

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Katedra biochemie



Struktura a vizualizace proteinových tunelů

DIPLOMOVÁ PRÁCE

Autor:	Bc. Anna Špačková
Studijní program:	N1406 - Biochemie
Studijní obor:	Bioinformatika
Forma studia:	Prezenční
Vedoucí práce:	doc. RNDr. Karel Berka, Ph.D.
Rok:	2022

Prohlašuji, že jsem diplomovou práci vypracovala samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním diplomové práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, ve znění pozdějších předpisů.

V Olomouci dne

.....
Podpis studenta

Poděkování

Chtěla bych poděkovat vedoucímu své diplomové práce doc. RNDr. Karlovi Berkovi, Ph.D. za odborné vedení, čas a úsilí které věnoval konzultacím ohledně mé diplomové práce. Také bych chtěla poděkovat rodině za podporu po celou dobu studia.

Bibliografická identifikace

Jméno a příjmení autora	Bc. Anna Špačková
Název práce	Struktura a vizualizace proteinových tunelů
Typ práce	Diplomová
Pracoviště	Katedra fyzikální chemie
Vedoucí práce	doc. RNDr. Karel Berka, Ph.D.
Rok obhajoby práce	2022

Abstrakt

Práce analyzuje struktury tunelů v proteinech v databázi ChannelsDB a z výsledků MOLEonline spolu s jejich vlastnostmi a predikcí možných procházejících molekul. Dalším cílem bylo napsat program, který vizualizuje tunely ve 2D reprezentaci spolu s aminokyselinovým zastoupením na povrchu tunelu v proteinech. Dosažení výsledků bylo uskutečněno především pomocí vlastních kódů napsaných v programovacím jazyce Python, výsledků výpočtu z MOLEonline a dat uložených v ChannelsDB. Výsledky umožňují jednodušší integraci dat o tunelech v proteinech do databáze ChannelsDB.

Klíčová slova	Tunely, póry, kanály, protein, aminokyselina, enzym, substrát, transport, fyzikálně-chemické vlastnosti, úzké hrdlo, molekula, ionty, ligandy.
Počet stran	81
Počet příloh	1
Jazyk	Český

Bibliographical identification

Author's first name and surname	Bc. Anna Špačková
Title	Structure and Visualization of Protein Tunnels
Type of thesis	Diploma
Department	Department of physical chemistry
Supervisor	doc. RNDr. Karel Berka, Ph.D.
The year of presentation	2022

Abstract

This work analyzes the structures of tunnels in proteins in the ChannelsDB database and results from MOLEonline, along with their properties and predictions of possible passing molecules. Another goal was to write a program that visualizes tunnels in a 2D representation along with the amino acid representation on the tunnel surface in proteins. Achieving the results was primarily accomplished using custom codes written in the Python programming language, computational results from MOLEonline, and data stored in ChannelsDB. The results allow for easier integration of the protein tunnel data into the ChannelsDB database

Keywords	Tunnels, pores, channels, protein, aminoacid, enzyme, substrate, transport, physico-chemical properties, bottleneck, molecule, ions, ligands.
Number of pages	81
Number of appendices	1
Language	Czech

OBSAH

1	ÚVOD	1
2	SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY	3
2.1	Kanály, tunely a póry v proteinech	3
2.1.1	Akvaporiny	6
2.1.2	Iontové kanály	8
2.1.2.1	Napětím ovládané iontové kanály	10
2.1.3	Mechanosenzitivní kanály	11
2.1.4	Enzymové kanály	12
2.1.4.1	Struktura enzymového kanálu	14
2.1.4.2	Residua v okolí enzymového kanálu	17
2.1.4.3	Selekce látek enzymatickými tunely	18
2.1.4.4	Typy vazby ligandu na tunel	20
2.1.4.5	Průzkum enzymatických tunelů v proteinech	20
3	EXPERIMENTÁLNÍ ČÁST	22
3.1	MOLEonline	22
3.1.2	Algoritmus MOLE	23
3.1.3	Vypočítané fyzikálně-chemické vlastnosti tunelu	24
3.1.4	Vizualizace a analýza výsledků v MOLEonline	26
3.2	Databáze tunelů ChannelsDB	27
3.3	Python	28
3.3.1	Metody, knihovny, moduly	28
3.3.2	Knihovna Pandas	29
3.3.3	Python Arcade	30
3.3.4	Python Matplotlib	31
3.3.5	Qt Designer + knihovna PyQt5	31
4	VÝSLEDKY	32
4.1	Analýza zkoumaných dat v MOLE	32
4.1.1	Analýza input.xml	32
4.1.2	Analýza tunnels.xml	37
4.1.3	Analýza zkoumaných struktur	40
4.2	Program sloužící k 2D vizualizaci tunelu proteinu	41
4.2.1	Načtení dat a jejich modifikace	42
4.2.2	Vykreslení os	45
4.2.3	Vykreslení obrysu tunelu	45
4.2.4	Vykreslení aminokyselin	48
4.2.5	Barva vrstev korelující s jejich vlastnostmi	52
4.2.6	Úprava parametrů	54
4.3	Analýza látek procházejících tunely	56
4.4	Predikce procházejících látek vzhledem k vlastnostem tunelu	62
4.4.1	Výpočet tvaru molekuly	62
4.4.2	Průchod molekul tunelem podle vypočítaných vlastností	65
4.4	Doplňování informací o tunelech do ChannelsDB	66
4.5	GUI sloužící k ovládání programů	67
5	DISKUZE	70
6	ZÁVĚR	72
7	SUMMARY	73
8	LITERATURA	74
9	SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK	78

CÍLE PRÁCE

Teoretická část

- Vypracování literární rešerše v oblasti typů proteinových tunelů v různých typech proteinů.
- Prostudování uvedené odborné literatury.

Praktická část

- Doplnování databáze kanálů ChannelsDB pomocí softwaru MOLE doplnovanou informacemi o odpovídajících ligandech.
- Vizualizace proteinových tunelů ve 2D reprezentaci spolu s aminokyselinovým zastoupením na povrchu proteinového tunelu.
- Analýza látek procházejících tunely a snaha o predikci těchto látek vzhledem ke struktuře proteinového tunelu.

1 ÚVOD

Proteiny zajišťují rozmanité funkce v živých organismech. Vyskytují se jako stavební materiál buněk, hrají důležitou roli v regulaci genové exprese, signalizaci, transportu látek a enzymatické katalýze malých molekul. Skládají se z jednoho nebo více řetězců. Uspořádávají se do složitých struktur prostorové konformace, v níž obsahují mnoho výčnělků nebo dutin. Prázdná místa jsou z hlediska stability nevýhodná, a proto jejich výskyt bývá spojen s funkcí ať už jde o aktivní či alosterická místa, tunely, kanály a póry. Anatomie a vlastnosti těchto podstruktur významně ovlivňují interakce bílkovin s malými molekulami.

Kanál je obecné pojmenování pro tunely a póry. Tunely spojují aktivní místo s povrchem biomakromolekuly, zatímco póry typicky prochází skrz molekulu a mají otvory na obou stranách. Póry hrají důležitou roli v transportu iontů, nebo jiných molekul prostřednictvím buněčných membrán. Tunely se vyskytují především v globulárních proteinech s katalytickou funkcí a póry pak u membránových proteinů. (Stourac et al., 2019) Kanály (tunely a póry) jsou velmi důležitými strukturálními rysy biomakromolekul úzce spjatých s jejich biologickou funkcí. (Pravda et al., 2017)

Pochopení transportních drah, jejich vlastností a dopadu na průchod ligandů je důležité pro dešifrování funkce proteinu i pro praktické aplikace v oblasti proteinového inženýrství a návrhu léčiv. Studium přístupových cest a transportních procesů ligandů pomocí experimentálních technik není zdaleka triviální. Kvantitativní popis může být získán nepřímo pomocí kinetických měření. Může probíhat také přímo pomocí dostupných přímých metod, jako je časově rozlišená krystalografie pod tlakem xenonu, která poskytuje pouze informace o přítomnosti kanálu, ale za cenu enormních nákladů. Funkce kanálů se proto častěji studují *in silico* (např. Caver a MOLE programy).

Vyvíjené nástroje umožňují identifikaci a analýzu kanálů ve známých strukturách proteinů. Informace o strukturách proteinu se povětšinou získávají z dostupných databází, data o zjištěných tunelech ve strukturách se dále ukládají do databází určených pro proteinové kanály.

Vzhledem k důležitosti proteinových kanálů je dobré vědět jaké látky jimi mohou procházet, průchod ovlivňují vlastnosti kanálů a molekul samotných. Proto je součástí diplomové práce provedení analýzy kanálů uložených v databázi ChannelsDB a získání

informací o procházejících látkách. Následně také snaha o automatickou predikci procházejících látek s ohledem na jejich vlastnosti.

Získané informace o tunelech a ligandech budou vepisovány do databáze ChannelsDB, aby se podpořilo její rozšíření, které do budoucna bude znamenat dostupnější přehled o proteinových kanálech na jednom místě.

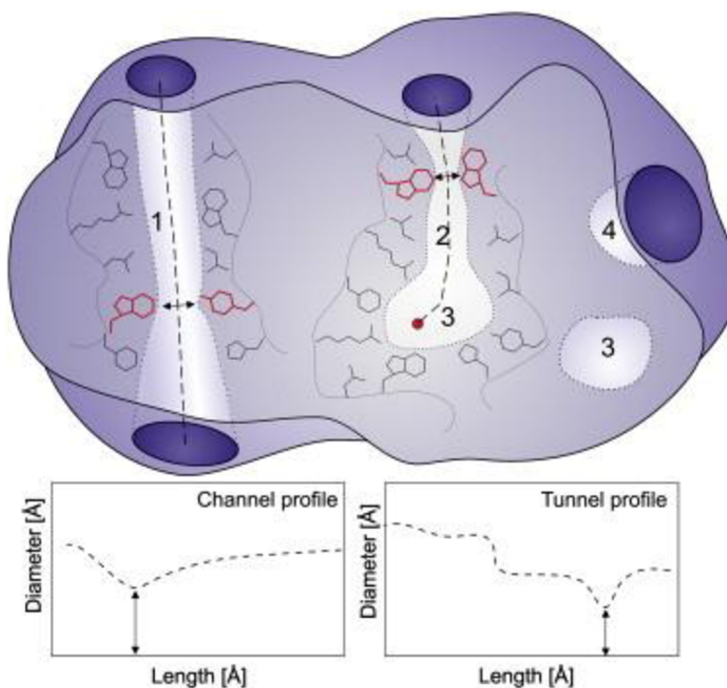
V MOLEonline (<https://mole.upol.cz/>) dochází k 3D vizualizaci proteinu a nalezených kanálů. Spolu s tím je zde zobrazená také vizualizace kanálů ve 2D. Protože tato vizualizace neobsahuje informace o okolních residuích kanálu bude v rámci diplomové práce naprogramován kód, který umožňuje 2D vizualizaci na základě vypočítaných kanálů v MOLEonline.

2 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

2.1 Kanály, tunely a póry v proteinech

Kanály (Obr. 1) se rozdělují vzhledem k jejich struktuře na i) póry, které se otevírají na opačných stranách proteinu a vedou skrz něj a ii) tunely, které spojují aktivní místo nacházející se ve vnitřní části proteinu s vnějším prostorem, putuje přes něj molekula substrátu a produkt. (Hollander et al., 2021) Selektivita kanálů je řízena velikostí, tvarem a fyzikálně-chemickými vlastnostmi zbytků, které je lemují (Brezovsky et al., 2013). Velké množství proteinů obsahuje tunely nebo póry, které mají důležitou fyziologickou roli. Konkrétní případy představují proteinové iontové kanály, pumpy, poriny nebo transportéry. Ačkoli některé struktury mají jeden kanál s jednoduchou strukturou, ve většině jde o komplikovanější struktury (Coleman & Sharp, 2009).

Vnitřní enzymové kanály jsou nezbytné pro katalýzu komplexních reakcí, kdy mutace vyskytující se v místě tunelu může ovlivnit jak aktivitu enzymu, tak substrátovou specifitu. Specifičnost enzymu je určena nejen interakcemi protein-substrát v aktivním místě, ale je ovlivněna i selektivitou těchto přístupových tunelů. (Chaloupková et al., 2003)



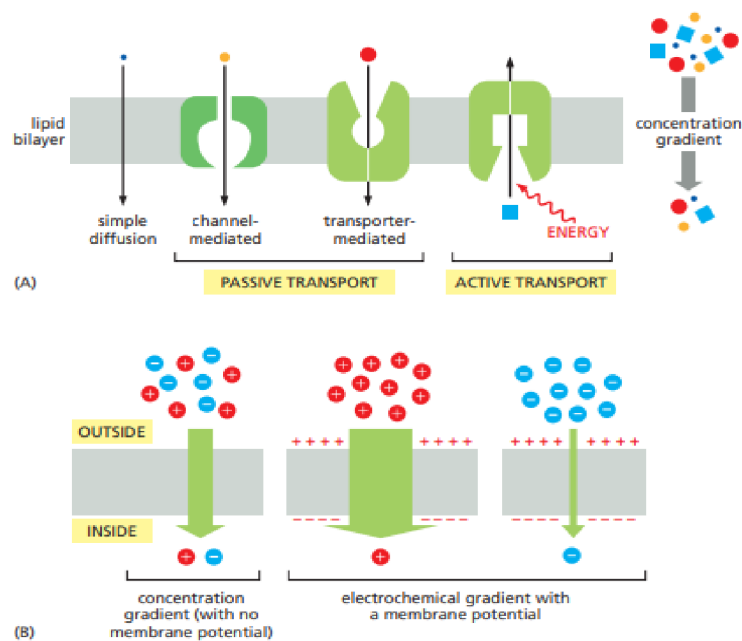
Obr. 1: Vizualizace proteinu s 1) pórem, 2) tunelem, 3) vnořenou dutinou a 4) povrchovou štěrbinou či rýhou. Světlé oblasti představují vnitřní část prohlubně, tunelu nebo póru. Šipky označují úzké hrdlo (*bottleneck*). Grafy ukazují, jak se od sebe odlišují profily kanálu a tunelu. Převzato od (Brezovsky et al., 2013).

Póry (Obr. 1) hrají důležitou roli ve velkém množství biologických procesů, například se podílí na kontrole transportu vody, iontů a ostatních molekul skrz biomembrány (Pravda et al., 2018 a). Celkově jsou důležité pro řízení transportu přes buněčnou membránu (Pravda et al., 2018 b). Takové proteiny se vyskytují v mnoha formách a ve všech typech biologických membrán.

Zlepšující se techniky, především kryoelektronové mikroskopie, sloužící ke zjištění struktury proteinu umožnily celkově kvalitnější analýzu kanálů v proteinech. Krystalografická analýza a simulace molekulární dynamiky potvrdily úspěšné zavedení strukturálně definovaného a funkčního transportního tunelu. (Brezovsky et al., 2016)

Znalost struktury proteinu je důležitá především u vývoje léků (vyhledávání inhibitorů), dále také v enzymologii, kde se identifikují substrátové molekuly (na principu zámku a klíče), a v chemii pro studium proteinových interakcí (Petrek et al., 2006). Kvůli velkému zájmu o tyto struktury, vznikají nástroje k identifikaci a analýze kanálů v proteinech. Pro identifikaci pórů je jeden z nejpoblárnějších nástrojů HOLE, pro tunely je to program Caver a pro obojí MOLE a MOLEonline. Je dostupných mnoho dalších programů a softwarů s odlišnou funkčností a různou vizualizací. Bylo vytvořeno několik paralelních databází shromažďující informace o proteinových kanálech, většina z nich se ale zaměřuje na vlastnosti a ontologii proteinů, místo zaměření na vlastnosti struktury kanálů. (Pravda et al., 2018 b)

Jednou z funkcí pórů je transport. Existují dvě hlavní třídy membránového transportu, a to přes kanály, nebo transportéry (Obr. 2). Transportéry (nosiče) vážou konkrétní látku, která se má transportovat a podstoupí řadu konformačních změn, které střídavě vystavují vazebná místa pro transportovanou látku na jedné straně membrány a pak na druhé, aby se přes něj přenesla. Póry se naopak vzájemně ovlivňují mnohem slaběji s danou látkou, která má být přenášena. Když jsou póry otevřené umožňují průchod membránou specifickým látkám. Průchod póry je vždy pasivní a mnohem rychlejší než průchod transportéry. (Alberts et al., 2015)



Obr. 2: Různé formy membránového transportu, pasivní transport (A) koncentračním gradientem, nebo (B) elektrochemickým gradientem. K transportu dochází spontánně přes lipidovou dvojvrstvu přes póry, tedy pasivní transportéry. Naopak aktivní transport vyžaduje vstup energie a je zprostředkován transportéry. Převzato od (Alberts et al., 2015).

Každý protein často transportuje pouze konkrétní molekulární druh nebo třídu molekul (např. ionty, cukry, aminokyseliny). Póry vedoucí skrz transmembránové proteiny (Obr. 1) umožňují selektivní transport iontů, malých molekul, a dokonce i velkých makromolekul přes biologické membrány. Selektivita kanálu je určena jeho geometrií a složením aminokyselinových zbytků v okolní oblasti. (Gouaux & Mackinnon, 2005) Všechny membránové transportní proteiny jsou transmembránové proteiny, jejichž polypeptidový řetězec je zanořen do lipidové struktury membrány. Dysfunkce kanálů souvisí s různými nemocemi. (Niemeyer et al., 2001)

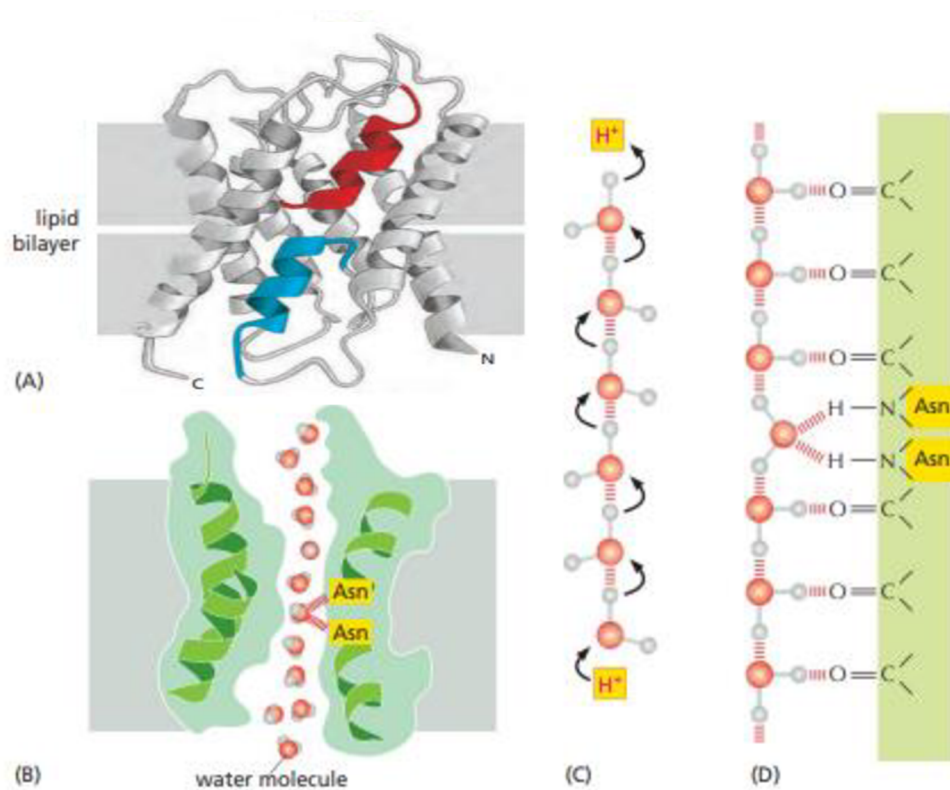
Vzhledem k odlišným vlastnostem kanálů je lze rozdělit do podskupin podle toho, jaké látky transportují. V těchto množinách jsou pak zařazeny kanály se stejnými, nebo podobnými vlastnostmi, aby umožnili průchod specifické molekuly. Typickými třídami, které jsou v textu dále podrobněji popsány mohou být akvaporiny, schopné transportovat vodu, které vzhledem k vysokému podílu vody v organismech hrají zásadní roli v rychlém transportu molekul vody. Důležitou skupinou jsou také iontové kanály a jejich podtřída napětím ovládané iontové kanály, obě skupiny transportují ionty v organismech.

Za zmínku stojí mechanosenzitivní kanály, udržující správný osmotický tlak buňky. Další významnou skupinou jsou tunely vyskytující se v enzymech, které svými fyzikálně chemickými vlastnostmi selektují látky schopné vstupovat do blízkosti ukrytého aktivního místa v tunelu.

2.1.1 Akvaporiny

Buňky si udržují osmotický tlak v reakci na okolní podmínky přímou difúzí. Krom difúze mají některé prokaryotické a eukaryotické buňky vodní kanály neboli akvaporiny. Ty leží v jejich plazmatické membráně, umožňují rychlejší pohyb molekul vody směrem z a do buněk. Akvaporiny představují selektivně propustné kanály, které jsou nepropustné pro ionty, ale propustné pro vodu. Všichni ostatní obratlovci, bezobratlí, mikrobi a rostliny jsou také tvořeni primárně vodou. (Agre, 2004) Akvaporiny se hojně vyskytují především v živočišných buňkách, které musí transportovat vodu vysokou rychlostí např. epitelální buňky ledvin nebo exokrinní buňky, které musí transportovat nebo vylučovat velké objemy tekutin. Rodina akvaporinů obsahuje funkčně různorodé členy, kteří nepřepřavují jen vodu, ale také glycerol, močovinu, arsenit a oxid uhličitý. (Hedfalk et al., 2006)

Aby se zabránilo narušení iontových gradientů transportem přes membrány, musí akvaporiny umožnit rychlý průchod molekul vody a zároveň zcela blokovat průchod iontů. Trojrozměrná struktura akvaporinu odhaluje, jak dosahuje takovéto selektivity. Struktura úzkého póru umožňuje molekulám vody procházet membránou v jediném kroku po dráze karbonylových kyslíků lemující jednu stranu póru (Obr. 4). Druhou stranu póru tvoří řada hydrofobních aminokyselin. Pór je příliš úzký, aby do něj mohl vstoupit jakýkoliv hydratovaný iont a energetické náklady na dehydrataci iontu by byly obrovské, proto hydrofobní stěna póru nemůže interagovat s dehydratovaným iontem, aby kompenzovala ztrátu vody. Tento popis snadno vysvětluje, proč akvaporiny nemohou vést K^+ , Na^+ , Ca^{2+} ani Cl^- ionty. (Alberts et al., 2015) Tyto kanály jsou také nepropustné pro H^+ , který je přítomný v buňkách hlavně jako H_3O^+ .



Obr. 3: Na obrázku (A) jsou vidět dvě krátké α šroubovice akvaporinu (červená a modrá), které prochází skrz membránu. Každá z těchto šroubovic prochází pouze přes polovinu membrány. (B) Představuje podélný průřez jedním monomerem v rovině centrálního póru. Lze rozeznat jednu stranu, která je tvořena z hydrofilních aminokyselin poskytující přechodné vodíkové vazby vodě. (C) Spolu s (D) vysvětluje proč jsou akvaporiny nepropustné pro H^+ , karboxylové skupiny dvou strategicky umístěných asparaginů vážou centrální vodu. Převzato z (Alberts et al., 2015).

