovnání vyplývá, že program HOLLOW neposkytuje plně informativní popis tunelů, zvláště v případě jejich shlukování.

MOLE a MolAxis pracují sobě navzájem velmi podobně. MOLE nebere při výpočtu v úvahu rozdíly ve Van der Waalsově poloměru atomů. Je možné hledat póry ve struktuře bez započítání vodíkových atomů, které mají menší poloměr než atomy uhlíku, dusíku a kyslíku, a tím minimalizovat zkreslení, které program MOLE může vnést (Brezovsky a kol., 2013). Rozhodnutí by mohlo ovlivnit i to, s jakými programy na vizualizaci výsledků je program spojen a jaké má uživatel k dispozici. Pro ověření biologické relevance výsledků je možné použít program dxTuber nebo analýzu molekulární dynamiky.

V případě, že víme, odkud hledané tunely mají vycházet, CAVER 3 poskytuje nástroje k nalezení tunelů i k jejich validaci v kontextu molekulární dynamiky bez nutnosti srovnávání výstupů z více programů.

Pokud hledáme póry, známe jejich přibližný směr a umístění a víme, že ve struktuře je jeden pór (popřípadě postačuje najít pouze hlavní pór), program HOLE poskytuje spolehlivý výstup a kvalitní nástroje k měření póru. Při porušení některého z těchto předpokladů lze ve výše uvedených nalézt program, který je navržen pro danou situaci, mohou však být zatíženy dalšími podmínkami použití (například PoreWalker je použitelný pouze pro transmembránové proteiny).

Při hledání pórů bez jakýchkoli informací nebo při porušení více výše uvedených podmínek může být užitečné použít programy na hledání tunelů CAVER nebo MOLE a (pokud službu poskytují a pokud je to možné) nalezené tunely spojit do pórů procházejících skrz proteinovou strukturu. Při tomto přístupu dokážeme najít více pórů procházejících přes proteinovou strukturu, aniž o pórech cokoliv věděli.

Použití HOLLOW je na místě, pokud již známe hlavní informace o tunelu či póru a chceme detailně prozkoumat jeho vnitřní povrch, například k určení povrchové struktury aktivních míst uvnitř.

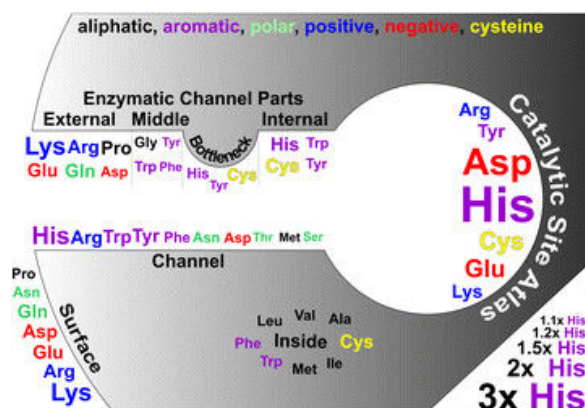
Porovnání výstupů výše popsaných programů, popřípadě programů samotných, je obtížné, protože zatím neexistuje žádné objektivní měření podobnosti. Částí řešení tohoto problému by mohla být níže popsaná ChannelsDB.

4. Příklady využití programů

4.1 Použití při zkoumání proteinů

Velkým přínosem analýzy proteinů *in silico* je možnost hledat podobné prvky ve velkých datasetech. Zkoumanými prvky mohou být tunely jako takové, v nichž můžeme zkoumat zastoupení aminokyselin, délku tunelu a další prvky. Na obrázku 4.1 je ilustrováno zastoupení aminokyselin různých vlastností v enzymech. Mezi nalezenými tunely jsou častější hydrofilní (k povšimnutí na obrázku 4.1 - zastoupení alifatických a nepolárních aminokyselin je výrazně nižší než zastoupení ostatních aminokyselin)

Kromě zkoumání vlastností enzymatických tunelů práce Pravda a kol. (2014) odhaluje, že regulace přístupu k aktivnímu místu enzymu je běžně regulována pomocí tunelu - tunel dlouhý alespoň 5 Å je přítomen ve více než 86% zkoumaných enzymů.