Každý monomer akvaporinu obsahuje jakýsi filtr tvaru „přesýpacích hodin“, které skrývá konzervované strukturální jádro skládající se ze šesti transmembránových α - helixů obklopujících jedinou úzkou dráhu přes membránu. Sedmá transmembránová oblast je tvořena smyčkami B a E, které tvoří poloviční šroubovice (Obr. 3) a skládají se do kanálu z opačných stran membrány. N-terminální konce těchto polovičních šroubovic obsahují signaturní motiv akvaporinu Asn-Pro-Ala a setkávají se uprostřed dvojvrstvy (Obr. 3). Dalším strukturálním znakem, který se vyskytuje v celé rodině akvaporinů, je zužující se oblast tvořena argininem a třemi dalšími aminokyselinami, z nichž jedna je obvykle aromatická. (Sasaki, 2012) Akvaporiny obsahují strategicky umístěné asparaginy (Obr. 3), vážající se na atom kyslíku centrální molekuly vody procházejících pórem, zavádějí bipolaritu na celý sloupec molekul vody. To znemožňuje tvorbu a rozbíjení vodíkové vazby, aby se dostaly přes centrální vodu vázanou asparaginem.

Protože obě valence tohoto centrálního kyslíku nejsou k dispozici pro vodíkovou vazbu, centrální molekula vody se nemůže účastnit přenosu H^+ a pór je tak pro H^+ nepropustný. (Alberts et al., 2015)

AQP0, AQP1, AQP2 a AQP4 jsou dostupné strukturální modely akvaporinů u savců, oproti tomu bakteriální struktury akvaporinů jsou GlpF, AqpM a AQPZ a rostlinné SoPIP2, SoPIP1 (Hedfalk, 2006).

AQP1 je homotetramer obsahující čtyři nezávislé vodné kanály. Vyskytuje se v proximálních tubulech a Henleově kličce. Po začlenění do lipidových dvojvrstev tvoří protein dvourozměrné mřížky s jednotkovou buňkou obsahující dva tetramery v opačné orientaci. (Walz et al., 1997) Změna orientace je způsobena interakcemi s elektrostatickým potenciálem oblasti Asn-Pro-Ala tvořeným makrodipóly polovičních šroubovic smyček B a E, tato elektrostatická bariéra je nyní považována za hlavní příčinu vyloučení protonů (Sasaki, 2012).

AQP2 je vazopresinem regulovaný vodní kanál, který určuje propustnost vody ledvinovým sběrným kanálem. U suchozemských zvířat včetně lidí je pro přežití v nepříznivých podmínkách na souši nezbytné zabránit ztrátě vody, proto je tento sběrný kanálek důležitý. AQP2 je protein s vyšší molekulovou hmotností 35-50 kDa. U lidského proteinu AQP2 byla potvrzena homo-tetramerní struktura. (Sasaki, 2012)

2.1.2 Iontové kanály

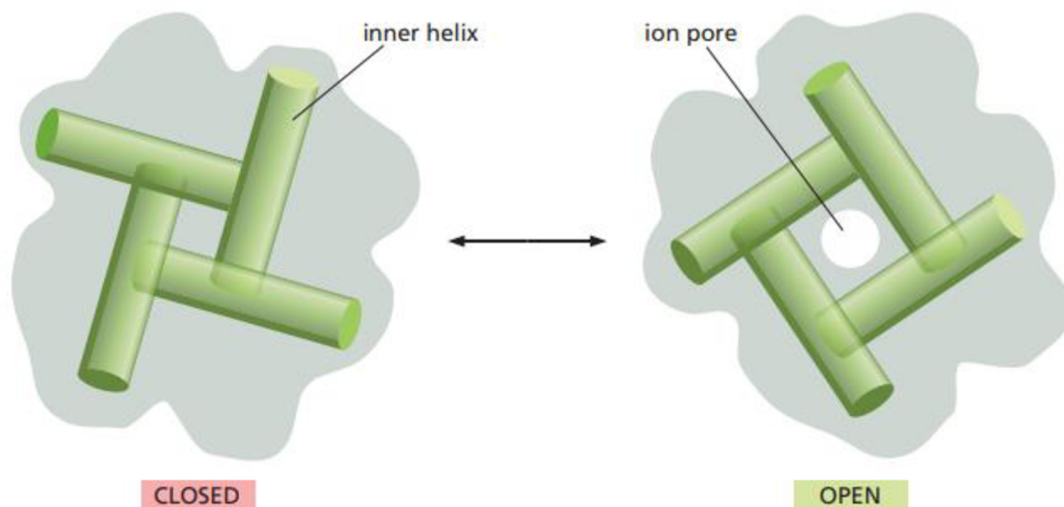
Většina kanálů v plazmatické membráně má nutně úzké a vysoce selektivní póry, které se rychle otevírají a zavírají. Protože takové proteiny se týkají specificky anorganického iontového transportu označují se jako iontové kanály. Každou sekundu může projít až 100 miliónů iontů jedním kanálem (což je 10⁵ - krát rychlejší přenos než jakýmkoliv transportérem). Kanály nelze propojit se zdrojem energie, prováděný transport je vždy pasivní (Obr. 2). Funkce iontových kanálů spočívá především v umožnění průchodu specifickým anorganickým iontům (nejčastěji Na^+ , K^+ , Ca^{2+} nebo Cl^-), aby rychle difundovaly přes lipidovou dvojvrstvu. (Alberts et al., 2015) Schopnost kontroly toku iontů přes kanály je nezbytná pro mnoho buněčných funkcí. Zejména nervové buňky se specializovaly na používání iontových kanálů. Elektrické vlastnosti biologické tkáně poskytují důležité diagnostické informace v rámci rádiových a mikrovlnných frekvencí a hrají také důležitou roli při výpočtu specifické rychlosti absorpce. (Zhang et al., 2013)

Iontové kanály jsou obvykle multimerní proteiny umístěné v plazmatické membráně. Každý z těchto proteinů tvoří průchod rozprostírající se z jedné strany membrány na druhou (Moreau et al., 2008). Jde o důležitou třídu biologických molekul, umožňující vést ionty (jako Na^+ , K^+ , Cl^-) úzkým tunelem pevného náboje. Buňky exprimují rozmanitou řadu proteinů s iontovými kanály významně se lišící ve své iontové selektivitě a v ovlivnění ligandy (jako jsou neurotransmitery) nebo membránovým napětím. (Obr. 4) Většina iontových kanálů jsou proteiny s více podjednotkami a jako takové podstupují složitou sérii posttranslačních dějů skládání, modifikace a oligomerizace, aby se dosáhlo jejich správné funkční kvartérní struktury (Eisenberg, 1998).

Iontové kanály mohou přepínat konformaci mezi uzavřeným a otevřeným stavem, vykazují iontovou selektivitu, díky které je umožněn průchod jen některým anorganickým látkám. Simulace na modelových nanopórech odhalují, že úzká ($<4 \text{ \AA}$) hydrofobní oblast může tvořit funkčně uzavřenou bránu v kanálu a může být otevřena buď malým ($\sim 1 \text{ \AA}$) zvětšením poloměru pórů nebo zvýšením polaritu. Studie modelování a simulace potvrzují důležitost hydrofobního otevírání brány v K^+ kanálech a podporují model, ve kterém je ohnutí šroubovice, která slouží jako dvířka s výstelkou pórů M2 (nebo S6 v K^+ kanálech), základem kanálového uzavírání a zavírání. Simulace jednoduchého proteinu vnější membrány naznačují, že brána může být také tvořena interakcemi nabitých postranních řetězců v póru, jak je tomu také v případě ClC kanálů (Beckstein et al., 2003).

Jejich póry musí být místy dostatečně úzké, aby přinutily prostupující ionty do těsného kontaktu se stěnami kanálu tak, že pouze ionty o správném náboji a velikosti můžou projít. Omezení průchodu a selektivitu určuje zejména nejužší místo kanálu nazývané filtr selektivity. Iontové kanály mají možnost krátce se otevřít a pak znovu zavřít. Při prodloužené stimulaci (chemické nebo elektrické), přejde většina iontových kanálů do uzavřeného stavu, ve kterém jsou odolné vůči dalšímu otevírání až do odstranění stimulu.

Ve většině případů se „brána“, kterou představuje úzké hrdlo selektivity, otevře jako odpověď na konkrétní podnět. Hlavními typy podnětů způsobujících otevření iontových kanálů jsou napětí na membráně (napěťově řízené kanály), mechanický signál (mechanicky řízené kanály), nebo vazba ligandu (ligandem řízené kanály), kanály extracelulárního mediátoru (neurotransmitter), nebo iontově řízené kanály (intracelulární mediátor je iont), nebo nukleotidové kanály (nukleotid). Kromě toho fosforylace a defosforylace proteinů reguluje aktivitu mnoha iontových kanálů (Alberts et al., 2015).



Obr. 4: Model brány bakteriálního K^+ kanálu. Kanál je vidět v řezu, čtyři vnitřní transmembránové šroubovice, které lemují póry na cytosolové straně selektivního filtru mění uspořádání, aby se zavřel cytosolový vstup do kanálu. Převzato z (Alberts et al., 2015).

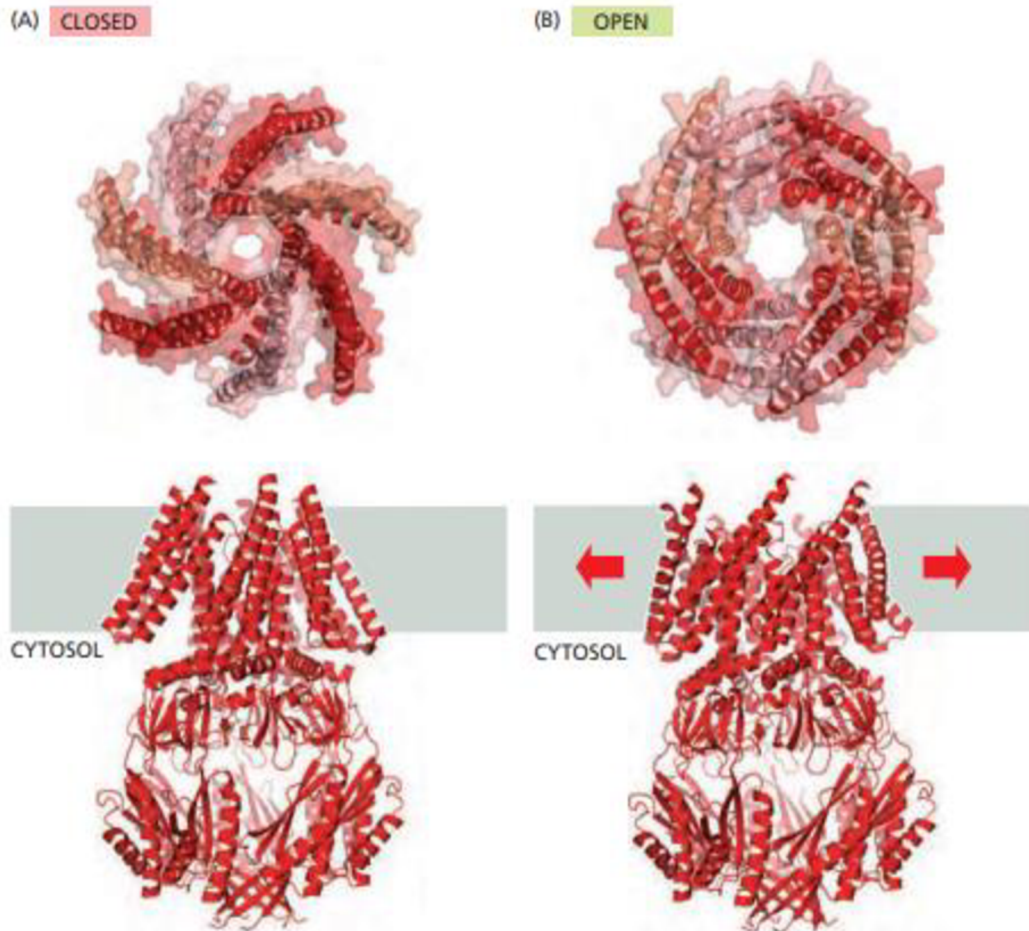
2.1.2.1 Napětím ovládané iontové kanály

Napětově řízené kanály reagují na odchylky v potenciálu buněčné membrány a jsou vysoce selektivní pro konkrétní ionty Na^+ , K^+ , Ca^{2+} a Cl^- . Dále se dělí do rodin na základě hlavního procházejícího iontu. Nicméně aminokyselinové složení známých, napětím řízených kanálů vykazuje značnou podobnost, což dokazuje, že všechny patří do nadrodiny evolučně příbuzných proteinů. Lidé, kteří zdědí mutantní geny kódující iontové kanály mohou trpět na různé nervové, svalové, mozkové nebo srdeční choroby. (Alberts et al., 2015) Napětově řízený Na^+ kanály jsou zodpovědné za vytváření akčního potenciálu s dlouhým trváním, a proto jsou cílem lokálních anestetik, jako je lidokain a benzokain. Argininové náboje na uzavíratelné bráně vytvářejí mnohonásobné hydrofilní interakce uvnitř napětového senzoru, včetně neočekávaných vodíkových vazeb k proteinové páteři. (Jiang et al., 2003) Napětově řízené Ca^{2+} kanály regulují intracelulární koncentrace Ca^{2+} , a jsou tak zodpovědné za širokou škálu biochemických procesů v buňkách. Jedním z nejdůležitějších procesů regulovaných těmito kanály je uvolňování neurotransmiterů na synapsích neuronů. Blokátory kalciových kanálů jsou cenné při léčbě různých stavů, od srdečních chorob po úzkostné poruchy. Napětově řízené K^+ kanály představují největší a nejrozmanitější třídu napětově řízených iontových

kanálů. Jsou nezbytné při vytváření klidového membránového potenciálu. Kanál obsahuje centrální iontový vodivý pór obklopený napěťovými senzory, které tvoří „lopatky napěťových senzorů“ hydrofobní, kationtové, struktury helix – turn – helix se nachází na vnějším obvodu kanálu. Flexibilní panty naznačují, že napěťové sensorové lopatky se pohybují v reakci na změny membránového napětí a přenášejí svůj kladný náboj přes membránu. (Jiang et al., 2003) Napěťově řízené Cl⁻ kanály jsou přítomny v každém typu neuronu a podílejí se na regulaci vzruchu a objemu buněk. Je také známo, že přispívají ke klidovému membránovému potenciálu.

2.1.3 Mechanosenzitivní kanály

Mechanosenzitivní kanály jsou ovládány napětím zprostředkovaným okolními lipidy v membráně. Napětí lipidové dvojvrstvy je distribuováno zcela nehomogenně (Gullingsrud & Schulten, 2004). Mechanosenzitivní (MS) kanály (Obr. 5), které poskytují ochranu před hypoosmotickým šokem, se nacházejí v membránách organismů ze tří oblastí života: bakterií, archaea a eukarya. Jsou rozpoznány dvě rodiny všudypřítomných MS kanálů a ty byly označeny jako rodiny MscL a MscS (Obr. 5). Pro členy rodiny MscL je k dispozici rentgenová krystalografická struktura s vysokým rozlišením a rozsáhlé molekulárně genetické, biofyzikální a biochemické studie provedené v mnoha laboratořích umožnily postulovat uzavíratelný mechanismus „bran“ umožňující vzájemnou konverzi těsně uzavřeného stavu a otevřeného stavu, který řídí transmembránové toky iontů a metabolitů. Na rozdíl od proteinů kanálu MscL, které mají jednotnou topologii, mnohem větší rodina MscS zahrnuje proteinové členy s topologiemi, u kterých se předpokládá, že se budou lišit od 3 do 11 α -helikálních transmembránových segmentů (TMS) na polypeptidový řetězec. (Pivetti et al., 2003)



Obr. 5: Struktura mechanosenzitivního kanálu. Krystal struktur MscS v jeho (A) uzavřené a (B) otevřené konformaci. Boční pohled ukazuje celkový protein, včetně intracelulární domény zanořené do membrány. Pohled z vrchu ukazuje pouze transmembránovou doménu, kdy zabírá větší prostor při otevřeném stavu než uzavřeném. Převzato z (Alberts et al., 2015).

2.1.4 Enzymové kanály

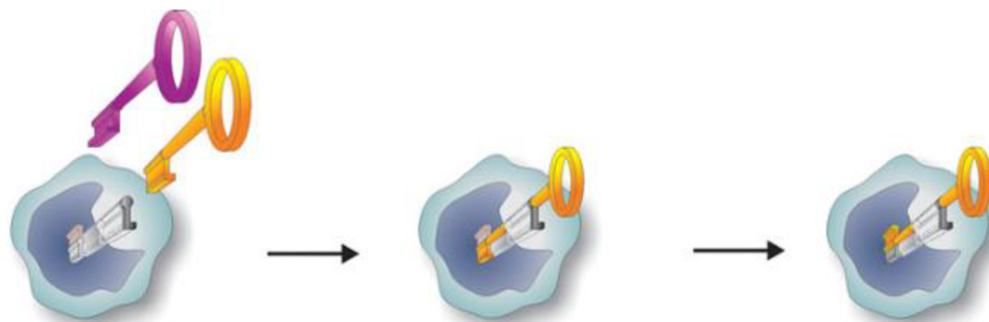
Enzymy jsou účinné a specifické katalyzátory pro mnoho základních funkcí v organismech. Jedná se o proteiny, které katalyzují chemickou reakci a mění substrát na produkt. Enzymatické reakce se objevují v enzymaticky aktivním místě. Díky mnoho analýzám enzymatických reakcí, došlo k lepšímu porozumění, jak fungují chemicky aktivní místa a jaké aminokyseliny jsou přítomny v těchto aktivních místech. Relativně málo prozkoumán je vstup substrátu do aktivního místa a jak jej příslušný produkt opouští. (Pravda et al., 2014)

Zatímco některé aktivní místa jsou umístěny na proteinovém povrchu, v trhlínách nebo kapsách, mnoho enzymů s aktivními místy schovanými ve struktuře musí být schopné přenášet substráty a produkty do a z okolí za účelem uplatnění jejich katalytické aktivity.

Tvar a velikost tunelů tak může významně přispívat k vazbě ligandu v proteinech se skrytými aktivními místy. Aminokyseliny ležící v okolí tunelu jsou důležité v selektivě těchto enzymů, flexibilitě kanálů jako jsou otevírací a zavírací pohyby, přispívá k široké substrátové specifitě (Pravda et al., 2014). Ve většině případů fungují tunely jako prodloužení vazebného místa enzymu a podílejí se na vstupu a výstupu ligandu, také na stabilizaci ligandu v jeho přirozeném vazebném režimu. Fyzický tvar a velikost tunelu nemusí určovat pouze maximální velikost odpovídajícího ligandu, ale může také stanovit počet potenciálních interakcí nebo bariér, na které může ligand narazit, když migruje směrem k aktivnímu místu. (Kingsley & Lill, 2015)

Byl navržen model „keyhole–lock–key“, který se konkrétně zabývá tím, jak tunely nebo „klíčové dírky“ v proteinech ovlivňují enzymatickou aktivitu, specifitu a stereoselektivitu (Prokop et al., 2012). V takovém modelu lze rozeznat substrát (klíč), který musí projít tunelem (klíčovou dírkou), aby dosáhl aktivního místa (zámku), kde se přemění na produkt. Díky takovému modelu lze říct, že tunely obsahující aktivní místo jsou důležitými strukturálními jednotkami, které by mohly přispět k regulaci enzymatických funkcí a dalších biologických procesů. (Marques et al., 2016) Jako takové lze považovat enzymové tunely za potenciální hotspoty pro modulaci fungování biomolekul uvnitř živých buněk. (Obr. 6)

Enzymové tunely často obsahují molekulární brány, které mohou specificky regulovat transport látek těmito biomolekulárními systémy. Brány fungují jako filtry v širokém spektru enzymů a řídí řadu substrátů (Gora et al., 2013). Molekulární otevírání a zavírání brány je dynamický proces, při kterém je jednotlivým látkám s malým rozměrem molekuly umožněn nebo odepřen přístup ke konkrétním místům v makromolekule. Různé procesy otevírání a zavírání brány mohou mít velmi specifické časové osy a role. Brána je umístěna směrem k cytoplazmě, bývá často jasně vidět v krystalové struktuře. (Batiza et al., 1999) Brány, které se nacházejí v mnoha enzymech, nebyly důkladně prozkoumány a nejsou tak dobře pochopeny. Molekulární brána je dynamický systém skládající se z jednotlivců nebo skupin zbytků, smyček, sekundárních strukturních prvků nebo dokonce domén, které reverzibilně přepínají mezi otevřenými a uzavřenými konformacemi, a tím řídí pohyb malých molekul – substrátů, produktů, iontů nebo rozpouštědla – do nebo z proteinové struktury. (Marques et al., 2017)



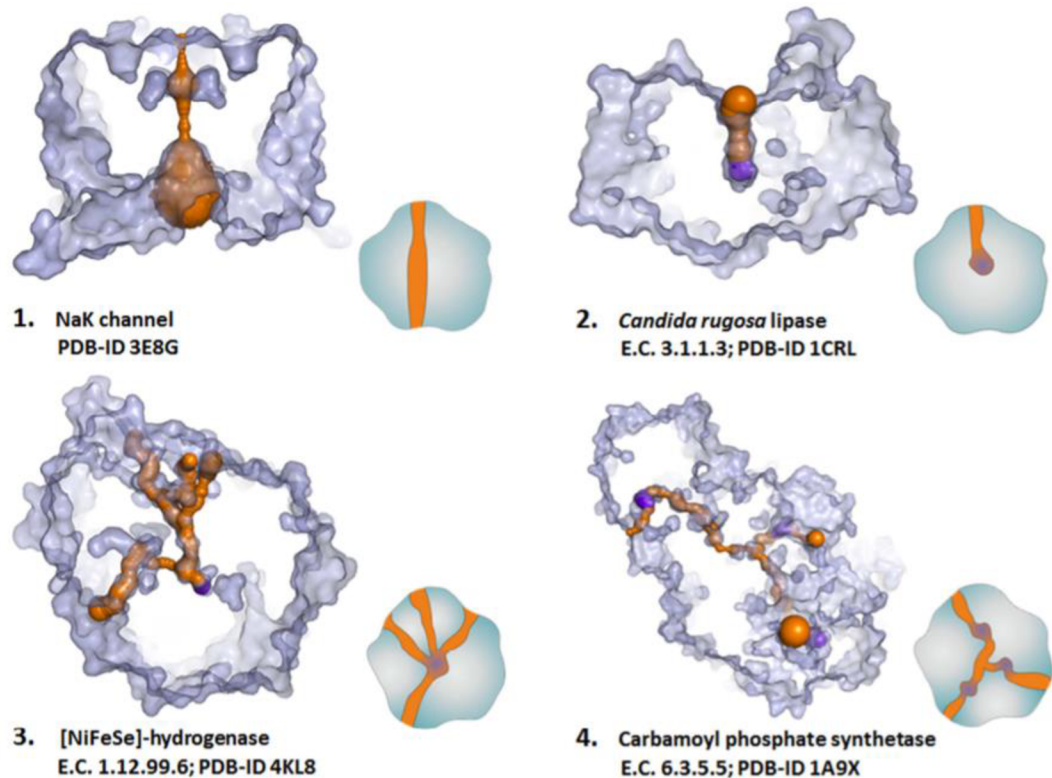
Obr. 6: Model „keyhole–lock–key“, kde klíč představuje substrát, který musí projít klíčovou dírkou ilustrující tunel, aby dosáhl zámku, tedy aktivního místa a reagoval. Převzato od (Marques et al., 2017).

2.1.4.1 Struktura enzymového kanálu

Ve většině případů fungují tunely jako rozšíření vazebného místa enzymu a jsou zapojeny do vstupu a výstupu ligandu, jakož i do stabilizace ligandu v jeho přirozeném vazebném režimu. Schovaná aktivní místa mohou umožnit velmi těsnou kontrolu nad katalytickým procesem na různých úrovních, ale vyžadují určité prostředky komunikace s hromadným rozpouštědlem, to znamená způsob přenosu substrátů, produktů, kofaktorů a molekul rozpouštědla mezi vnějším prostředím a aktivním místem (nebo mezi jedním aktivním místem a druhým). Toto je primární role enzymového tunelu.

Sekundární a související role tunelů je vybírání druhů ligandů, které mají povoleno podstoupit tento přenos. Například přítomnost molekul vody může bránit enzymaticky katalyzované reakci, a tak by enzym musel přísně kontrolovat jejich průchod. V jiných případech může být nutné zabránit uvolňování toxických meziproductů z vnitřku enzymu a zároveň je přenášet mezi různými aktivními místy. (Marques et al., 2017)

Existence tunelu obecně usnadňuje výběr, kterým látkám je povolen přístup ke konkrétním částem proteinu. Pokud jsou geometrie a fyzikálně chemické vlastnosti tunelu dobře uzpůsobeny, může být kanál schopen vyloučit všechny kromě požadovaných substrátů a zajistit tak substrátovou specifitu enzymu. Bylo zjištěno, že tunely obsahující skryté aktivní místo se vyskytují ve všech šesti hlavních klasifikačních třídách enzymů. (Prokop et al., 2012) Na základě strukturních prvků a molekulárních funkcí lze vymezit čtyři typy kanálů (Obr. 7):



Obr. 7: Příklady proteinů obsahujících pór (1.), jediný tunel (2.), více tunelů spojující aktivní místo (3.), nebo více aktivních míst s vnějším prostorem proteinu (4.). Kanály a tunely jsou znázorněny oranžovou barvou, aktivní místa fialovou. Převzato (Marques et al., 2016).

- 1) Póry fungují jako vysoce selektivní, obousměrný transportér pro obrovské množství proteinových či iontových nákladů, pro takový transport existuje většinou dobře regulovaný mechanismus (Wente & Rout, 2010). Nacházejí se například v iontových kanálech (Obr. 7), které umožňují průchod specifických iontů skrz membránové proteiny (K^+ kanál, Ca^{2+} kanál atd.), v iontových pumpách (Na^+/K^+ , adenosintrifosfatáza (EC 3.6.1.3), transportér neurotransmiterů atd.), nebo v porinech (Marques et al., 2016).
- 2) Jednotlivý tunel spojující skryté aktivní místo s rozpouštědlem (Obr. 7). Tunel je jedinou cestou pro výměnu činidel, produktů, rozpouštědel nebo iontů mezi skrytým aktivním místem a okolím proteinu (Marques et al., 2017). Mnoho enzymů má také jeden trvalý tunel a několik přechodných tunelů, které lze odhalit pouze studiem dynamiky proteinů (Marques et al., 2016). Jedním z příkladů takových enzymů je triacylglycerol acylhydroláza (lipáza EC 3.1.1.3) z *Candida rugosa*.

- 3) Spojení mezi aktivním místem enzymu skrytým uvnitř proteinového tunelu a rozpouštědlem může být zajištěno více tunely (Obr. 7) (Prokop et al., 2012). Substráty, produkty a rozpouštědlo mohou mít různé preference pro konkrétní tunely. Ve většině případů existuje jeden hlavní tunel a jeden nebo více vedlejších tunelů, které fungují jako alternativní nebo pomocné trasy.
- 4) Cytochrome-c3 hydrogenáza (G3DSA 1.10.645.10) je enzym, který obsahuje mnohočetné tunely. (Marques et al., 2017) Jedním z takových příkladů je cytochrom P450, který má hlavní 22 Å dlouhý hydrofobní tunel důležitý v přístupu k substrátu a výstupu produktu, zatímco 12 další sekundární tunelů umožňuje výměnu molekul kyslíku a rozpouštědla a také zajišťují alternativní cesty pro uvolnění produktu. Podobně mají halogenalkandehalogenázy (HLDs, EC 3.8.1.5) hlavní tunel, používaný pro výměnu halogenovaného substrátu, alkoholu a halogenidových produktů a několik sekundárních tunelů se používá pro uvolňování alkoholu a výměnu vodního rozpouštědla. (Klvaňa et al., 2009)
- 5) Tunely spojující různá katalytická místa v multifunkčních enzymech nebo enzymových komplexech s více aktivními centry (Obr. 7). Představují dvě oddělená aktivní místa vzájemně spojená molekulárním tunelem. (Raushel et al., 2003) Tunely mohou zvýšit katalytickou účinnost enzymu, snížením doby průchodu z a do vazebných míst. Navíc, průchod substrátu tunelem může sloužit k účelu těsné regulace řady po sobě jdoucích reakcí, které mohou být součástí metabolické dráhy nebo vícestupňového katalytického cyklu (Prokop et al., 2012). Mohou také pomoci předcházet vedlejším reakcím tím, že udržují nestabilní meziprodukty mimo rozpouštědlo, nebo dokonce zabraňují uvolňování toxických produktů do intracelulárního prostředí. Takový jev je běžně známý jako „kanálování substrátu molekulárními tunely“. (Miles et al., 1999) Nemělo by se zaměřovat s elektrostatickým kanálem, který se týká řízení meziproduktů mezi různými katalytickými místy poháněnými elektrostatickými poli a nemusí nutně zahrnovat enzymové tunely. Dobrým příkladem je karbamoylfosfátsyntáza (EC 6.3.4.16) multifunkčního enzymu s několika tunely spojujícími různá aktivní místa. (Marques et al., 2017)

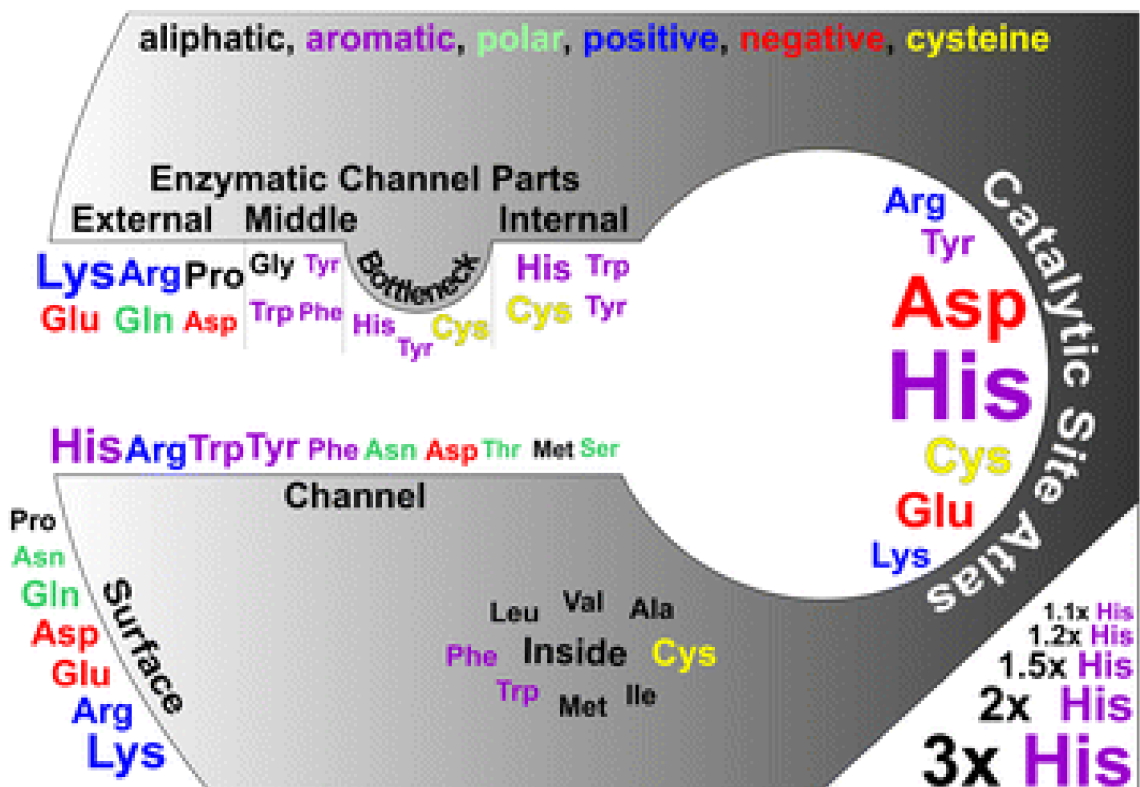
2.1.4.2 Residua v okolí enzymového kanálu

Frekvence aminokyselin v aktivních místech, na povrchu proteinu a uvnitř proteinu, nebo v kanálech a jejich aminokyselinové složení je rozdílné. Výsledné vlastnosti tunelu, se výrazně liší podle aminokyselin v jeho okolí. Aminokyseliny zaujímající větší prostor (s větším postranním zbytkem) a aromatické aminokyseliny se vyskytují 1,25 krát častěji v tunelech s délkou 15 Å a větší než jinde v celém enzymu. Jiné aminokyseliny (Asp, Phe, Asn, Thr, Met, Ser) se také vyskytují v trochu větší frekvenci v kanálech než ve zbytku proteinu. Naopak nepolární, alifatické aminokyseliny (Pro, Gly, Ile, Leu, Ala a Val) se výrazně méně vyskytují ve stěnách kanálu. (Pravda et al., 2014) Liší se od typických katalytických aminokyselinových zbytků.

Enzymy také využívají širokou škálu chemikálií, včetně mnoha různých kovových iontů, kovové kofaktory, kovové/nekovové komplexy a posttranslační modifikace 20 přirozeně vyskytujících se aminokyselinových zbytků (Holliday et al., 2009).

U tunelů přijímající plynné substráty nacházející se v několika enzymech se vyskytují modifikované aminokyseliny. Třída převládajících plynných substrátů jsou alkany, jako je methan a ethan, které bývají oxidované. Na konci tunelu, v blízkosti aktivního místa je několik zbytků, které jsou posttranslačně modifikované. Methylované aminokyseliny, kterými jsou S-methylcystein, 3-methylisoleucin, 2(S)-methylglutamin a N2-methylhistidin lemují tunel. Je pravděpodobné, že tyto zbytky pomohou selekci ethanu vytvořením velmi hydrofobního prostředí a zabránit hydrofilním molekulám podobné velikosti od dosažení aktivního centra. (Singh & Anand, 2021)

Hrdlo kanálu obsahuje výrazně více cysteinu, histidinu a tyrosinu a mnohem méně alifatické aminokyseliny (Pro, Gly, Ala) (Obr. 8). Histidin a cystein mají unikátní vlastnosti vazby, díky tomu jde předpovědět, že se tyto vazebné vlastnosti mohou podílet na kontrole hrdla kanálu. Zatímco malé, alifatické residua nemůžou vydržet velké změny a nemohou sloužit ke kontrole hrdla kanálu. Aktivní místo obsahuje výrazně více aminokyselin, které mohou být součástí katalytického cyklu (His, Asp, Cys, Glu, Arg, Tyr, Lys) (Obr. 9) a poskytují proton a elektron, následně dochází k reorganizaci kovalentní vazby. Frekvence méně reaktivních aminokyselin (Trp, Thr, Gln, Phe) nebo aminokyselin s nereaktivním postranním řetězcem (Met, Ala, Pro, Ile, Val, Leu) je v aktivních místech menší. (Pravda et al., 2014)



Obr. 8: Zastoupení frekvence aminokyselin v různých částech struktury enzymu. Znárodnění částěji frekventovaných aminokyselin se odráží ve velikosti písma, kterou jsou napsány (legenda vpravo dole). Aminokyseliny jsou znázorněny odlišnými barvami podle jejich vlastností. Převzato (Pravda et al., 2014).

2.1.4.3 Selekcce látek enzymatickými tunely

Vliv „uzavíratelných bran“ na substrátovou specifitu lze připsat lokalizovanému znaku nebo souboru znaků, ale při určování substrátové specifity může hrát roli také celková dynamika enzymu (Kingsley & Lill, 2015). Úzké hrdlo tunelu – jeho nejužší místo, je často ohniskem selektivity, protože určuje maximální velikost látek, které mohou projít tunelem. Další důležitou součástí tunelu je jeho vchod nebo ústí. To má za následek prvotní interakci s objemovým rozpouštědlem a může hrát zásadní roli při rozpoznávání substrátu. Podobně skupina zbytků tvořící úzké hrdlo na vstupu do tunelu může hrát hlavní roli při určování jeho funkce. (Alberts et al., 2015)

Pro pochopení funkce tunelu je důležité znát jeho rozměry, jako je délka, průměr ve všech místech a zakřivení. Důležitou roli hrají také aminokyseliny vyskytující se v okolí tunelu a jejich postranní zbytky. Jako jedny z nejselektivnějších míst se považují hrdlo a vstup tunelu, aminokyseliny v okolí a velikost tunelu současně zaručují jeho specifické

vlastnosti. Zatímco hlavní tunely jsou většinou trvalé a lze je snadno identifikovat zkoumáním krystalických struktur, další přechodné tunely lze identifikovat pouze studiem dynamických změn v rámci biomolekuly.

Přechodné tunely přepínají mezi otevřeným a uzavřeným stavem v závislosti na konformaci proteinu. Navzdory své pomíjivosti mohou být nezbytné pro správnou funkci enzymu. Mnoho enzymů, o kterých se v současné době předpokládá, že mají pouze jeden tunel, mohou ve skutečnosti mít jiné funkční přechodné tunely, které lze identifikovat pouze zkoumáním dynamiky proteinů. (Sanson et al., 2011) V některých případech lze vlastnosti tunelů správně charakterizovat pouze tehdy, když jsou studovány v přítomnosti ligandů, které mohou vyvolat jejich otevření a ovlivnit jejich tvorbu a perzistenci (Marques et al., 2017).

Mutace v místě aktivního místa, vstupu, nebo hrdla kanálu může měnit jeho geometrii, fyzikálně chemické vlastnosti a tím i jeho selektivitu vstupujících molekul. S využitím genetického inženýrství mohou být vytvořeny modifikované a někdy účinnější enzymy pro požadované účely. Vytvoření modifikovaných enzymů však závisí na několika experimentech *in vitro*, což může vést k tomu, že proces bude drahý a časově náročný. Proto by výpočetní přístupy mohly snížit náklady a urychlit objevování nových technologických produktů. (Mariano et al., 2019)

Dlouhodobý výzkum v oblasti proteinové analýzy poukazuje na to, že reaktivita struktury proteinu závisí na přítomnosti tunelů. Tyto struktury jsou velmi důležité hlavně v procesu hledání nových léčiv. Vizualizace tunelu je dalším velmi důležitým krokem analýzy, protože umožňuje biochemikům určit klíčové oblasti tunelu, které mohou mít podstatný vliv na tunel. (Kozlíková et al., 2007) Hlavním předpokladem pro predikci vzájemných se látek je, že podobná vazebná místa pro proteiny vážou podobné ligandy, a tak známý ligand z jednoho proteinu může být transponován na podobné vazebné místo v jiném proteinu, o kterém dříve nebylo známo, že váže tento ligand. Transpozice ligandů je založena na přesném zarovnání trojrozměrných aminokyselinových vzorů nebo jejich odpovídajících funkčních skupin ve vazebných místech proteinů. Kvůli jejich lokální povaze ve vazebných místech nemusí být taková srovnání možná se standardními přístupy pro srovnání sekvencí nebo struktur. (Konc et al., 2015)

2.1.4.4 Typy vazby ligandu na tunel

Ligand, který se váže na tunel nebo bránu v enzymu a tím inhibuje nebo modifikuje aktivitu enzymu, tak může učinit interakcí následovně podle (Marques et al., 2017):

- 1) Ligand se váže přímo na katalytické místo. Tento typ vazby je mezi známými enzymovými inhibitory nejčastěji pozorovaným typem.
- 2) Ligand se může vázat na tunel v proteinu a blokovat jeho hlavní funkci, sloužící jako transportní cesta. Takové ligandy se mohou rozprostírat po celé délce tunelu a vytvářet mnoho interakcí se zbytky obložení tunelu. To může vést k tvorbě velmi stabilních komplexů kvůli velkému počtu interakcí mezi ligandem a tunelem, což z ligandu dělá velmi silný inhibitor. Je běžnější, že se ligand váže na specifickou oblast tunelu, kde vytváří silné stabilizační interakce s několika zbytky, než aby se rozprostíral po celé jeho délce. Zbytky úzkého hrdla mohou být méně konzervované než zbytky v aktivním místě, co nabízí příležitosti k vývoji vysoce specifických a selektivních inhibitorů. (Marques et al., 2017)
- 3) Pokud se ligandy vážou na bránu enzymu, nejběžnější vazebná poloha je v bodě kontaktu mezi pohybujícími se prvky. Ligandy tohoto druhu mohou cílit buď na otevřenou nebo uzavřenou konformaci brány, stabilizovat a uzamknout konformaci, na kterou se vážou, a tím narušit vtokový mechanismus a biologickou funkci cíle. (Marques et al., 2017)
- 4) Ligandy, které interagují s tunely a branami, se mohou vázat i více způsoby. Například může dojít k vazbě jak na katalytické místo, tak na některé zbytky tvořící tunel.

2.1.4.5 Průzkum enzymatických tunelů v proteinech

Na dostupných 4306 strukturách enzymů z Catalytic Site Atlas (<http://www.ebi.ac.uk/thornton-srv/databases/CSA>). Bylo zjištěno, že více než 64 % enzymů, má aktivní místo dostupné tunelem delším než 15 Å. (Pravda et al., 2014) Tunely alespoň 5 Å dlouhé obsahovalo 86,6 % zkoumaných enzymů. Dalo by se očekávat, že větší proteiny obsahují větší počet kanálů, ale ukázalo se, že počet delších kanálů nekoreluje s velikostí proteinu. Některé proteiny mají více než 5 kanálů a největší zjištěný počet je 68. Velikost enzymů neomezuje délku, kterou kanály mohou mít. Nejdelší kanál měl délku 172 Å (*Pennicillium janthinellum*). Model průměrného enzymového kanálu

obsahoval především aminokyseliny histidin, arginin, tryptofan, tyrosin a v menší míře fenylalanin, asparagin a kyselinu asparagovou. Průměrná délka byla okolo 27,7 Å. Průměrná hydropatie kanálů je -0,92 tedy cca hodnotou odpovídala tyrosinu a tryptofanu (hydropatie aminokyselin se pohybuje od -4,5 pro arginin až po 4,5 u isoleucinu). Hydrofilní kanály se objevovaly častěji než kanály hydrofatické. Průměrná polarita kanálu byla 16,5, přičemž tato hodnota spadá mezi hodnoty vysoce polárních aminokyselin jako jsou asparagin, glutamin, lysin, arginin a histidin. Ukázalo se také, že kanály jednotlivých enzymatických tříd se mezi sebou liší. Největší podíl struktur s kanály byl identifikován u oxidoreduktáz (77,8 %), oproti tomu nejméně takových kanálů obsahuje enzymatická skupina hydrolázy (51,8 %). Počet kanálů dané délky je mírně zvýšen u oxidoreduktás, transferás a izomerás. Oxidoreduktázy mají medián délky kanálu okolo 2 Å, ostatní enzymy, jako transferázy, ligázy mají průměrnou délku kanálu kratší. Oxidoreduktázy nejen, že mají nejdelší kanály z enzymatických tříd, ale také největší výskyt tunelů v proteinech. (Pravda et al., 2014)

3 EXPERIMENTÁLNÍ ČÁST

3.1 MOLEonline

Jedná se o volně dostupné webové rozhraní (<https://mole.upol.cz>) sloužící k analýze kanálů, tunelů a pórů ve strukturách proteinů, algoritmus je napsaný v programovacím jazyce Java Script. Aplikace umožňuje dva základní módy, jeden z nich je věnován analýze kanálů, zatímco druhý byl speciálně navržen pro transmembránové póry. Výsledky jsou prezentovány jasným způsobem, což usnadňuje jejich interpretaci. Pro každý kanál MOLEonline zobrazí 3D grafické znázornění kanálu, jeho profil doplněný seznamem zbytků v okolí tunelu i jeho základní fyzikálně-chemické vlastnosti. (Berka et al., 2012)

MOLEonline přijímá vstupní soubory ve formátu PDB i mmCIF, nebo umožňuje vybrat biologicky aktivní jednotku z PDBe, následkem toho lze v aplikaci zkoumat široké spektrum biomakromolekul. Výsledky zkoumaných struktur pochází z technik NMR, X - ray, nebo kryo - EM. MOLEonline je propojen s dalšími bioinformatickými nástroji, jako jsou PDBe, CSA, ChannelsDB, OPM a UniProt, pro zlepšení výsledků analýzy. MOLEonline umožňuje analýzu pro detekci a charakterizaci struktury kanálů, ale zároveň také číselné informace o fyzikálně chemických vlastnostech. V nynější verzi umožňuje MOLEonline plnou a interaktivní funkci pro analýzu tunelů a transmembránových pórů. Navíc umožňuje počítání geometrie detekovaných kanálů, také poskytuje informaci o aminokyselinových residuích ležících na povrchu tunelu. MOLEonline je propojen s databází ChannelsDB pro usnadnění porovnání strukturních vlastností nalezených kanálů s těmi uloženými v databázi (Pravda et al., 2018 a) (<https://webchemdev.ncbr.muni.cz/ChannelsDB/>).

MOLE lze také používat jako samostatnou verzi pro příkazový řádek, také ve snadno použitelném grafickém prostředí, a to buď prostřednictvím pluginu pro široce používaný software PyMOL, nebo online MOLEonline verzi, která kombinuje algoritmus MOLE s grafikou Litemol (<https://www.litemol.org/>). Software MOLE také obsahuje plně automatický algoritmus shlukování usnadňující analýzu lokalizovaných tunelů. (Petrek et al., 2007)

3.1.2 Algoritmus MOLE

Uživatel může využít defaultní nastavení parametrů MOLE, nebo je modifikovat, dle vlastního uvážení. Parametry dutin (*Cavity parameters*), *Probe radius* a *Interior threshold* definují α -tvar povrchu biomakromolekuly a malé β -tvary uvnitř nich. Parametry kanálů *Origin radius* a *Surface cover radius* mohou být optimalizovány pozicí počátečních bodů, aby uzavřel čtyřstěn nebo koncový bod na povrchu. *Pattern query* (dotaz) může být použit nejen pro definici počátečních a koncových povrchů, ale také pro vyloučení specifických skupin nebo residuí, které mohou zablokovat nebo schovat kanál (Pravda et al., 2018 a).