Obrázek 4.1: Statistické zastoupení aminokyselin v různých částech tunelů a kavít enzymů, převzato z Pravda a kol. (2014).

Programy k detekci tunelů lze použít i jako dílčí nástroj při analýze konformačních změn ve vodě oproti organickému rozpouštědлу. Program MOLE byl použit spolu s programem k detekci dutin Fpocket k prozkoumání a porovnání tunelů a dutin v proteinových strukturách v různých médiích (Rueda a kol., 2018).

4.2 Použití při designu proteinů

Protože tunely, kavity a póry jsou funkčním prvkem proteinu, nástroje k jejich analýze *in silico* mohou být velmi užitečné, zvláště v případě, kdy strukturu navrhujeme nebo upravujeme pomocí mutantních residuů a nemáme možnost zkoumat fyzickou molekulu.

Program CAVER byl použit při designu cytochromu P450B6, který v játrech aktivuje cyklofosamid (CPA), cytotoxickou látku, která se používá při chemoterapii. Jaterní

cytochrom P4502B6 katalyzuje CPA na 4-OH-CPA, který je následně krví doručen k tumoru. Cytochrom P4502B6 má však k CPA poměrně malou afinitu, což bylo důvodem pro pokus o zlepšení této vlastnosti. Krystalografická struktura cytochromu P4502B6 byla v době výzkumu neznámá, proto byla struktura určena pomocí homologního modelování. Tunely a přístupové cesty k aktivnímu místu ukrytému uprostřed struktury byly vypočteny pomocí programu CAVER. CAVER byl také použit při zkoumání stability přístupových tunelů k aktivnímu místu mutovaného proteinu. Chen (2004)

4.3 ChannelsDB

ChannelsDB je novým projektem, jehož snahou je poskytnout sdružené informace o makromolekulárních pórech a tunelech, jejich geometrii v molekule a jejich fyzikálních a chemických vlastnostech. Je úzce propojena s PDBe a s Uniprotom. Kanály obsažené v proteinových strukturách získává buď z literatury (k dubnu 2018 obsahuje 294 manuálně anotovaných struktur a 15500 struktur s anotací z Catalytic Site Atlas) a v dalších strukturách byly póry vypočteny automaticky pomocí MOLE 2 programu (Pravda a kol., 2018). Tato práce umožňuje nalezení tunelu či póru v určitých strukturách bez nutnosti ladění parametrů programu nebo ověřit výsledky vlastní analýzy. V budoucnu by mohla nahradit program HOLE ve funkci benchmarku pro nové i stávající programy díky tomu, že sdružuje manuálně ověřené tunely póry.

Závěr

Prázdné prostory v proteinové struktuře jsou různorodé. Podle svého tvaru se dělí na póry, tunely, kavity a kapsy. Automatická detekce každé z těchto podskupin spoléhá na specifické algoritmické přístupy.

Při detekci pórů v proteinu se lze zaměřit na geometrii struktury, na analýzu povrchu či na specifické vlastnosti transmembránových proteinů, v nichž se póry běžně vyskytují jakospojení dvou stran membrány. Detekce tunelů je obvykle omezena jen na geometrické přístupy nebo na analýzu povrchu proteinu. Výskyt tunelů v proteinu není omezen na třídu o specifických vlastnostech, kterých by se při detekci dalo využít.

Výběr programu k analýze může záviset na množství výchozích informací o póru i o proteinu. Při hledání pórů mohou být s dobrým výsledkem použity i programy primárně cílené na hledání tunelů. Dosud však neexistuje univerzální program, který by dokázal bez výchozích informací spolehlivě nalézat póry a tunely ve všech typech proteinových struktur. Obecně je k vyslovení relevantní predikce nutná kombinace programů, kombinace programu a dalšího zdroje (jakým může být například Catalytic Site Atlas) nebo náhled člověka.