MOLEonline algoritmus obsahuje několik kroků (Sehnal et al., 2013):

- 1) Vypočítání Delaunayovy triangulace / Voronného diagram pro střed atomů
- 2) Konstrukce molekulárního povrchu
- 3) Identifikace dutin (*cavities*)
- 4) Identifikace možných počátečních bodů kanálů
- 5) Identifikace možných koncových bodů kanálů
- 6) Umístění tunelu
- 7) Filtrování určených kanálů

Výpočet kanálu je dokončen po definici počátečních a koncových bodů pro všechny dutiny, Dijkstrův algoritmus hledající nejkratší cesty se používá k nalezení kanálů mezi všemi dvojicemi počátečních a koncových bodů (nebo molekulárním povrchem, pokud ho algoritmus dosáhne dříve, než najde výstup z kanálu). Funkce váhy hrany použitá v algoritmu bere v úvahu vzdálenost k povrchu nejbližší Van der Waalovy sféry a délku hrany. Středová čára kanálu je reprezentována jako 3D přirozená spline reprezentace definovaná vrcholy Voronoiova diagramu, které tvoří cestu nalezenou pomocí Dijkstrova algoritmu. Algoritmus může najít duplicitní kanály. Proto je v posledním kroku zařazena filtrace detekovaných kanálů (<https://mole.upol.cz/method>).

3.1.3 Vypočítané fyzikálně-chemické vlastnosti tunelu

Pokud je tunel vypočítán, soubor aminokyselin (postranních řetězců) lemujících kanál může být použit pro výpočet vlastností (dále popsané podle <https://mole.upol.cz/method>). Vzhledem ke složení postranních řetězců se vypočítá soubor fyzikálně-chemických vlastností, kde se berou v potaz pouze aminokyseliny obložení daného tunelu.

Některé aminokyseliny mohou být (de)protonovány na základě pH, a tudíž nabité. Při fyziologickém pH jsou lysin a arginin nabitý kladně, naproti tomu kyselina asparagová a glutamová jsou nabitý záporně. Na druhé straně je protonační stav histidinu závislý na jeho mikroprostředí. V této studii jsou všechny histidiny považovány za pozitivně nabité. Vlastnosti náboje jsou součtem všech kladně a záporně nabitých aminokyselin. (<https://webchemdev.ncbr.muni.cz/ChannelsDB/methods.html>) *Náboj* se počítá jako součet náboje postranních zbytků. (<https://mole.upol.cz/method>)

Výpočet *Hydropatie* probíhá spočítáním průměru indexu hydropatie přiřazeného zbytkům podle metody Kyte a Doolittle (index hydropatie je spojen s hydrofilitou/hydrofobicitou aminokyselin). *Hydrofobicita* a *hydrofilita* jsou dva extrémní spektra, běžně označované jako hydropatie a vztahují se k tendenci molekuly interagovat s vodou. Existuje několik stupnic hydropatie, aby bylo možné popsat celkový charakter proteinů nebo jejich částí. Kyte-Doolittlova stupnice (Kyte & Doolittle, 1982) je široce používaným měřítkem pro vyjádření hydrofobnosti aminokyselinových zbytků. Hydrofobní oblast se pohybuje nad 0. Stupnice je symetrická v intervalu od -4,5 (Arg) do 4,5 (Ile). (<https://webchemdev.ncbr.muni.cz/ChannelsDB/methods.html>) *Hydrofobnost* se vypočítá jako průměr normalizovaných stupnic hydrofobicity podle (Cid et al., 1992). Nejvíce hydrofilní aminokyselina podle hodnoty hydrofobnosti je Glu (-1,140) a nejvíce hydrofobní je Ile (1,810). (<https://mole.upol.cz/method>)

Polarita je vlastnost molekuly daná oddělením rozložení elektrického náboje, což vede k tomu, že molekula má elektrické póly. Obecně řečeno, polární molekuly jsou hydrofilní, zatímco nepolární molekuly jsou obvykle hydrofobní, ale mohou existovat výjimky. Z hlediska aminokyselinových zbytků lze hydrofilní/polární a hydrofobní/nepolární považovat za synonyma. Stupnice se pohybuje od 0 pro malé alifatické aminokyseliny (Ala, Gly) do 51,6 (His). (<https://webchemdev.ncbr.muni.cz/ChannelsDB/methods.html>) *Polarita* se vypočítá jako průměr polarit aminokyselin přiřazených podle metody.

(Zimmerman et al., 1968) Polarita sahá od zcela nepolárních aminokyselin (Ala, Gly = 0,00) přes polární zbytky (např. Ser = 1,67) až po nabitě zbytky (Glu = 49,90, Arg = 52,00).

Lipofilita (škála *LogP*) se vypočítá jako rozdělovací koeficienty oktanol/voda fragmentů postranních řetězců a hlavního řetězce (-0,86) prostřednictvím (www.chemicalize.org).

Hodnota *lipofility LogD* se vypočítá jako distribuční koeficienty oktanol/voda fragmentů postranních řetězců a hlavního řetězce (-0,86) při pH 7,4 prostřednictvím (www.chemicalize.org). Distribuční koeficient zohledňuje ionizaci sloučenin

Škála *rozpuštnosti LogS* se počítá jako rozpustnost fragmentů postranních řetězců a hlavního řetězce ve vodě (0,81) při pH 7,4 za použití (www.chemicalize.org). Odhadovaná hodnota *LogS* je jednotkový dekadický logaritmus rozpustnosti měřené v mol/litr, představuje míru, jak dobře mohou jednotlivé zbytky reagovat s molekulami vody. Nejnížší hodnoty *LogS* se vyskytují u aromatických aminokyselin, následovaných alifatickými a aminokyseliny obsahující síru až po nejvyšší rozpustnost u polárních aminokyselin. Většina molekul vody bude přitahovat nabitě zbytky. Glycin nemá žádný postranní řetězec, proto je jeho hodnota přednastavena na 0. (<https://mole.upol.cz/method>)

Mutabilita se počítá jako průměr relativního indexu mutability (Jones et al., 1992). Výpočet relativní mutability je založen na tvorbě empirických substitučních matic mezi podobnými proteinovými sekvencemi. (<https://mole.upol.cz/method>) Každá z aminokyselin může být zmutována, avšak některé z aminokyselin mohou být mutovány bezpečněji vzhledem ke struktuře, stabilitě a skládání proteinu. Index relativní mutability odráží tyto vlastnosti. Zároveň je založen na analýze proteinových sekvencí (podobně jako dobře známé matrice BLOSUM a PAM používané při zarovnání sekvencí). (<https://mole.upol.cz/FAQ>)

3.1.4 Vizualizace a analýza výsledků v MOLEonline

Vypočítané kanály jsou tříděny podle jejich typů na tunely, póry a cesty (spojující dva vybrané body uživatelem) a následně je kategorizují podle jejich délky. Kanály pocházející z jedné dutiny (*cavity*) jsou přiděleny do jednoho setu. Kanály mohou být vizualizovány v LiteMol Viewer a při výběru jsou jejich vlastnosti vizualizovány v levém spodním rohu webové stránky. Dva nezávislé grafy ukazují profil kanálu ve 2D (poloměr, vzdálenost nebo délku). (Pravda et al., 2018 a)

2D mapa kanálu nabízí znázornění ve třech variantách a to *Radius* (označuje poloměr koule, která je omezená třemi nejbližšími atomy), *Free Radius* (představuje poloměr koule omezený třemi nejbližšími atomy, aby byla možná flexibilita postranních řetězců) a *BRadius* (poloměr + RMSF, vypočtený z B-faktorů residuí v jednotlivých vrstvách $((3*B)/(8*\pi*\pi))^{-0.5}$) (<https://mole.upol.cz/method>) Každý profil (vrstva) může být obarven, podle fyzikálně chemických vlastností aminokyselin v okolí tunelu. Uživatel může vybírat z *Charge*, *Hydropathy*, *Hydrophobicity*, *Polarity*, *Lipophilicity* (*LogP*, *LogD*), *Solubility* (*LogS*) nebo *Mutability*. (Pravda et al., 2018 a)

Aby byla zajištěna bezpečnost, výsledky každého výpočtu jsou uloženy jako samostatné a odeslané se všemi ostatními parametry v rámci hashovaného ID úkolu. Individuální historie hledání může být procházena v pravém horním rohu. Uživatel také může ukládat své výsledky do databáze ChannelsDB, pokud je k dispozici (<https://webchemdev.ncbr.muni.cz/ChannelsDB/>). Je-li k dispozici anotace residua získaná z UniProt API (https://www.uniprot.org/help/programmatic_access), tak je také ukázaná. Uživatel může poslat svoje výsledky, jako vstup anotace kanálu, do ChannelsDB používající jednoduchý formulář dostupný na stránce MOLEonline. Veškeré výsledky mohou být stáhnuty v různých formátech, buď jako makromolekula samotná v mmCIF formátu, nebo souřadnice tunelu v PDB souboru, nebo kombinace obou v PyMol, VMD a Chimera vizualizační skripty, nebo jako JSON nebo ZIP soubory. ZIP soubor obsahuje všechny výsledky odlišných formátů pro další analýzy a vizualizaci (PyMol, Chimera, VMD). Zpráva o výsledcích může být také stáhnuta také jako PDF soubor. (Pravda et al., 2018 a)

3.2 Databáze tunelů ChannelsDB

ChannelsDB je komplexní a pravidelně aktualizovaný zdroj kanálů, pórů a tunelů nalezených v biomakromolekulách uložených v Protein Data Bank (<https://www.wwpdb.org/>). Kanály byly detekovány softwarem MOLE, založeným na informacích z literatury, databáze obsahuje pozice schované katalytické sítě, kofaktory, nebo pozice transmembránových proteinů v membráně. Jako taková je to jedinečná služba pro analýzy související s kanálem. Databáze obsahuje informace o plochách kanálů, geometrii a fyzikálně-chemických vlastnostech. Všechny záznamy jsou navíc propojeny s databází UniProt (<https://www.uniprot.org/>), která je komplexním vysoce kvalitním zdrojem informací o funkci proteinů. V neposlední řadě jsou všechny výsledky zobrazeny přehledným interaktivním způsobem, co dále usnadňuje interpretaci dat. (<https://webchemdev.ncbr.muni.cz/ChannelsDB/>)

Individuální PDB identifikátor záznamu je použit jako klíč pro biologickou strukturu, pod kterou jsou informace o kanálech uloženy, zatímco UniProt mapuje unikátní informaci o studovaném proteinu. Anotace kanálů v ChannelsDB je získávána následujícím postupem.

- 1) Stovky záznamů byly procházeny manuálně, s ohledem na dostupné informace ve vědecké literatuře a následně mapovány přes předem vypočítané kanály v MOLE.
- 2) Bylo detekováno zhruba 15 000 záznamů o enzymech v tunelech, spojujících zakrytou katalytickou síť anotovanou v Catalytic Site Atlas (<http://www.ebi.ac.uk/thornton-srv/databases/CSA>) s povrchem proteinu.
- 3) Bylo identifikováno asi 12 000 proteinových záznamů z MOLEonline, které představují tunely spojující vybrané kofaktory s proteinovým povrchem.
- 4) Stovky záznamů byly predikovány v MOLEonline jako transmembránové proteiny.

ChannelsDB domovská stránka umožňuje dotazování pro struktury s anotací kanálů používající PDB příbuzná metadata, jako PDB ID, jméno proteinu, proteinová rodina, kofaktory a jiné ligandy, autory nebo žurnály. (Pravda et al., 2018 b)

3.3 Python

Python patří mezi oblíbené programovací jazyky, byl vytvořen v roce 1991 Guidem van Rossumem. Jeho osvojení je snadnější než u jiných programovacích jazyků, a to hlavně díky tomu, že je velmi podobný přirozenému jazyku s matematickými prvky. Kód napsaný v Pythonu je tedy srozumitelný pro čtení i zápis. Python je interpretovaný, objektově orientovaný, vysokoúrovňový programovací jazyk s dynamickou sémantikou. Díky vestavěným datovým strukturám na vysoké úrovni v kombinaci s dynamickým psaním a dynamickou vazbou je velmi atraktivní pro rychlý vývoj aplikací a také pro použití jako skriptovací nebo spojovací jazyk pro spojení existujících komponent dohromady. (<https://www.python.org/doc/essays/blurb/>)

V dnešní době je nejpoužívanější, nejnovější verze programovacího jazyka a to Python 3, ale stále se používá i starší verze Python 2. Python 3 se dodává s pomocným skriptem nazvaným `2to3`, který jako vstup přebírá zdrojový kód napsaný pro Python 2 a automaticky převádí vše, co dovede, do podoby pro Python 3. (Pilgrim, 2010) K definování cyklů, funkcí a tříd se spoléhá na odsazení oproti jiným programovacím jazykům které používají například složené závorky.

Python je multiplatformní jazyk. Obecně lze říci, že program napsaný v jazyku Python lze spustit ve Windows i v unixových systémech, jako je Linux, BSD a Mac OS X, pouhým zkopírováním souboru či souborů, které tvoří daný program, na cílový stroj, aniž by je bylo nutné „sestavovat“ nebo kompilovat pro dotyčnou hardware a software konfiguraci. Je možné vytvářet programy napsané v Pythonu, které používají funkčnost specifickou pro určitou platformu. To je ale jen zřídka nezbytné, protože téměř celá standardní knihovna Pythonu a většina knihoven třetích stran jsou plně a transparentně multiplatformní. (Summerfield, 2013)

3.3.1 Metody, knihovny, moduly

Knihovna v programování představuje předkompilované kódy, které lze použít při programování vlastního kódu pro specifické operace. Kromě předkompilovaných kódů může knihovna obsahovat dokumentaci, konfigurační data, šablony zpráv, třídy a hodnoty. Knihovna Pythonu je sbírka souvisejících modulů. Obsahuje více kódů, které lze opakovaně používat ve vlastních programech k určené činnosti. Usnadňuje tím práci programátora, protože nepotřebuje psát stejný kód znovu pro různé programy. Knihovny

Pythonu hrají velmi důležitou roli např. v oblastech strojového učení, datové vědy, vizualizace dat. (<https://www.geeksforgeeks.org/libraries-in-python/>)

Standardní knihovna Pythonu je velice rozsáhlá a umožňuje širokou škálu zařízení. Knihovny obsahuje vestavěné moduly (některé napsané v programovacím jazyce C), které poskytují přístup k funkcím systému, jako jsou souborové I/O, které by jinak byly pro programátory Pythonu nedostupné. Také moduly napsané v Pythonu, které poskytují standardizovaná řešení mnoha problémů se používají v každodenním programování. Některé z těchto modulů jsou výslovně navrženy tak, aby podporovaly a zlepšovaly přenositelnost programů Python tím, že abstrahují specifika platformy do platformou neutrálních API. (<https://docs.python.org/3/library/>)

Mezi knihovny používané v diplomové práci patří knihovny Pandas, Matplotlib, Arcade, NumPy, Math, os, glob, itertools, operator, cmd, zipfile, PyQt5, sys, selenium, time, win32api, cmd, PyQt5, PIL, pathlib, xml.dom, multiprocessing.

3.3.2 Knihovna Pandas

Pandas je open source balíček Pythonu, který se nejčastěji používá pro analýzu dat a úlohy strojového učení. Je postaven na jiném balíčku s názvem NumPy, který poskytuje podporu pro vícerozměrná pole. Mezi jeho výhody patří intuitivní manipulace s daty pomocí minimálních příkazů, optimalizovaná rychlost a dosahuje velmi flexibilního spojení s jinými knihovnamí Pythonu.

Pandas slouží pro analýzu dat, které lze reprezentovat 2D tabulkou. Takto reprezentovaná data mohou představovat informace v SQL databázích, souborech CSV nebo tabulkových procesorech. Knihovna Pandas umožňuje funkce jako má Excel, zároveň má širší možnosti funkcí, a hlavně umožňuje analýzu automatizovat.

Základní datový typ, který Pandas nabízí, se nazývá *DataFrame* a představuje tabulku. Informace o jednotlivých položkách představují řádky, sloupce pak vlastnosti dané položky v řádku. Nejčastější způsob, jak vložit informace do *DataFrame*, je načtení ze souboru. Pandas obsahuje sadu funkcí, která umožňuje toto načítání, příkaz začíná `pd.read_`, některé z nich potřebují další knihovny, viz dokumentace. Datové typy (`dtypes`) se doplňují automaticky podle zadaných hodnot (ze souboru). Pandas je používá hlavně pro šetření paměti, objekt Pythonu typu `bool` zabírá v paměti desítky bytů, ale v `bool` sloupci si každá hodnota vystačí s jedním bytem. Na rozdíl od NumPy jsou typy dynamické, pokud se do sloupce zapíše „nekompatibilní“ hodnota, kterou

Pandas neumí převést na daný typ, typ sloupce se automaticky zobecní. Některé automatické převody ovšem nemusí být úplně intuitivní (např. převod `None` na `NaN`).

Sloupec (*Series*) je dalším ze základních datových typů v Pandas. Obsahuje sérii hodnot, jako seznam, ale navíc má jméno, datový typ a index, který jednotlivé hodnoty pojmenovává. Sloupce se dají získat selekcí z tabulky. S informacemi ve sloupcích lze počítat samostatně. Základní aritmetické operace, jako sčítání či dělení sloupcem a skalární hodnotou (číslem, řetězcem, ...) provedou danou operaci nad každou hodnotou ve sloupci. Výsledek je nový sloupec.

Protože Python neumožňuje předefinovat chování operátorů `and` a `or`, logické spojení operací se tradičně dělá přes bitové operátory `&` (a) a `|` (nebo). Ty mají ale neintuitivní prioritu, proto se jednotlivé výrazy hodí uzavřít do závorek

Sloupce mají zabudovanou celou řadu operací, od základních (např. `column.sum()`, která bývá rychlejší než vestavěná funkce Pythonu `sum()`) až po statistické metody. Povědomí o operacích, které sloupce umožňují, je základní znalost datového analytika. (<https://naucse.python.cz/lessons/intro/pandas/>)

3.3.3 Python Arcade

Knihovna Arcade jazyku Python slouží ke grafické vizualizaci. Jedná se objektově orientovanou knihovnu. Knihovnu Arcade napsal Paul Vincent Craven, profesor počítačových věd na Simpson College v Iowě, USA. Je postavena na Pyglet knihovně oken a multimédií, avšak Arcade obsahuje různá vylepšení a modernizace.

Arcade kód běží téměř na každé platformě, která podporuje Python. To vyžaduje Arcade vypořádat se s abstrakcemi pro různé hardwarové rozdíly na těchto platformách. Vzhledem k tomu, že pracuje s různými platformami, je nutné provést inicializační krok před použitím. Tento krok je automatický a dochází k němu při každém importu Arcade.

Vše v Arcade se děje v okně, které se vytváří pomocí příkazu `window()` nebo `open_window()`. V současné době Arcade podporuje pouze jedno okno displeje. Před otevřením okna lze měnit velikost. Arcade používá kartézský souřadnicový systém. Okno leží v kvadrantu I s počátečním bodem (0, 0) umístěným v levém dolním rohu okna. Souřadnice x se zvyšuje směrem doprava, a souřadnice y se zvyšuje směrem nahoru. (<https://realpython.com/arcade-python-game-framework/>)

3.3.4 Python Matplotlib

Matplotlib.pyplot je kolekce funkcí, které umožňují knihovně Matplotlib pracovat jako MATLAB. Každá `matplotlib.pyplot` funkce provede nějakou změnu na obrázku: např. vytvoří obrázek, vytvoří kreslicí plochu v obrázku, vykreslí některé čáry do kreslicí oblasti.

`matplotlib.pyplot` sleduje věci, jako je aktuální obrázek a vykreslovací plocha, a vykreslovací funkce jsou směřovány k aktuálním osám („osy“ na většině míst v dokumentaci odkazují na osovou část obrazce a ne na striktní matematický termín pro více než jednu osu). (<https://matplotlib.org/3.5.1/tutorials/introductory/pyplot.html>)

3.3.5 Qt Designer + knihovna PyQt5

Jednotlivé kódy byly napsány jako moduly, které se importují z jednoho kódu, k ovládní byla využita knihovna Pythonu PyQt5, a GUI vytvořené v nástroji Qt Designer. PyQt5 je napsaná v programovacím jazyce C++ a umožňuje ovládat Qt GUI framework z kódu napsaném v Pythonu. (<https://build-system.fman.io/pyqt5-tutorial>)

GUI (grafické uživatelské rozhraní) umožňuje ovládat počítač pomocí zobrazených oken na monitoru představující výstup pro program. Uživateli usnadňuje práci pomocí tlačítek, vepisováním textu do řádku, výběru z možností v *comboboxu* a jiných, vše za použití klávesnice a myši. Tím pádem se uživatel nemusí orientovat v kódu a měnit hodnoty proměnných podle syntaxe programovacího jazyka.

Qt Designer je uživatelský nástroj sloužící pro návrh a vytváření grafických rozhraní (GUI). Widgets a formuláře vytvořené pomocí Qt Designer se hladce integrují s naprogramovaným kódem pomocí mechanismu signálů a slotů, může se tak grafickým prvkům snadno přiřadit funkce v kódu, která určuje jejich chování. (<https://doc.qt.io/qt-5/qtdesigner-manual.html>)

4 VÝSLEDKY

4.1 Analýza zkoumaných dat v MOLE

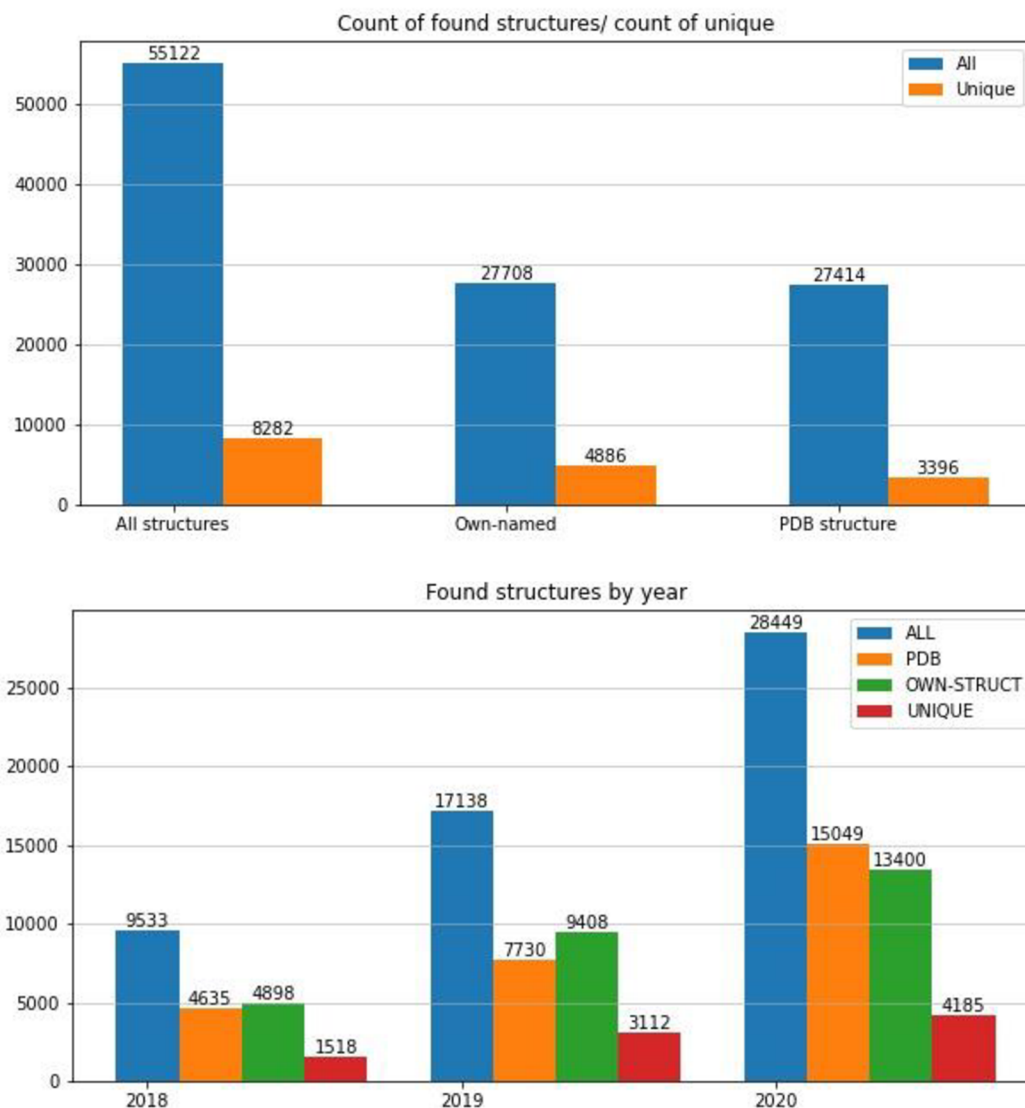
Nejprve došlo ke zpracování dat, které byly poskytnuty vedoucím diplomové práce z historie výpočtů nástrojem MOLEonline. Šlo o informace vyhledávání a zkoumání tunelů v proteinech pomocí softwaru MOLEonline. Od roku 2018 do 2020. Záznamy o vyhledávání jsou uloženy do tří složek podle let, ty dále obsahují podsložky podle jednotlivých vyhledávání. Takové složky obsahují vstupní *.xml soubor s parametry vyhledávání. V další složce se nachází výsledky také ve formátu *.xml jednotlivých vyhledávání a to tunely, póry, počáteční body, dutiny, cesty. Dále jsou dostupné informace pro vizualizaci identifikovaných tunelů v proteinech.

4.1.1 Analýza input.xml

Pomocí kódu napsaném v programovacím jazyce Python byly vybrány všechny dostupné soubory input.xml. Tyto soubory obsahují informace o zadávaném proteinu k identifikaci v MOLE, z tohoto souboru lze zjistit `PDBId` (PDB identifikátor), nebo zda zadávaný protein představuje vlastní strukturu. Dále lze pomocí napsaného programu vyčíst, nebo extrahovat informace o nastavení parametrů pro vyhledávání. Těmito parametry jsou *ProbeRadius*, *InteriorThreshold*, *IgnoreHETAtoms*, *IgnoreHydrogens*, *BottleneckRadius*, *BottleneckTolerance*, *MaxTunnelSimilarity*, *OriginRadius*, *SurfaceCoverRadius*, *WeightFunction*, *UseCustomExitsOnly*. V samotném nástroji probíhá vyhledávání s defaultními hodnotami, nebo s uživatelem změněnými parametry. Získané informace pomocí programu byly z *DataFramu* knihovny Pandas uloženy do *.csv souboru. To následně umožňuje rychleji zpracovávat dané informace, bez opakovaného vyhledávání jednotlivých souborů z mnoha složek uložených na disku.

Následně byl daný *.csv soubor načten v dalším programu, který data zpracovává a vyhodnocuje. Došlo k identifikaci počtu zkoumaných proteinů v jednotlivých třídách, jako je struktura s PDB, nebo vlastní struktura (bez `PDBId`) a také počet všech struktur. Celkem bylo zkoumáno 55122 struktur, z toho bylo 27414 s PDB identifikátorem a zbytek 27708 bylo struktur vlastních (Obr. 9). Informace o tom, zda je struktura PDB, nebo vlastní byly získány ze sloupce tabulky `isPDB` (True/False), který byl vyplněn z *.csv souboru, po předchozím určení podle délky názvu (4 = PDB). Vzhledem k tomu, že některé struktury byly zkoumány víckrát než jednou, bylo pro každou třídu spočítáno

kolik unikátních struktur obsahuje, celkově bylo zkoumáno 8282 struktur s odlišným pojmenováním. U výsledků unikátních struktur může být zkrácený počet stejným pojmenováním jiné struktury (především u vlastních struktur). Získané informace byly pomocí knihovny Matplotlib programovacího jazyka Python vizualizovány v grafu, jako obrázek *.jpg (Obr. 9).



Obr. 9: Celkový počet vyhledávaných struktur za tři roky (2018-2020) v MOLEonline činil 55 122 struktur. Vzhledem k tomu, že hodně struktur bylo vyhledáváno opakovaně je vyhledávaných struktur s odlišným jménem celkem 8 282. Vyhledávání struktur pod PDB identifikátorem a pod vlastním názvem bylo ve skoro stejném zastoupení a to 27 414 (PDB)/27 708 (vlastní struktury). Tyto počty byly získány sečtením řádků, které mají ve sloupci `isPDB` True (PDB), nebo False (vlastní struktura). Unikátních PDB struktur bylo 3 396 a unikátních vlastních struktur bylo 4 886. Spodní graf ukazuje zpracování všech input.xml souborů podle rozdělení do složek, tedy podle jednotlivých let vyhledávání v MOLE.

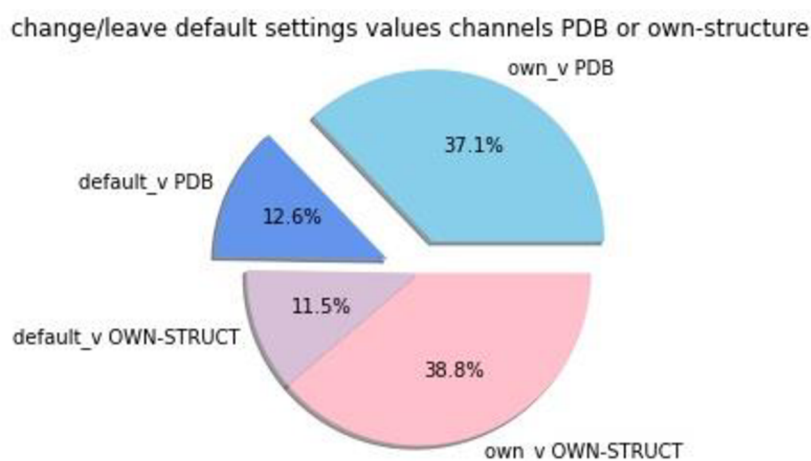
Dále byly informace rozděleny do jednotlivých let, podle dostupných složek na disku. Šlo o tři po sobě jdoucí roky a to 2018, 2019 a 2020. Pomocí programu byly extrahovány input.xml soubory z disku, následně zpracovávány k získání daných informací a ty ukládány do tří *.csv souborů podle jednotlivých roků. Další program načel všechny tři předchozí soubory a vypočítal kolik bylo zkoumáno struktur v jednotlivých letech celkově (obr. 9) takovou informaci uložil do grafu jako ALL. Dále byl pro každý rok identifikován počet zkoumaných PDB struktur a struktur vlastních v grafu jako PDB a OWN-STRUCT. Jako poslední ukazatel byl identifikován počet zkoumaných struktur s unikátním jménem (UNIQUE). Z grafu lze vyčíst, že počet vyhledávání napříč jednotlivými roky se zvyšuje, a to každým rokem více než 1,4krát. Spolu se zvyšujícím počtem zkoumaných molekul roste i počet v jednotlivých skupinách, z čehož by se dalo říct, že rostou úměrně k celkovému počtu zkoumaných biomakromolekul.

Jako další byly identifikovány nejčastěji se vyskytující struktury podle jejich názvu. Takové struktury byly také rozděleny podle toho, jestli jde o PDB identifikátor či vlastní strukturu. V tabulce (Tab. 1) je vypsáno 7 nejčastějších z každé skupiny. Některé struktury se vyskytují v tabulce dvakrát, jak ve sloupci pro všechny struktury, tak pro specifickou skupinu PDB nebo vlastních struktur. Z tabulky lze vyčíst, že nejčetnější struktury jsou z množiny PDB struktur. Zároveň struktura s PDB identifikátorem 1TQN se vyskytuje více než dvakrát častěji než ostatní struktury. To je následkem toho, že tato struktura je přednastavena jako defaultní příklad. Dalšími významnými zkoumanými strukturami jsou 1EVE (acetylcholinesteráza) a 4CZH (MreB aktinový protein).

Tab. 1: Vypsáno sedm nejčastěji zkoumaných struktur v MOLEonline (z let 2018, 2019 a 2020) ze skupin všech struktur, s PDB identifikátorem a struktur vlastních.

Struktura (vše)	Počet	Struktura (vlastní)	Počet	Struktura (PDB)	Počet
1tqn	3345	6VCH_Co-CND-APO	476	1tqn	3345
4czh	1394	6VCF_Fe-CND	215	4czh	1394
1eve	1247	Aga_lig	207	1eve	1247
1k4c	801	final	189	1k4c	801
4csk	778	CYP	171	4csk	778
5vkq	730	Model	159	5vkq	730
6VCH_Co-CND-APO	476	T12_R230Q	153	1ymg	228

Pro následnou analýzu dat byly použity hodnoty *Bottleneck tolerance* (defaultní nastavení 3 Å, maximální délka segmentu kanálu užší než poloměr úzkého hrdla (*Bottleneck radius*), může být použit pro detekci uzavřeného kanálu), *Max tunnel similarity* (defaultní nastavení 0,7, limit podobnosti čtyřřstěňů mezi kanály, delší podobný kanál je vyřazen), *Origin radius* (defaultní nastavení 5 Å, optimalizace počátečního bodu ve sféře), *Bottleneck radius* (defaultní nastavení 1.2 Å, minimální poloměr kanálu) a *Surface cover radius* (defaultní nastavení 10 Å, jsou zobrazeny pouze kanály s koncovými body dál než tento parametr). (<https://mole.upol.cz/documentation/>) Z příslušných *.csv souborů s vlastnostmi ze souboru input.xml byly pomocí naprogramovaného kódu určeny nejčastější hodnoty pro jednotlivých pět parametrů a následně vyneseno do tabulek nejčastějších pět hodnot pro daný parametr u struktur PDB a struktur vlastních. Ze všech získaných informací byly vytvořeny grafy, které odlišují barevně vlastní strukturu a PDB. Tyto grafy ukazují kolikrát se vyskytovala daná hodnota parametru v *datasetu*. Vzhledem k tomu, že četnost výskytu je velice proměnlivá pro jednotlivé hodnoty (činí rozdíl až 17 000) bylo použito logaritmické měřítko pro počet detekovaných hodnot v ose y. U všech vyhodnocených parametrů lze na první pohled zjistit, že nejčastěji je vyhledávání spouštěno s defaultními hodnotami.

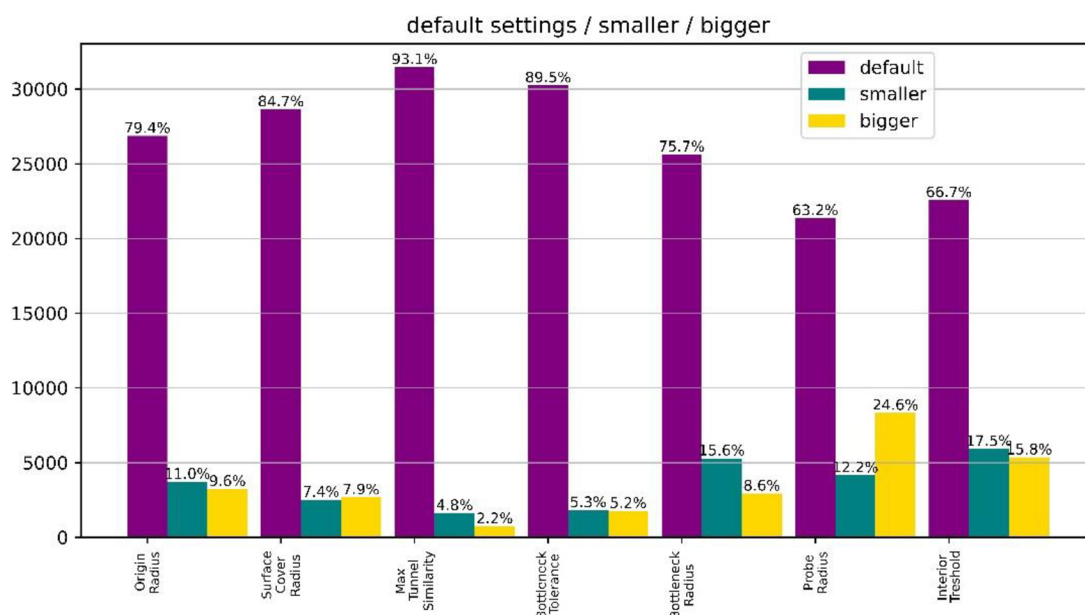


Obr. 10: Na obrázku je znázorněn graf, který představuje vyhledávané struktury v MOLEonline v letech 2018, 2019 a 2020. Celék je rozdělen do segmentů podle vlastností struktur, tedy pokud je struktura vlastní nebo s PDB identifikátorem (OWN-STRUCT/ PDB). Dále byly tyto skupiny rozděleny podle toho, jestli byly zanechány všechny defaultní hodnoty (default_v), nebo (own_v) kdy byla pozměněna alespoň jedna hodnota parametru z defaultní hodnoty na hodnotu vlastní podle uvážení uživatele. Z grafu lze vyčíst, že nejvíce zkoumaných v daných letech, 38,8 % struktur bylo vlastních spolu s alespoň jedním parametrem pozměněným, dále bylo skoro stejné množství a to 37,1 % struktur mělo PDB identifikátor a také alespoň jednu nebo více úprav parametrů. Defaultní hodnota byla ponechána pouze u 24,1 % struktur.

Jako další bylo zkoumáno procentuální zastoupení struktur analyzovaných s defaultním nastavením parametrů a v kolika případech uživatel změnil tyto parametry. Z daného souboru *.csv byly získány informace o tom, zda je struktura PDB nebo vlastní. Struktury byly rozděleny tedy na tyto dvě skupiny. Ty se dále dělily každá do dvou dalších podmnožin tak, že pokud obsahovalo vyhledávání pouze defaultní hodnoty (uživatel nezměnil ani jeden parametr) pak patří do skupiny defaultních hodnot (`default_v/ default value`) (Obr. 10). V opačném případě, a tedy pokud byla změněna alespoň jedna hodnota parametru, pak spadá vyhledávání do podmnožiny druhé, a to vlastních hodnot (`own_v/ own value`) (Obr. 10). Informace byly vyneseny naprogramovaným kódem do grafu pomocí knihovny Matplotlib. Z výsledků lze vyčíst, že nejvíce zkoumaných struktur spadalo do skupiny vlastních struktur s pozměněnými defaultními hodnotami a to 38,8 % (Obr. 10), jako další skupina podle velikosti zastoupení je množina PDB struktur se změněnou alespoň jednou hodnotou parametru a to v 37,1 % (Obr. 10). V 75,9 % byl alespoň jeden parametr jiný než u defaultního nastavení (Obr. 10).

Změna defaultního nastavení uživatelem byla zkoumána ještě z pohledu jednotlivých parametrů, které může uživatel snížit nebo zvýšit oproti defaultnímu nastavení. To bylo zkoumáno u sedmi parametrů, kterými jsou *Origin radius*, *Surface cover radius*, *Max tunnel similarity*, *Bottleneck tolerance*, *Bottleneck radius*, *Probe radius* a *Interior threshold* (Obr. 11). U všech hodnot v dostupných datech bylo provedeno porovnání s defaultní hodnotou a informace o jednom parametru byly rozděleny do tří skupin (podle toho, zda je hodnota přímo rovna defaultní hodnotě, hodnota je menší než ta defaultní, nebo naopak větší). Tato skutečnost byla vynesena do grafu (Obr. 11) pomocí vytvořeného programu.

U všech sedmi zkoumaných parametrů je více než 60 % vyhledávání prováděno s defaultním nastavením, pro daný parametr (rozmezí 63,2 % *Probe radius* – 93,1 % *Max tunnel similarity*) (Obr. 11). V 5 parametrech ze 7 bylo vícekrát změněno na menší hodnotu, než je ta defaultní. Ostatní dva případy jsou pro *Probe radius* a u *Surface cover radius*, u nichž došlo více ke zvyšování hodnoty. V některých případech nedává součet celých 100 %, což je způsobeno zaokrouhlením na jedno desetinné místo.

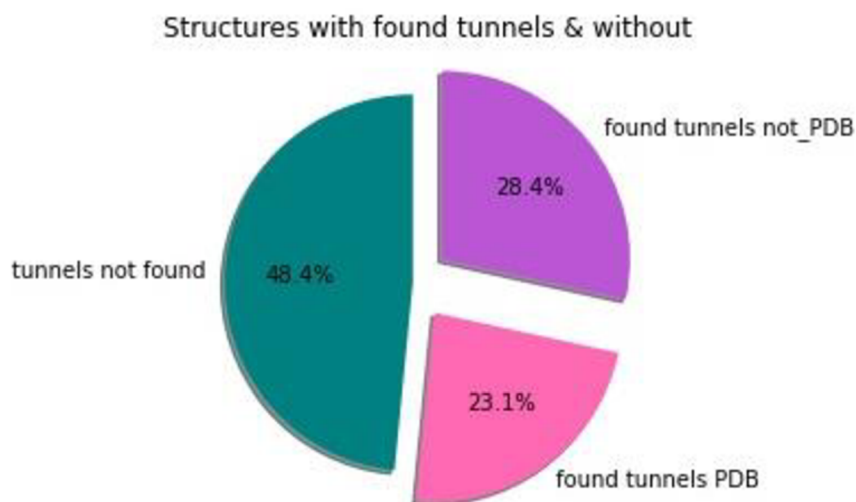


Obr. 11: Graf získaný pomocí naprogramovaného kódu. Ukazuje, jak často uživatelé měnili defaultní hodnoty parametrů při vyhledávání v MOLEonline (2018 – 2020). Z grafu lze vyčíst, zda uživatelé defaultní hodnoty spíše snižovali nebo zvyšovali. Ve více než polovině případů u všech zkoumaných parametrů zůstávala přednastavená hodnota. Z dostupných sedmi parametrů se nejméně měnil parametr Max tunel similarity. Naopak nejvíce se měnil parametr *Probe radius*, což odpovídá tomu, že tento parametr hraje roli při hledání různých velikých tunelů.

Z těchto dvou výzkumů s defaultní hodnotou, lze říct, že ve většině případů dochází ke změně alespoň jednoho parametru a to v 75,9 % (obr. 10). Ve 24,1 % nedošlo ke změně žádného parametru při vyhledávání. V průměru zůstávají jednotlivé parametry jako defaultní v 78,9 %. Lze tedy říci, že ve většině vyhledávání probíhá změna alespoň jednoho parametru nebo více, ale málokdy probíhá změna všech možných parametrů.

4.1.2 Analýza tunnels.xml

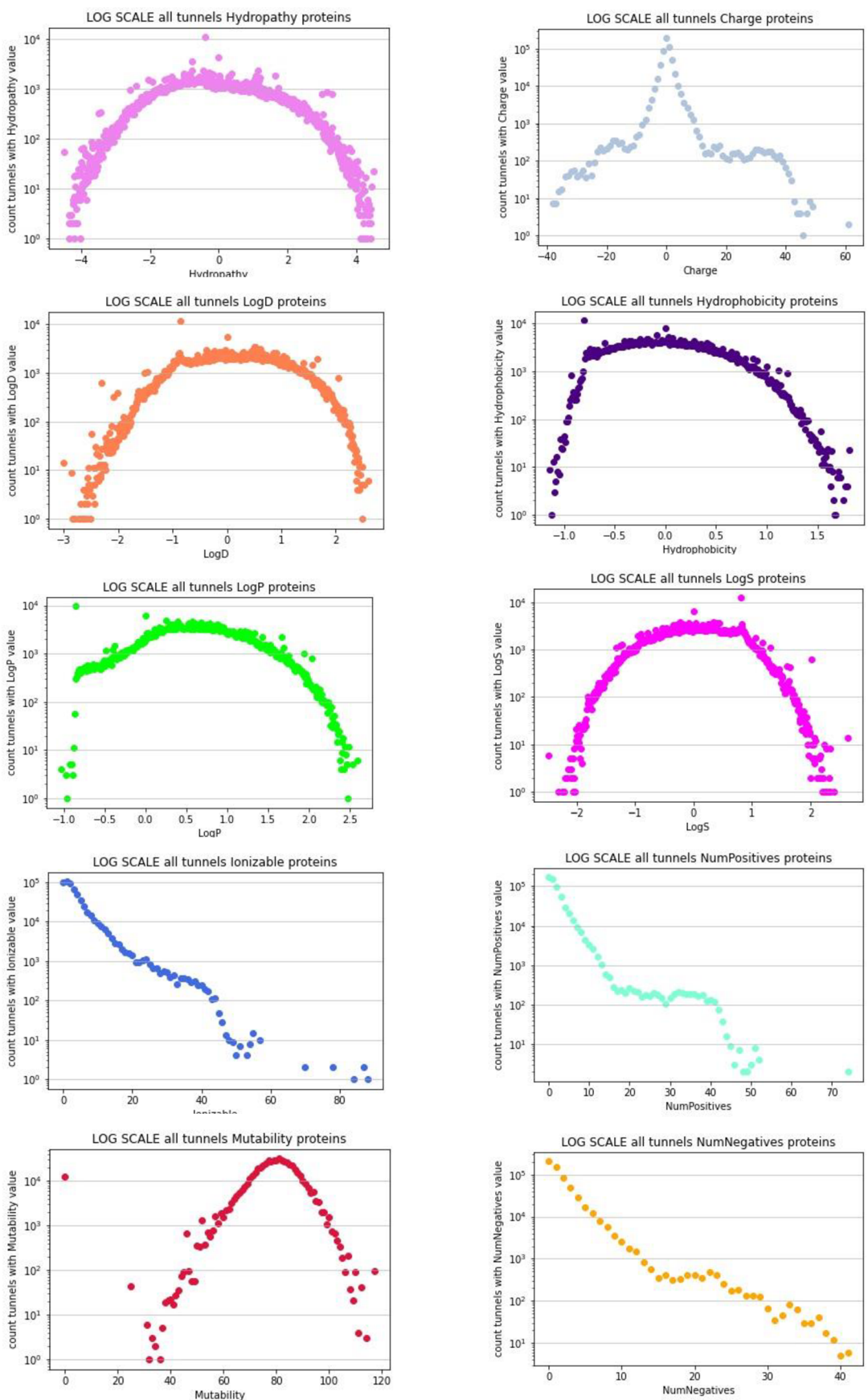
Následně proběhla analýza vyhledaných tunelů v MOLEonline v letech 2018-2020. V dostupných datech byly pomocí programu vyhledány soubory tunnels.xml, které popisují nalezené tunely v proteinech. Získané informace byly jako v předchozím případě uloženy do *.csv souboru. Nejdříve proběhlo zkoumání všech tunelů v proteinové struktuře a jejich minimální, průměrná a maximální hodnota spolu s mediánem byly uloženy do příslušného souboru. To znamená, že pro každou strukturu proteinu byly čtyři hodnoty ke každému parametru tunelu. Parametry v případě tunelů jsou *Hydrophobicity*, *Hydropathy*, *Charge*, *Ionizable*, *LogD*, *LogP*, *LogS*, *Mutability*, *Num Positives*, *Num Negatives*.