Seznam použité literatury

- ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K. a WALTER, P. (2002). Membrane Proteins. In *Molecular Biology of the Cell, 4th edition*. Garland Science.
- BARBER, C. B., DOBKIN, D. P. a HUHDANPAA, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, **22**(4), 469–483. ISSN 00983500. doi: 10.1145/235815.235821.
- BERKA, K., HANÁK, O., SEHNAL, D., BANÁŠ, P., NAVRÁTILOVÁ, V., JAISWAL, D., IONESCU, C. M., SVOBODOVÁ VAŘEKOVÁ, R., KOČA, J. a OTYEPKA, M. (2012). MOLEonline 2.0: Interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Research*, **40**(W1), W222–W227. ISSN 03051048. doi: 10.1093/nar/gks363.
- BREZOVSKY, J., CHOVANCOVA, E., GORA, A., PAVELKA, A., BIEDERMANNNOVA, L. a DAMBORSKY, J. (2013). Software tools for identification, visualization and analysis of protein tunnels and channels. ISSN 07349750.
- CHEN, C.-S. (2004). Activation of the Anticancer Prodrugs Cyclophosphamide and Ifosfamide: Identification of Cytochrome P450 2B Enzymes and Site-Specific Mutants with Improved Enzyme Kinetics. *Molecular Pharmacology*, **65**(5), 1278–1285. ISSN 0026-895X. doi: 10.1124/mol.65.5.1278.
- CHOVANCOVA, E., PAVELKA, A., BENES, P., STRNAD, O., BREZOVSKY, J., KOZLIKOVÁ, B., GORA, A., SUSTR, V., KLVANA, M., MEDEK, P., BIEDERMANNNOVA, L., SOCHOR, J. a DAMBORSKY, J. (2012). CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Computational Biology*, **8**(10). ISSN 1553734X. doi: 10.1371/journal.pcbi.1002708.
- COJOCARU, V., WINN, P. J. a WADE, R. C. (2007). The ins and outs of cytochrome P450s. *Biochimica et Biophysica Acta (BBA) - General Subjects*, **1770**(3), 390–401. ISSN 0304-4165. doi: 10.1016/J.BBAGEN.2006.07.005.
- COLEMAN, R. G. a SHARP, K. A. (2009). Finding and characterizing tunnels in macromolecules with application to ion channels and pores. *Biophysical Journal*, **96**(2), 632–645. ISSN 15420086. doi: 10.1529/biophysj.108.135970.
- DIJKSTRA, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**(1), 269–271. ISSN 0029599X. doi: 10.1007/BF01386390.
- FURNHAM, N., HOLLIDAY, G. L., DE BEER, T. A. P., JACOBSEN, J. O. B., PEARSON, W. R. a THORNTON, J. M. (2014). The Catalytic Site Atlas 2.0: Cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Research*, **42**(D1). ISSN 03051048. doi: 10.1093/nar/gkt1243.
- HARRIES, W. E. C., AKHAVAN, D., MIERCKE, L. J. W., KHADEMI, S. a STROUD, R. M. (2004). The channel architecture of aquaporin 0 at a 2.2-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(39), 14045–50. ISSN 0027-8424. doi: 10.1073/pnas.0405274101.