Obr. 12: Graf ukazuje poměry struktur zkoumaných v MOLEonline v letech 2018-2020. Žádný tunel nebyl nalezen u 48,4 % (tunnels not found). Zbytek tvoří struktury, u kterých byl nalezen alespoň jeden tunel a tato část je rozdělena podle toho, jestli jejich struktura je vlastní nebo PDB. (Je způsobena chyba zaokrouhlením, součet není 100 %.)

V některých případech byly soubory tunnels.xml prázdné, což se projevilo i v získaných datech. Takový fakt může být způsoben několika faktory, a to jak nesprávným nastavením *Probe radius*, *Interior treshold* nebo kritérii filtrování, nesprávným umístěním počátečního bodu kanálu, substrát může blokovat kanál, nebo tam žádný kanál není. (<https://mole.upol.cz/FAQ>) Dále proběhl výpočet programem, který počítal, jaké zastoupení tvoří nalezené tunely a v jakém množství nebyly nalezeny žádné tunely v proteinové struktuře. Výsledkem bylo (Obr. 12), že ve 48,4 % strukturách nebyly nalezeny žádné tunely. V 51,6 % byl nalezen alespoň jeden tunel ve struktuře a větší část z nalezených tunelů tvoří proteiny s vlastní strukturou a to 55 % ze struktur s nalezenými tunely.

Ze získaných dat parametrů kanálů ve struktuře proteinu (pro každý protein čtyři hodnoty - minimální, maximální, průměrná hodnota a medián), byly následně vytvořeny grafy pomocí vlastního programu. Kód byl naprogramován v programovacím jazyce Python a pomocí knihovny Matplotlib byly vyneseny informace do grafů. Pro každý parametr u tunelů je vytvořen graf, který na ose x obsahuje hodnoty parametru, na ose y je počet kolikrát se daná hodnota vyskytovala v *datasetu*. Osa y je v logaritmickém měřítku, kvůli velkým rozdílům ve frekvenci jednotlivých hodnot. Grafy zobrazují tři hodnoty pro každou proteinovou strukturu, a to minimální a maximální hodnotu parametru v dané jedné struktuře proteinu a hodnotu mediánu vypočítanou ze všech tunelů v dané struktuře proteinu, nejsou tak zastoupeny hodnoty všech tunelů z dostupných.



Obr. 13: Sada grafů pro vlastnosti specifické pro nalezené tunely v MOLEonline v letech 2018 - 2020. Grafy zobrazují na ose x hodnotu parametru a osa y ukazuje kolikrát byla tato hodnota zastoupena v daném *datasetu*, y osa je v logaritmickém měřítku.

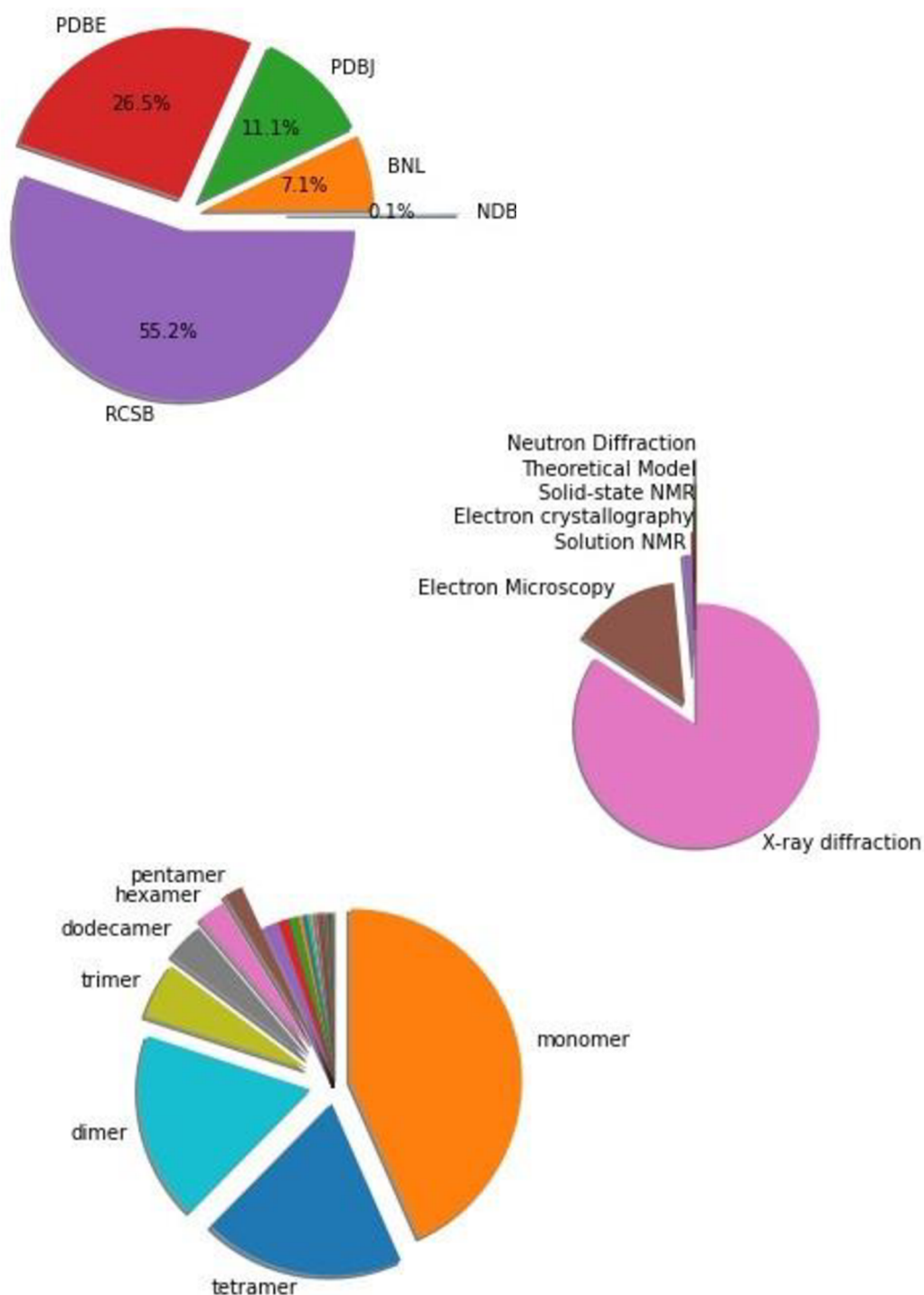
Byla provedena analýza všech nalezených tunelů ve všech zkoumaných strukturách proteinů. K tomu sloužil kód, který z dostupných dat extrahoval hodnoty parametrů pro všechny tunely a zapisoval je do *.csv souboru. Tento soubor byl dále zpracován a výsledky byly vyneseny do grafů, pro každý parametr zvlášť (Obr. 13). Z grafů lze vyčíst, že u většiny hodnot dochází k normálnímu Gaussovu rozdělení, avšak u hodnot *NumNegatives*, *NumPositives* a *Ionizable* dochází k celkovému klesání od nejčastěji zastoupené nejnižší hodnoty.

4.1.3 Analýza zkoumaných struktur

Proběhla analýza zkoumaných struktur, podle jejich PDB identifikátoru (Obr. 14). Takové vlastnosti byly pomocí kódu získány přímo s použitím RCSB PDB. Všechny informace jsou výsledkem pozorování zkoumaných struktur v nástroji MOLEonline za roky 2018, 2019 a 2020.

Jako další proběhla analýza struktur, z jakých metod byla struktura získána, a to u struktur zkoumaných v MOLEonline. Více než tři čtvrtiny z těchto struktur byly získány metodou rentgenové difrakce. Jako další metoda, která tvoří méně, než čtvrtinu zkoumaných struktur je elektronová mikroskopie. Nejmenší část z frekventovanějších struktur představují struktury z NMR. Počet struktur zkoumaných proteinů získaných pomocí rentgenové difrakce koreluje s počtem struktur obecně dostupných v databázi PDB (86,9 %). Dostupné struktury v PDB získané metodou NMR představují 7,2 % struktur, dalších 5,7 % pochází z elektronové mikroskopie. Méně než 0,1 % struktur byla získána ostatními metody.

Byla také použita informace o tom z kolika podjednotek se struktura zkoumaného proteinu v MOLEonline skládá ve chvíli výpočtu. Výsledkem bylo zjištění, že méně než polovina struktur tvoří monomery, poté tvoří skoro stejné zastoupení struktury tetrameru a dimeru. Následuje trimer, dodecamer, hexamer, pentamer a další.



Obr. 14: Grafy ukazující informace struktur z RCSB PDB, podle jejich identifikátoru. První graf zobrazuje informaci, přes jakou část databáze byly jednotlivé data uloženy. Druhý graf zobrazuje z jakých metod pochází, která část zkoumaných hodnot a třetí graf ukazuje z kolika podjednotek se struktura skládá. Vše je zobrazeno koláčovým grafem.

4.2 Program sloužící k 2D vizualizaci tunelu proteinu

2D vizualizace tunelů v proteinech byla vyřešena naprogramováním kódu v programovacím jazyce Python 3. Ten využívá výstupní soubor z nástroje MOLEonline tunnels.xml k získání informací o tunelech, spolu s ním také vstupní soubor *.cif, kvůli

znalosti identifikátoru zkoumaného proteinu. V souboru tunnels.xml jsou popsány všechny nalezené tunely zvlášť, spolu s jejich číselným identifikátorem. Dále soubor obsahuje celkové vlastnosti pro každý tunel. Mezi tyto vlastnosti patří *Mutability*, *Polarity*, *LogS*, *LogD*, *LogP*, *Hydropathy*, *Hydrophobicity*, *NumNegatives*, *NumPositives*, *Ionizable*, *Charge*. Pod nimi jsou vypsané rozměry pro tvar jednotlivých tunelů, těmi jsou souřadnice *X*, *Y* a *Z*, *Distance*, *T*, *BRadius*, *FreeRadius* a *Radius*. Ostatní informace jsou vypsané dále, jedná se o všechny aminokyseliny vyskytující se v okolí tunelu. Následují také vážené hodnoty *Mutability*, *Polarity*, *LogS*, *LogD*, *LogP*, *Hydropathy*, *Hydrophobicity*. Jako klíčové informace jsou rozepsány vlastnosti jednotlivých vrstev tunelů. K těm se řadí aminokyseliny vyskytující se v dané vrstvě, *StartDistance*, *EndDistance*, *MinBRadius*, *MinFreeRadius*, *MinRadius* spolu s vlastnostmi vrstvy tunelu *Mutability*, *Polarity*, *LogS*, *LogD*, *LogP*, *Hydropathy*, *Hydrophobicity*, *NumNegatives*, *NumPositives*, *Ionizable* a *Charge* (Obr. 15).

4.2.1 Načtení dat a jejich modifikace

Kód zpracovává soubor tunnels.xml, spolu se souborem input.xml. Prvně dojde k vyhledání daného souboru, nebo souborů vyskytujících se v zadané cestě. Tato cesta je zadána uživatelem. Uživatel tedy upřesní cestu ke složce ve svém počítači, program vyhledá soubor tunnels.xml, způsobem, že prochází všechny jména souborů ve složce, dokud nenarazí na jméno tunnels.xml. To je zajištěno pomocí modulu `os` v Pythonu, který poskytuje funkce pro interakci s operačním systémem. `os` spadá pod standardní obslužné moduly Pythonu. Tento modul poskytuje přenosný způsob používání funkcí závislých na operačním systému. Moduly `os` a `os.path` obsahují mnoho funkcí pro interakci se systémem souborů. (<https://www.geeksforgeeks.org/os-module-python-examples/>)

Pro takto nalezený soubor je dále vyhledán soubor *.cif. Který dále slouží k zjištění identifikátoru proteinu ať už PDB ID, nebo vlastního identifikátoru, zadaného uživatelem. Tento soubor by se měl nacházet ve stejné cestě jako soubor tunnels.xml.

Python umožňuje analyzovat soubory XML pomocí `xml.dom.minidom`, což je minimální implementace rozhraní *Document Object Model* (DOM). (<https://docs.python.org/3/library/xml.dom.minidom.html>) Díky němu lze získat text, který je napsaný v souboru input.xml. Pro toto použití bylo požadováno získat pouze

identifikátor, který se vyskytuje v uzlu Input. PDB identifikátor je získán z názvu souboru *.cif.

Následně probíhá získávání dat ze souboru tunnels.xml. Výsledné informace jsou obdrženy pomocí modulu `xml.etree.ElementTree`, který implementuje jednoduché a efektivní API pro XML (<https://docs.python.org/3/library/xml.etree.elementtree.html>). Informace jsou převedeny do textového formátu, který udržuje vzhled souboru tunnels.xml (Obr. 15). Získaná proměnná typu `str` se dále rozdělí do pole v místech textového řetězce "*Tunnel*" tj. podle jednotlivých tunelů. Následně probíhá selekce vlastností pro jednotlivé tunely, tedy na určité pozici v poli vlastností tunelů se vyhledává textový řetězec určující danou vlastnost. Tento textový řetězec předchází samotné hodnotě, která je dále klíčová. Tedy pro příklad máme textový řetězec: `<Node BRadius="4.190" Distance="7.246" FreeRadius="3.322" Radius="3.286" T="0.75000" X="3.087" Y="60.156" Z="71.403" />` z kterého chceme získat informaci o poloměru, program vyhledá v textovém řetězci pozici, kde je napsáno „*Radius*“, hodnotu získá rozdělením textového řetězce pomocí uvozovek do pole na kratší textové řetězce. Hodnota potom bude na dané pozici.

Takové vyhledávání se provádí pro text vlastností, určující jednotlivý tunel, zvlášť aby byly tunely odlišeny a zároveň se najdou všechny takové vlastnosti vyskytující se v textu. Výsledkem je tedy pole, které obsahuje stejný počet polí s hodnotami jako je počet tunelů v proteinu, každé toto pole nese hodnoty daného parametru pro specifický tunel. Takový postup je v programu proveden hned pro několik vlastností, mezi ně patří *Mutability*, *Polarity*, *LogS*, *LogD*, *LogP*, *Hydropathy*, *Hydrophobicity*, *NumNegatives*, *NumPositives*, *Ionizable*, *Charge*. Dále je toto aplikováno na informace o tvaru a velikosti tunelu a to *Radius*, *Distance*, *EndDistance*, *StartDistance*. Důležitý je také krok pro získání dat o výskytu aminokyselin v okolí tunelu.

Protože všechny residua vyskytující se v jedné vrstvě tunelu jsou umístěna v poli pro daný tunel jako textový řetězec (např. `[[,TYR 121 A,PHE 330 A,TYR 334 A`, ,TYR 121 A,PHE 330 A,TYR 334 A`], [,GLY 117 A Backbone,GLY 118 A Backbone,GLU 199 A`]]`). Dále jsou zde informace týkající se jedné aminokyseliny jako je číslo, řetězec a název aminokyseliny vloženy do pole o třech (čtyřech pokud jde o Backbone) textových řetězcích.

```

<?xml version="1.0" encoding="UTF-8"?>
- <Tunnels Timing="973" FoundOrigin="0" Version="2.5.18.1.14">
  - <Tunnel Auto="1" Cavity="1" Id="1">
    <Properties Mutability="46" Polarity="1.76" LogS="-1.23" LogD="1.67" LogP="1.67"
      Hydropathy="0.01" Hydrophobicity="0.86" NumNegatives="0" NumPositives="0"
      Ionizable="0" Charge="0"/>
    - <Profile>
      <Node Z="68.555" Y="65.803" X="3.801" Distance="0.000" T="0.00000"
        BRadius="3.530" FreeRadius="3.287" Radius="2.691"/>
      <Node Z="68.535" Y="65.774" X="3.850" Distance="0.061" T="0.01087"
        BRadius="3.516" FreeRadius="3.257" Radius="2.677"/>
      <Node Z="68.516" Y="65.742" X="3.898" Distance="0.121" T="0.02174"
        BRadius="3.503" FreeRadius="3.229" Radius="2.666"/>
      <Node Z="68.499" Y="65.708" X="3.944" Distance="0.180" T="0.03261"
        BRadius="3.483" FreeRadius="3.202" Radius="2.659"/>
    </Profile>
  - <Layers>
    <ResidueFlow> TYR 121 A,PHE 330 A,TYR 334 A,PHE 331 A,PHE 290 A,TRP 279 A,PHE
      288 A Backbone,ARG 289 A Backbone,SER 286 A Backbone,ILE 287 A
      Backbone,TYR 334 A Backbone</ResidueFlow>
    <HetResidues> E20 2001 A</HetResidues>
    <LayerWeightedProperties Mutability="39" Polarity="1.95" LogS="-1.06" LogD="1.45"
      LogP="1.45" Hydropathy="-0.17" Hydrophobicity="0.72"/>
  - <Layer LocalMinimum="0" EndDistance="0.31976" StartDistance="0.00000"
    MinBRadius="3.48652" MinFreeRadius="3.20960" MinRadius="2.66061">
    <Residues> TYR 121 A,PHE 330 A,TYR 334 A</Residues>
    <FlowIndices> 0,1,2</FlowIndices>
    <Properties Mutability="50" Polarity="1.19" LogS="-1.56" LogD="2.28" LogP="2.28"
      Hydropathy="0.07" Hydrophobicity="1.19" NumNegatives="0" NumPositives="0"
      Ionizable="0" Charge="0"/>
    </Layer>
  - <Layer LocalMinimum="1" EndDistance="11.55414" StartDistance="8.29744"
    MinBRadius="4.10819" MinFreeRadius="3.38164" MinRadius="3.16213">
    <Residues> TRP 279 A,SER 286 A Backbone,TYR 334 A Backbone</Residues>
    <FlowIndices> 5,8,10</FlowIndices>
    <Properties Mutability="25" Polarity="2.95" LogS="-0.29" LogD="0.29" LogP="0.29"
      Hydropathy="-0.57" Hydrophobicity="0.04" NumNegatives="0" NumPositives="0"
      Ionizable="0" Charge="0"/>
    </Layer>
  </Layers>
</Tunnel>

```

Obr. 15: Obrázek ukazující jeden ze souborů tunnels.xml, přesněji jde o soubor pro protein Ieve. Obrázek je spojen z výřezů ze souboru, tak aby byla vidět struktura zápisu pro jeden tunel. Ze souboru lze vyčíst identifikátor tunelu, jeho celkové vlastnosti, dále pak informace o velikosti a celkovém tvaru tunelu. Pod nimi jsou uvedeny informace ohledně jednotlivých vrstev tunelu.

Jako další byl vypsán slovník dvaceti aminokyselin, které představují klíče a hodnoty klíčů jsou vlastnosti daných aminokyselin a to polární, nepolární, bazické a kyselé. Pro dané vlastnosti aminokyselin je v kódu zadaný další slovník, který udává vlastnosti i barvy pro vykreslení aminokyselin a tím se vizuálně odliší i jejich vlastnosti. Dále je v kódu vypsáno pole barev, které se mohou používat při vizualizaci. Jednotlivé aminokyseliny zastupující daný řetězec v proteinu jsou také odlišeny barevně, to je zajištěno tím, že kód projde všechny načtené aminokyseliny a vytvoří slovník kde jsou klíči jednotlivé identifikátory řetězců a ke každému klíči je přiřazena barva ze slovníku barev.

4.2.2 Vykreslení os

Pomocí kódu probíhá vykreslení os do obrázku, tyto osy se vykreslují podle toho, jestli je orientace tunelu horizontální nebo vertikální. U vodorovného vykreslení je na ose x délka tunelu, tato délka se získává z informací v souboru tunnels.xml, přesněji jde o hodnotu *Distance*. Tato hodnota vede od nuly až do celkové délky tunelu, stejně jako osa x, která se vykresluje. Současně vede tato osa z bodu (200, 150) do bodu (šířka_obrázku-200, 150) hodnoty v pixelech. Tím je zaručen stejný rozměr kreslící plochy jak pro krátké, tak pro dlouhé tunely. Na osu jsou zaznamenány intervaly, představující vzdálenost o jednom Ångströmu. Pod osu je současně vypsána délka tunelu zaokrouhlená na jedno desetinné.

Osa y představuje šířku tunelu, a to v jeho nejširším místě. Tato informace je získávána z tunnels.xml z hodnot *Radius*. Ze získaných informací je vybrán maximální poloměr pro daný tunel, vynásoben dvěma a zaokrouhlen na jedno desetinné místo. Výsledná délka v Ångströmech je vepsána do obrázku.

U obrázku s tunelem zobrazeným vertikálně probíhá postup podobně, jen jsou prohozeny osy x a y.

4.2.3 Vykreslení obrysu tunelu

Pro vykreslení tvaru tunelu jsou důležité vlastnosti, které se získávají ze souboru tunnels.xml. Mezi tyto hodnoty patří *Radius*, který udává poloměr v daném místě tunelu, *Distance* to je vzdálenost od začátku tunelu. Spolu s nimi také hodnoty *StartDistance* a *EndDistance*, které představují začátek a konec vrstvy tunelu. Tunel je tedy vykreslován jako 2D obrázek a jeho vlastnosti jako je šířka určuje poloměr (*Radius*) a jeho délku určuje *Distance*.

Pro každý tunel je obrázek vykreslován zvlášť, funkce se volají v cyklu pro každý tunel, nebo jen jednou pro jeden konkrétní. Jsou procházeny všechny *Distance* zjištěné u zkoumaného tunelu. To způsobuje vykreslování částí tunelu, pro každou hodnotu *Distance* je vykreslen lichoběžník, který má výšku rozdílů dvou po sobě jdoucích hodnot *Distance*, a široký je jako průměr odpovídající *Distance* u vrchní hrany a spodní hrana je široká jako *Radius* odpovídající druhé hodnotě *Distance*. Takové vykreslování probíhá do doby, kdy jsou *Distance* menší než první *EndDistance*. Pokud je zobrazení tunelu horizontální, pak probíhá vykreslení zleva doprava, délka tunelu je převedena do obrázku

na velikost šířky obrázku minus čtyři sta. Konec a začátek vrstvy je tedy počítán následovně:

$$((\text{picture_w}-400)/(\text{distances}[\text{len}(\text{distances})-1]))*(\text{distances}[i])$$

Šířka vykreslovací oblasti je vydělena hodnotou *Distance* dané vrstvy, která je následně vynásobena celkovou vzdáleností tunelu. Tato hodnota je ještě přičtena k hodnotě 200 od které program začíná vykreslovat tunel. Tím se zajistí prostor pro vykreslení os a jejich popisků. Pro vykreslení, kdy je tunel orientován vertikálně probíhá stejný výpočet, akorát je hodnota odečtena od hodnoty výšky obrázku minus dvě stě.

Šířka lichoběžníku, tedy poloměr tunelu, je také přizpůsobena velikosti obrázku. Dochází k výpočtu:

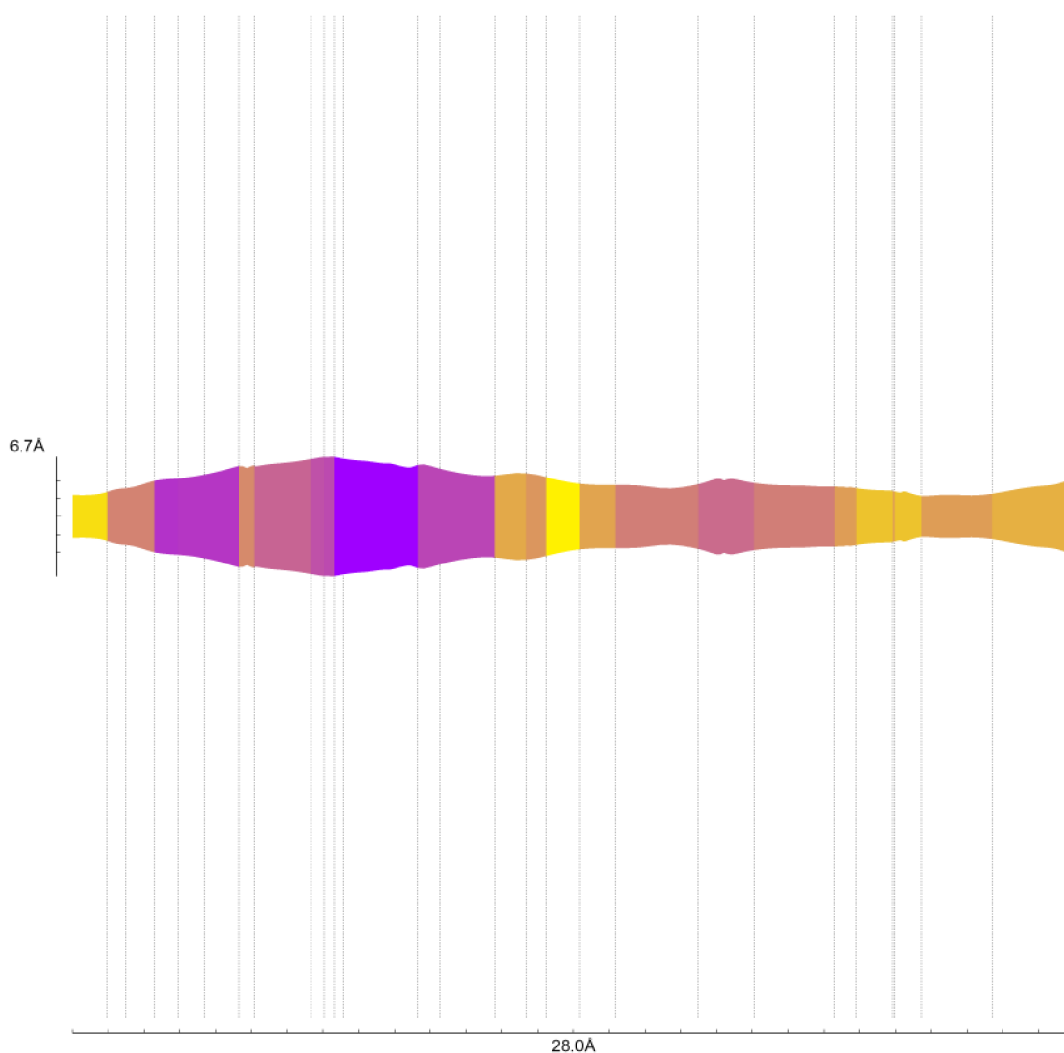
$$((\text{picture_w}-400)/2)/(\text{interval_r})*\text{float}(\text{radius}[i])$$

Polovina šířky obrázku bez čtyř set se vydělí intervalem pro vykreslení poloměru. Tento interval představuje délku tunelu, nebo hodnotu 10, pokud je délka tunelu menší než 10. To vychází z myšlenky, že čím delší je tunel, tím je v okolí více aminokyselin pro vykreslení a bude potřeba více prostoru. Čím delší je tedy tunel tím je užší vykreslení tunelu, právě kvůli zmíněným aminokyselinám na vykreslení. Pokud je ale délka tunelu menší než 10, tak je hodnota, kterou se dělí nastavena na 10. Tím se zabrání vykreslení šířky tunelu přes celý obrázek.

Pro správné vykreslení program vezme čtyři hodnoty, a to dvě po sobě jdoucí hodnoty *Distance* a dvě hodnoty *Radius* jim odpovídající (na stejném místě v poli, pro daný tunel). Na tyto hodnoty aplikuje předchozí výpočty, pro správné vykreslení v obrázku, je to jakési měřítko a následuje vykreslení, jak již bylo zmíněno.

Takové vykreslování probíhá barvou, která je vepsána ve slovníku, kde je k hodnotám vlastností přiřazená daná barva pro vykreslení (viz níže). Vykreslování probíhá stejně dokud program nenarazí na hodnotu *Distance*, která je větší než první hodnota *EndDistance*. V takovém případě dochází ke změně indexu vrstvy pro vykreslení a vykresluje se tak už druhá vrstva v pořadí, současně dojde ke změně barvy pro vykreslení, a to podle hodnoty vlastnosti dané vrstvy, změní se také hodnota *EndDistance* na koncovou hodnotu druhé vrstvy. Tento postup se opakuje, dokud se nevykreslí všechny vrstvy. (Obr. 16)

Společně s přechodem z jedné vrstvy do druhé dojde k vykreslení přerušovaných čar určujících konec jedné vrstvy a začátek druhé (Obr. 16). U horizontální orientace tunelu, k tomu dojde následovně. Vykreslování probíhá ze souřadnice x určující vzdálenost pro vykreslení dané vrstvy a 200 pixelů, která označuje začátek vykreslování přerušovaných čar (v rozmezí 200 pixelů až šířka obrázku-200 pixelů, 200 pixelů až výška obrázku -200 pixelů). Přerušování čáry probíhá v intervalu 5 a 2, tudíž 5 pixelů vykreslí a 2 ne. Současně dojde k přerušení čáry v oblasti vykresleného tunelu a následně dojde k vykreslení z druhé strany.



Obr. 16: Vykreslený obrázek pomocí program v Pythonu. Lze vidět obrys tunelu proteinu s rozdělením na jednotlivé vrstvy, ty jsou naznačeny přerušovanými čarami. Vrstvy se stejnou hodnotou vlastnosti jsou vyznačeny stejnou barvou. Současně jde vyčíst, jak je tunel dlouhý a jeho maximální šířka.

4.2.4 Vykreslení aminokyselin

Po vykreslení os a obrysu tunelu probíhá vykreslení aminokyselin. Protože v souboru není napsáno přesné umístění aminokyselin, ale je napsáno, v jaké vrstvě se vyskytují, které aminokyseliny, je vytvořen slovník. Ten obsahuje jako klíč název aminokyseliny spolu s jejím identifikátorem a názvem řetězce v kterém se vyskytuje. Prochází se postupně vrstvy a jejich příslušné aminokyseliny jsou vepsány do slovníku, k tomuto klíči je přiřazeno pole, které obsahuje poloměr (*Radius*), *StartDistance* a *EndDistance*. Pokud aminokyselina už ve slovníku je, pak se k předchozím informacím přidá informace z aktuálně procházené vrstvy. Pokud se jedná o aminokyselinu, která je popsána jako *Backbone*, pak se vypisuje bez této informace. To může být provedeno, protože jde o stejnou aminokyselinu, akorát v některé vrstvě je popsána jako *Backbone* a v jiné ne. Pokud by k tomu nedocházelo, pak by se mohlo stát, že ve výsledném obrázku dochází k duplikacím stejných aminokyselin, které obsahují název *Backbone* a které tento název neobsahují. K duplikacím by nedošlo pouze tehdy pokud by aminokyselina byla zapsaná v *tunnels.xml* pouze jako *Backbone* nebo naopak pouze bez této informace. Výsledkem je slovník, který obsahuje pro každou aminokyselinu vlastnosti potřebné k vykreslení.

Tento slovník se dále předává a aminokyseliny jako klíče jsou seřazeny podle jejich identifikátoru. Ze seřazeného pole aminokyselin se vybírají ty, které se vyskytují ve stejném řetězci a jejich rozdíl identifikátorů činí jedna. Takové residua jsou vloženy do separátních polí umístěných v poli dalším. Každé pole potom obsahuje jenom aminokyseliny, které jsou spojeny mezi sebou v řetězci. Jednotlivé aminokyseliny jsou rozděleny do dvou polí, tyto pole představují strany tunelu, na kterých se budou vykreslovat. Kvůli požadavku, aby aminokyseliny, které jsou spojeny za sebou v řetězci byly na stejné straně, proto jsou do těchto dvou polí přidávány aminokyseliny z předchozích separátních polí. Toto přidávání probíhá následovně. Do prvního pole jsou přidávány aminokyseliny, dokud je délka pole menší než polovina počtu aminokyselin. S každým přidáním dojde k vepsání všech aminokyselin spojených v řetězci (seřazené v jednom poli), proto je současně podmínkou, že počet přidávaných aminokyselin nesmí být větší než rozdíl délkou prvního pole a poloviny celkového počtu aminokyselin.

Pokud jsou aminokyseliny takto rozdělené na dvě části, pak probíhá výpočet umístění aminokyseliny. Následující postup popisuje kroky u horizontálního vykreslení. Pro každou aminokyselinu je vypočítáno umístění na y ose. To probíhá jako přičtení

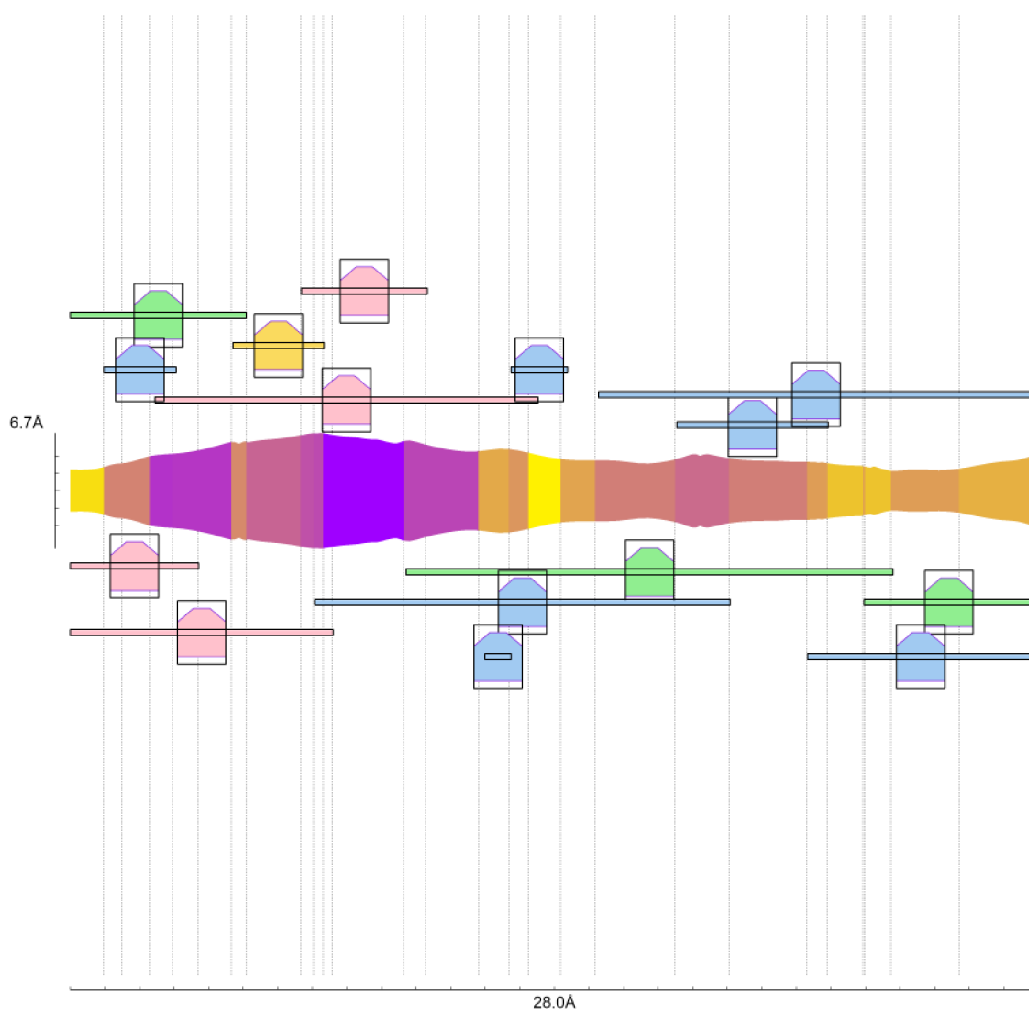
maximálního poloměru (z oblasti kde se aminokyselina nachází) k polovině výšky obrázku, pokud se aminokyseliny nachází v prvním poli (vrchní polovina obrázku), přičte se ještě polovina vykreslované aminokyseliny, aby vykreslená aminokyselina neležela přímo na struktuře tunelu. Jinak, když se aminokyselina nachází ve druhém poli (spodní polovině obrázku), pak se y souřadnice určí odečtením maximálního poloměru v oblasti tunelu, kde se daná aminokyselina nachází (aby se aminokyselina nepřekrývala s tunelem v nejširší části) spolu s výškou vykreslované aminokyseliny. Souřadnice x je vypočítána jako polovina vzdálenosti, kde se aminokyselina vyskytuje, tedy od nejmenší *StartDistance* po největší *EndDistance*. Hodnota y souřadnice slouží jako počáteční informace, která se dále může ještě změnit, x souřadnice představuje umístění aminokyseliny v tunelu, proto se dále nemění, nebo jen tak, aby její vykreslování začínalo o polovinu šířky aminokyseliny dřív, tudíž, aby polovina vykreslené aminokyseliny byla přesně v požadované hodnotě x.

Ze slovníku aminokyselin, kde jsou uloženy hodnoty *Radius*, *EndDistance* a *StartDistance* vrstev výskytu dané aminokyseliny je vybrána úplně počáteční hodnota *StartDistance* a koncová hodnota *EndDistance*. Tyto dvě hodnoty určují rozmezí, začátek a konec aminokyseliny, v délce tunelu.

Pro každou aminokyselinu a její teoretické umístění v obrázku probíhá kontrola překryvu residuí. Každá aminokyselina začíná s umístěním, které je jí dáno na začátku s vypočítanou hodnotou y a s danou hodnotou x podle toho, v jakých vrstvách se aminokyselina vyskytuje. Čím více je aminokyselin, tím je větší šance, že se na obrázku překryjí a nebudou čitelné, proto je zařazen následující postup. Prochází se aminokyseliny a jejich umístění, současně je vytvořeno pole, v kterém je uloženo 6 hodnot pro každou aminokyselinu (minimální hodnota vrstvy v které se daná aminokyselina vyskytuje, dále také tato maximální hodnota, spolu s y souřadnicí, kde se začíná kreslit aminokyselina a její výška, levá x souřadnice šestiúhelníku, pravá x souřadnice šestiúhelníku). Těchto šest hodnot tvoří dva obdélníky (Obr. 17), které obklopují obrázek aminokyseliny. Jeden z obdélníků obklopuje přímo šestiúhelník vykreslované aminokyseliny. Druhý typ obdélníku obklopuje linii, představující umístění aminokyseliny ve vrstvách.

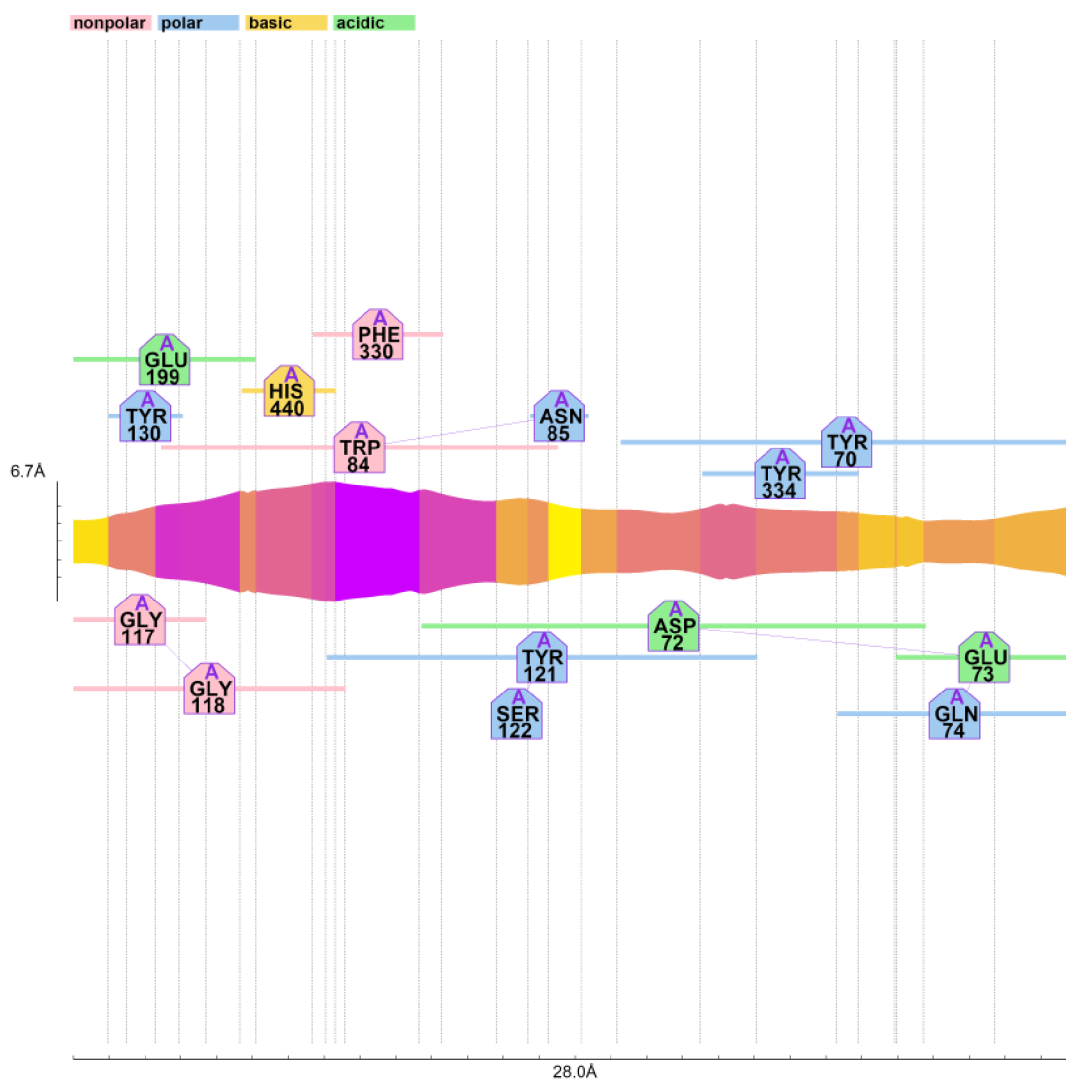
První procházená aminokyselina se uloží do pole, každá další zkoumaná aminokyselina, respektive její souřadnice prochází sérií úprav právě tehdy, když dochází k překryvu těchto obdélníků. Zkoumané residuum má prvotní vypočítané souřadnice,

pokud s takovými souřadnicemi (pro celý její obdélník) nedochází k překryvu, zůstávají souřadnice tak jak jsou a zapisují se do pole. V opačném případě, a tedy pokud dochází k překryvu obdélníku utvořeného okolo šestiúhelníku, poté se aminokyselina posouvá o výšku (šířku u vertikálního tunelu) šestiúhelníku. V případě obdélníku okolo vykreslené linie (určující vrstvy, kde se residuum nachází), kdy dvě porovnávané aminokyseliny, resp. jejich šestiúhelníky nemají žádný společný bod x , pak se aminokyselina s linií posouvají o velikost rozdílu linie jedné aminokyseliny a šestiúhelníku druhé aminokyseliny. Dojde-li u zkoumané aminokyseliny k překryvu s už zapsaným obdélníkem (pro jiné residuum) spolu s obdélníkem pro výskyt ve vrstvách, jsou tyto souřadnice upravovány do doby, kdy se nepřekrývají s žádnou již zapsanou aminokyselinou v poli.



Obr. 17: Ukázka, jak jsou rozmístěné aminokyseliny v okolí tunelu do obrázku. Pro aminokyseliny jsou postupně počítány souřadnice a porovnávány s už vypočítanými, jestli nedochází k překryvu. Jako jejich okolí je brán celý obdélník. Překryv se nesmí uskutečnit v oblasti obdélníku (okolo linie) a v ose x plus polovina vykreslované aminokyseliny (šestiúhelníku). Pokud dojde k překryvu je aminokyselina posunuta v ose y o výšku vykreslované aminokyseliny (šestiúhelníku).

V takovém případě dochází k posunu v ose y, a to o výšku rozdílu překryvu dvou porovnávaných aminokyselin spolu s 20 pixely. Pro aminokyseliny, které se vyskytují v oblasti kratší, než šířka vykreslované aminokyseliny je také zařazeno, že překryv se nesmí uskutečnit ani v případě, kdy obdélník prodloužíme v ose x o polovinu šířky vykreslované aminokyseliny v obou směrech (Obr. 18).



Obr. 18: Na obrázku jde vidět vykreslený obrys tunelu, který je rozdělený na jednotlivé vrstvy. Ty oddělují přerušované čáry. Na ose x je znázorněna délka tunelu a na ose y jeho největší šířka. Na obrázku jsou také vykresleny aminokyseliny, pro každou aminokyselinu je vykreslen jeden šestiúhelník, pod kterým vede úsečka, která udává rozmezí tunelu, kde se aminokyselina vyskytuje. barva šestiúhelníku a zmíněné úsečky je dána vlastností aminokyseliny. Úsečka vedená úhlopříčně pod aminokyselinou znamená, že v okolí tunelu se nachází aminokyselina/ny, které jsou v řetězci spojené za sebou.

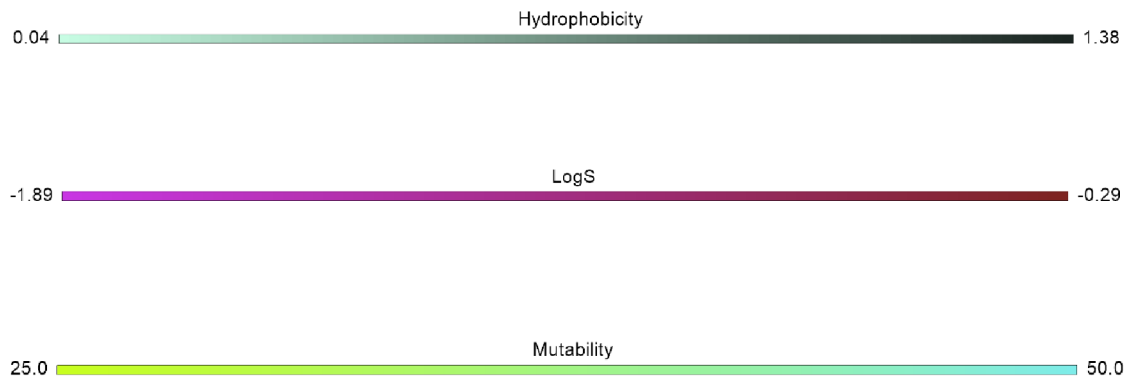
Pokud dojde k tomu, že se má aminokyselina vypisovat mimo vykreslovací plochu, tedy od 200 pixelů od každého kraje obrázku, pak se program snaží aminokyselinu přehodit na druhou stranu tunelu, kde by mělo být více místa, vypočítají se prvotní souřadnice a výpočet probíhá tak dlouho dokud program nevypočítá souřadnice na druhé straně tunelu.