- HO, B. K. a GRUSWITZ, F. (2008). HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures. *BMC Structural Biology*, **8**, 49. ISSN 14726807. doi: 10.1186/1472-6807-8-49.
- LEE, B. a RICHARDS, F. M. (1971). The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, **55**(3), 379–IN4. ISSN 00222836. doi: 10.1016/0022-2836(71)90324-X.
- PAVELKA, A., SEBESTOVA, E., KOZLIKOVÁ, B., BREZOVSKÝ, J., SOCHOR, J. a DAMBORSKÝ, J. (2016). CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **13**(3), 505–517. ISSN 15455963. doi: 10.1109/TCBB.2015.2459680.
- PELLEGRINI-CALACE, M., MAIWALD, T. a THORNTON, J. M. (2009). PoreWalker: A novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLoS Computational Biology*, **5**(7), e1000440. ISSN 1553734X. doi: 10.1371/journal.pcbi.1000440.
- PETŘEK, M., OTYEPKA, M., BANÁŠ, P., KOŠINOVÁ, P., KOČA, J. a DAMBORSKÝ, J. (2006). CAVER: A new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, **7**(1), 316. ISSN 14712105. doi: 10.1186/1471-2105-7-316.
- PETŘEK, M., KOŠINOVÁ, P., KOČA, J. a OTYEPKA, M. (2007). MOLE: A Voronoi Diagram-Based Explorer of Molecular Channels, Pores, and Tunnels. *Structure*, **15**(11), 1357–1363. ISSN 09692126. doi: 10.1016/j.str.2007.10.007.
- PORTER, C. T. (2004). The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Research*, **32**(90001), 129D–133. ISSN 1362-4962. doi: 10.1093/nar/gkh028.
- PRAVDA, L., BERKA, K., SVOBODOVÁ VAŘEKOVÁ, R., SEHNAL, D., BANÁŠ, P., LASKOWSKI, R. A., KOČA, J. a OTYEPKA, M. (2014). Anatomy of enzyme channels. *BMC Bioinformatics*, **15**(1), 379. ISSN 14712105. doi: 10.1186/s12859-014-0379-x.
- PRAVDA, L., SEHNAL, D., SVOBODOVÁ VAŘEKOVÁ, R., NAVRÁTILOVÁ, V., TOUŠEK, D., BERKA, K., OTYEPKA, M. a KOČA, J. (2018). ChannelsDB: database of biomacromolecular tunnels and pores. *Nucleic Acids Research*, **46**(D1), D399–D405. ISSN 0305-1048. doi: 10.1093/nar/gkx868.
- RAUNEST, M. a KANDT, C. (2011). dxTuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *Journal of Molecular Graphics and Modelling*, **29**(7), 895–905. ISSN 10933263. doi: 10.1016/j.jmgm.2011.02.003.
- RUEDA, A. J. V., MONZON, A. M., ARDANAZ, S. M., IGLESIAS, L. E. a PARISI, G. (2018). Large scale analysis of protein conformational transitions from aqueous to non-aqueous media. *BMC Bioinformatics*, **19**(1), 27. ISSN 14712105. doi: 10.1186/s12859-018-2044-2.
- SEHNAL, D., VAŘEKOVÁ, R. S., BERKA, K., PRAVDA, L., NAVRÁTILOVÁ, V., BANÁŠ, P., IONESCU, C. M., OTYEPKA, M. a KOČA, J. (2013). MOLE 2.0: Advanced approach for analysis of biomacromolecular channels. *Journal of Cheminformatics*, **5**(8), 39. ISSN 17582946. doi: 10.1186/1758-2946-5-39.

- SHRAKE, A. a RUPLEY, J. A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *Journal of Molecular Biology*, **79**(2), 351–371. ISSN 00222836. doi: 10.1016/0022-2836(73)90011-9.
- SMART, O. S., NEDUVELIL, J. G., WANG, X., WALLACE, B. a SANSOM, M. S. (1996). HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *Journal of Molecular Graphics*, **14**(6), 354–360. ISSN 0263-7855. doi: 10.1016/S0263-7855(97)00009-X.
- VOSS, N. R., GERSTEIN, M., STEITZ, T. A. a MOORE, P. B. (2006). The Geometry of the Ribosomal Polypeptide Exit Tunnel. *Journal of Molecular Biology*, **360**(4), 893–906. ISSN 00222836. doi: 10.1016/j.jmb.2006.05.023.
- YAFFE, E., FISHELOVITCH, D., WOLFSON, H. J., HALPERIN, D. a NUSSINOV, R. (2008a). MolAxis: Efficient and accurate identification of channels in macromolecules. *Proteins: Structure, Function and Genetics*, **73**(1), 72–86. ISSN 08873585. doi: 10.1002/prot.22052.
- YAFFE, E., FISHELOVITCH, D., WOLFSON, H. J., HALPERIN, D. a NUSSINOV, R. (2008b). MolAxis: a server for identification of channels in macromolecules. *Nucleic acids research*, **36**(Web Server issue), W210–5. ISSN 13624962. doi: 10.1093/nar/gkn223.