Po zjištění souřadnic pro všechny aminokyseliny, kde nedochází k překryvu program pokračuje v jejich vykreslení. Nejdříve dojde k vykreslení úsečky, která leží pod vykreslovanými aminokyselinami, ta má znázorňovat, že aminokyselina je spojena s další takovou, která se nachází také v oblasti daného tunelu, a jsou spojeny v řetězci (Obr. 18). Dále dochází k vykreslení linie, ta je v barvě, která určuje aminokyselinu podle jejích vlastností (s rozdělením na polární, nepolární, kyselé a bazické). Úsečka je vedena od minima kde se aminokyselina nachází až k maximální hodnotě vrstvy, kde tato aminokyselina leží (Obr. 18). Následně je vykreslen šestiúhelník ve stejné barvě, ten je lemovaný okrajem o odlišné barvě, která odpovídá barvě daného řetězce. Dále program vypíše do prostoru šestiúhelníku řetězec, v kterém se daná aminokyselina nachází, její trojmístný název a identifikátor, určující pořadí v daném řetězci (Obr. 18).

4.2.5 Barva vrstev korelující s jejich vlastnostmi

Barevným znázorněním jednotlivých vrstev lze zobrazit danou vlastnost tunelu. Uživatel zvolí požadovanou vlastnost, kdy je na výběr *Mutability*, *Polarity*, *LogS*, *LogP*, *LogD*, *Hydropathy*, *Hydrophobicity*, *NumNegatives*, *NumPositives*, *Ionizable*, *Charge*. Spolu s vlastností uživatel zvolí také dvě barvy, kde jedna z nich představuje minimální hodnotu vybrané vlastnosti v tunelu proteinu a další představuje hodnotu maximální (Obr. 19). Pro každý tunel zvlášť jsou seřazeny hodnoty vybrané vlastnosti od nejmenší po největší tak, že se hodnoty nevyskytují duplikátně.

Je vypočítán rozdíl mezi nejnižší a největší hodnotou vybrané vlastnosti pro daný tunel, tento interval je následně rozdělen na sto dílů, kde do jednoho ze sta intervalů směřují hodnoty vlastnosti, které budou mít stejnou barvu. V rozmezí obou barev jsou vypočítány hodnoty RGB pro jednotlivých sto intervalů. Následně pro každou hodnotu uživatelem zvolené vlastnosti jsou vypočítané tři čísla určující zastoupení červené, zelené a modré barvy. Tato informace je uložena ve slovníku kde klíčem je vlastnost a hodnota k ní přidružená představuje barvu v RGB.



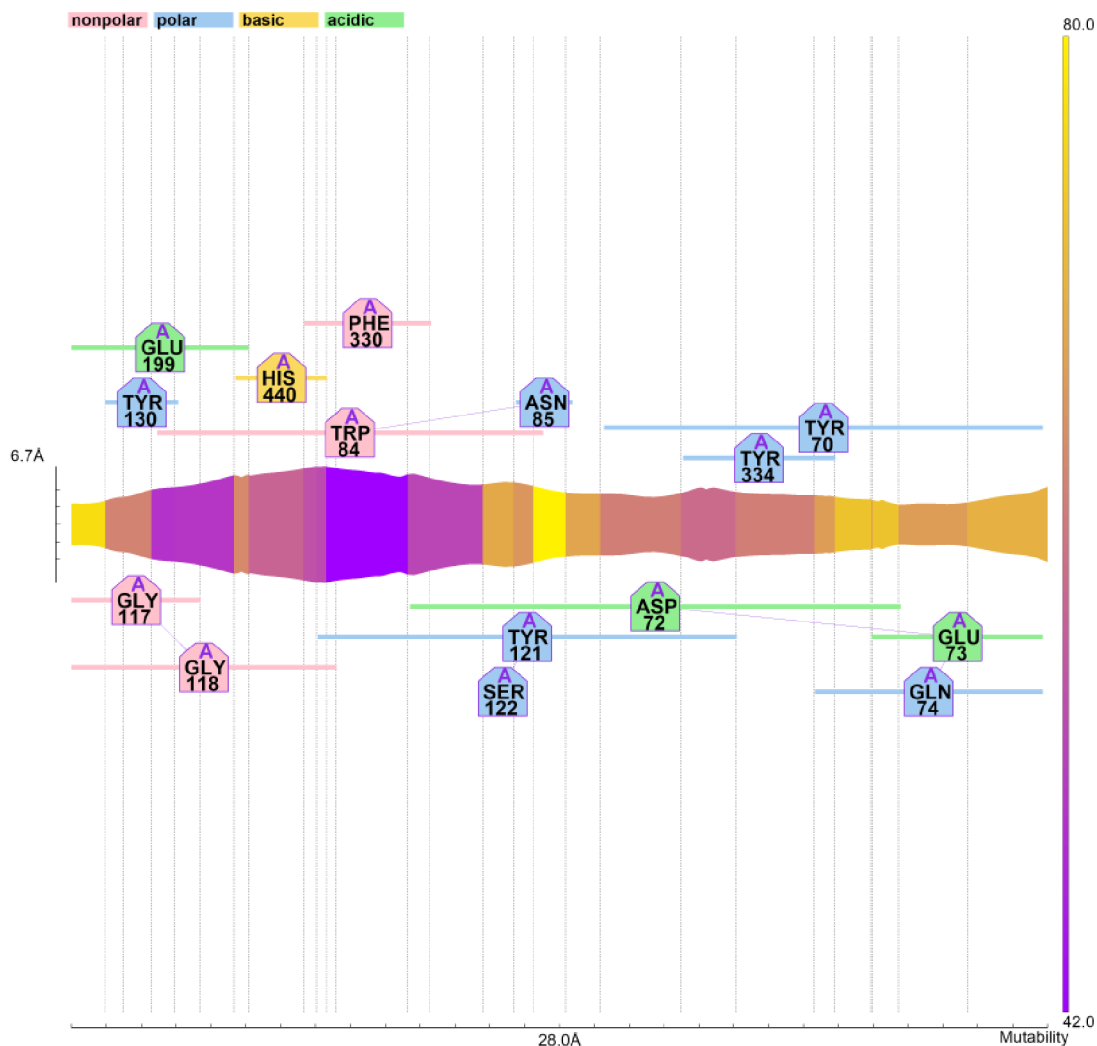
Obr. 19: Obdélník určující barvu pro hodnotu dané vlastnosti. Nad ním je napsáno, o jakou vlastnost se jedná a na okrajích jsou umístěné krajní hodnoty, kterými jsou minimum a maximum pro daný tunel.

Pro každý tunel probíhá výpočet zvlášť, tak je minimum i maximum zvoleno pro jednotlivé tunely, tudíž v jiných tunelech může tatáž barva představovat jinou hodnotu zvolené vlastnosti. Slovník barev je použit pro následující vykreslování jednotlivých vrstev (Obr. 19). Když se vykresluje vrstva s danou vlastností, vyhledá se ve slovníku její hodnota klíče a je použita její přiřazená barva ve formátu RGB.

Pro orientaci, v jakých hodnotách se vlastnosti pohybují je do obrázku přidán obdélník vedoucí ortogonálně k směru tunelu, tedy pokud je vykreslení horizontální, je obdélník vertikální (obr. 20), jinak je horizontální.

U vertikálního vykreslení představuje nejlevější barva minimum pro danou vlastnost v prohlíženém tunelu, nejmenší zjištěná hodnota je také vypsána vlevo. Stejně tak je na pravé straně zobrazena barva pro maximální hodnotu v tunelu a její hodnota. Nad obdélníkem je také vypsáno, o jakou vlastnost se jedná. Obdélník je rozdělen do sto dalších s odlišnou barvou, od barvy představující nejnižší hodnoty až po barvu reprezentující nejvyšší. Vyskytuje se v rozmezí 200 až šířka obrázku -200.

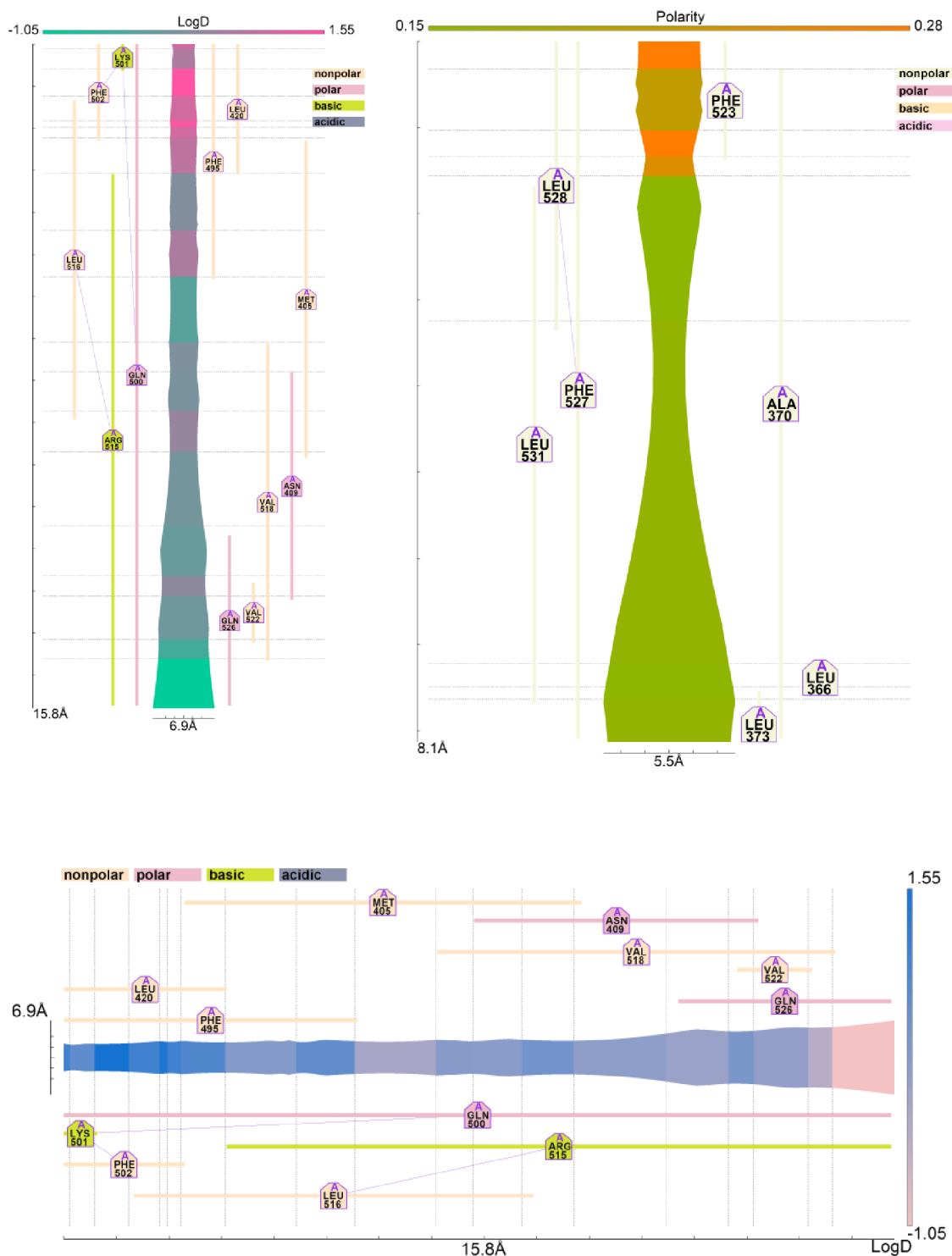
Pro horizontální zobrazení tunelu je obdélník určující hodnoty barev umístěn vertikálně, na pravé straně obrázku. Obdélník obsahuje barvu představující nejnižší hodnotu na spodní straně a barvu reprezentující maximální hodnotu vlastnosti navrchu. Současně je také vyobrazena nejnižší a nejvyšší hodnota detekovaná v daném tunelu. Pod nejnižší hodnotou je napsán parametr, který je zkoumaný.



Obr. 20: Výsledný obrázek, který je získaný pomocí kódu naprogramovaného v Pythonu, jako vstupní informace pro vykreslení je výstupní soubor tunnels.xml z MOLEonline. Je zobrazen tunel a aminokyseliny ležící v jeho okolí. Spolu se znázorněním vybrané vlastnosti pomocí barvy vrstev.

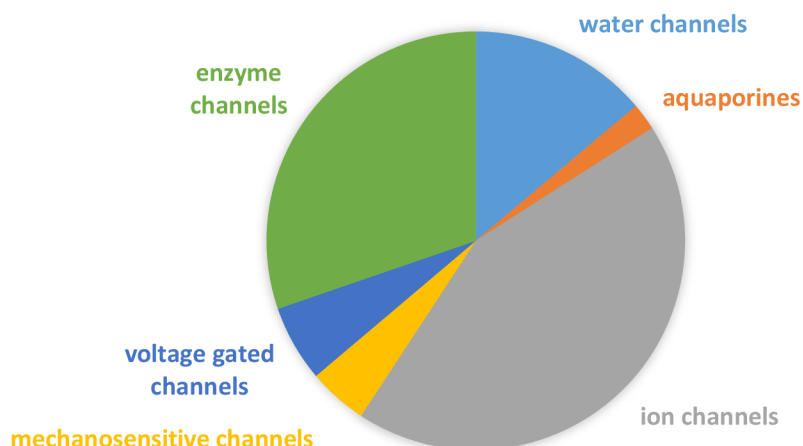
4.2.6 Úprava parametrů

V programu existuje možnost změny jednotlivých vlastností k vykreslení. První z nich je již zmíněná barva tunelu. Uživatel vybere danou vlastnost, kterou chce zobrazit a dvě barvy. Jako další možnost je volba velikosti obrázku, šířka a výška (Obr. 21). V této možnosti je doporučeno volit co největší čísla. Tyto čísla odpovídají počtu pixelům. Současně čím větší je obrázek, tím je kvalitnější a trvá delší dobu, než ho program vykreslí. Jako další parametry, které se dají volit je šířka a výška vykreslovaného šestiúhelníku představující aminokyseliny, spolu s tím je možnost zvolit velikost písma, které je do těchto šestiúhelníků vepsáno. Dále je také možnost měnit velikost popisků os a legendy. Pokud dojde k překryvu tunelu aminokyselinou je lepší zvětšit výšku obrázku, nebo zmenšit velikost vykreslovaných aminokyselin.



Obr. 21: Na obrázcích lze rozeznat, že byly zadané odlišné parametry pro vykreslení. Vrchní dva obrázky ukazují úzký a vysoký obrázek, přičemž u jednoho z nich je tunel orientován horizontálně a u druhého vertikálně. U spodního obrázku je zadána větší šířka než výška, zároveň si lze všimnout menších pětiúhelníků pro vykreslení aminokyselin v okolí tunelu.

4.3 Analýza látek procházejících tunely

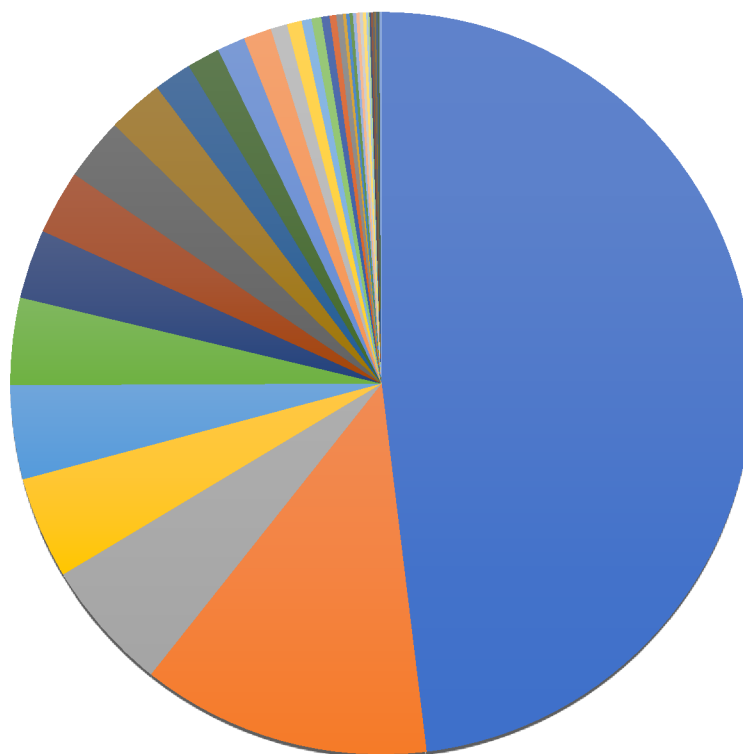


Obr. 22: Graf ukazující zastoupení proteinových struktur kanálů v databázi PDB.

Jednou z možností dělení kanálů je rozřazení do skupin podle látky, kterou transportují, nebo která se váže do aktivního místa. Mezi nejvíce zastoupenou třídou kanálů v PDB databázi (<https://www.rcsb.org/>) jsou iontové kanály (v podstatě póry) představují 43,28 % struktur, které jsou propustné pro ionty. Za ní následuje skupina enzymových kanálů se zastoupením 30,27 %. Další početnou skupinou 13,86 % v pořadí jsou kanály transportující molekuly vody. Se zastoupením 5,91 % to jsou napětím řízené kanály. Z 2,09 % jsou to akvaporiny. (Obr. 22)

Z databáze ChannelsDB byly získány uložené informace o kanálech, vyhledávání probíhalo pouze u proteinů, které mají v databázi záznam o kanálech v „*Channels description*“. Nejprve došlo ke zpracování popisu kanálu. V takovém popisu se objevovalo 28 různých typů, pro stejný typ (název) kanálu byl stejný popis jeho vlastnosti. Mezi nejčastěji zastoupené kanály v pořadí od nejfrekventovanějšího, patří kanál pro rozpouštědlo, kanál 2b, kanál pro molekuly vody, tunel do kterého putují substráty. Tyto vyjmenované tvoří více než polovinu zmíněných kanálů v „*Channels descriptions*“.

Následně byly získávány informace z ChannelsDB, které jsou v tabulce „*Channels properties*“. Tato tabulka obsahuje *jméno kanálu, délku, úzké hrdlo, hydropatii, náboj, polaritu a mutabilitu*. Zastoupení jednotlivých jmen bylo identifikováno a vyneseno do grafu (Obr. 23). Pomocí kódu byly získány i ostatní informace o vlastnostech kanálů, tyto informace byly vloženy do *DataFramu* díky knihovně Pandas.



- | | |
|--------------------------------|--------------------------------|
| ■ Tunel | ■ Iontový kanál |
| ■ Kanál pro rozpouštědlo | ■ Vodní kanál |
| ■ Kanál 2b | ■ Kanál 2e |
| ■ Substrátový tunel | ■ Kanál 2a |
| ■ Kanál 2c | ■ Channel 2f |
| ■ Channel 3 | ■ Channel 2ac |
| ■ Pór | ■ Kanál 5 |
| ■ Substrátový kanál | ■ Kanál 4 |
| ■ Substrátový/Produktový tunel | ■ Hydrofobní kanál |
| ■ Tunel pro rozpouštědlo | ■ Kanál 2d |
| ■ Peptidový kanál | ■ Transport metabolitů |
| ■ Kanál 2a | ■ Kanál 2b |
| ■ Kanál 2bs | ■ Kanál přístupný pro substrát |
| ■ Vodní tunel | ■ Nucleotidový kanál |
| ■ Přístupový kanál | ■ Kanál 1 |
| ■ Kanál 3a | ■ Produktový kanál |
| ■ GsPutA tunel | ■ Produktový tunel |
| ■ Kanál 1 | ■ Kanál 2d |
| ■ Mechanosenzitivní kanál | |

Obr. 23: Zobrazení poměru zastoupení jednotlivých kanálů z ChannelsDB zmíněných v “Channels properties”. Skoro polovina struktur nese pojmenování Tunel, ostatní hojně zastoupené jsou iontové kanály, kanály pro rozpouštědlo a vodu. Zobrazuje všechny uložené kanály v ChannelsDB, protože u všech se vyskytují popsane vlastnosti.

Z celkového *DataFrame* byly zpracovány informace, které se shodovaly v názvu kanálu, z nich byl vytvořen nový *DataFrame* (aby obsahoval pouze informace o daném typu kanálu). Následně došlo ke zpracování informací a vynesení do grafů. Pro náboj jednotlivých kanálů byl zvolen sloupcový graf, který ukazuje všechny náboje vyskytující se v ChannelsDB k danému názvu kanálu. V grafech jsou označeny záporné hodnoty náboje modrými sloupci, kladné zelenými a nula je označena červeným sloupcem.

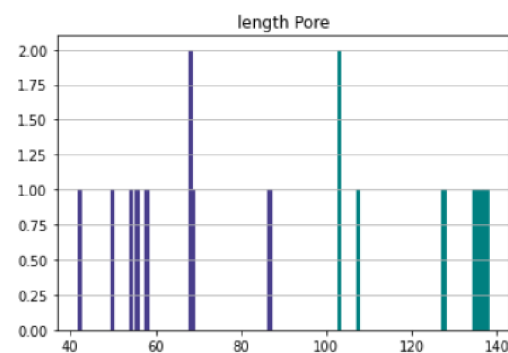
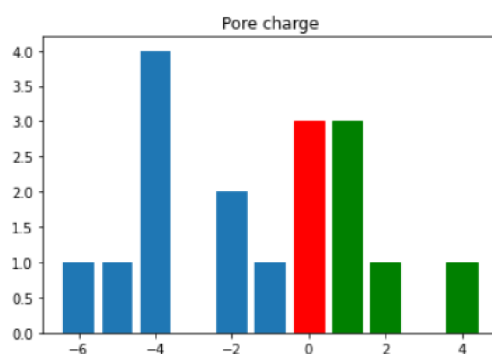
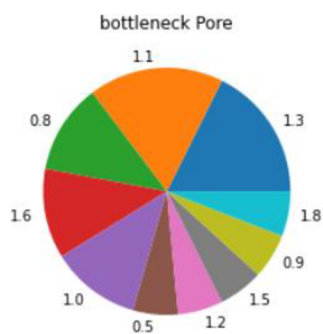
Pro znázornění informace úzkého hrdla byly vytvořeny koláčové grafy, takové grafy byly vztaženy na stejné pojmenování a bylo vypočítáno procento jednotlivých velikostí úzkého hrdla v kanálech daného typu. (Obr. 24)

Graficky došlo také ke znázornění délky kanálů, a to pomocí sloupcového grafu, který na x ose obsahuje jednotlivé délky a na y ose počet kolikrát byla tato délka identifikována v *datasetu* z ChannelsDB. Fialové sloupce představují délku menší, než je průměrná v daném typu kanálu, naopak další barva znázorňuje délky větší než průměrná hodnota. Je určena maximální velikost průměru molekuly, která může procházet danou skupinou tunelů a to tak, že je menší než nejširší úzké hrdlo ze skupiny, současně by tato látka měla být s nábojem vyhovujícím nejčastějšímu náboji kanálu ve skupině.

Pro průměrné hodnoty délky kanálu, jeho průměrný náboj a průměrnou šířku úzkého hrdla u každé skupiny byly vytvořeny grafy (Obr. 25) pomocí AMCHARTS, kde lze jednoduše porovnat odlišné skupiny (<https://www.amcharts.com/demos/column-with-rotated-series/>).

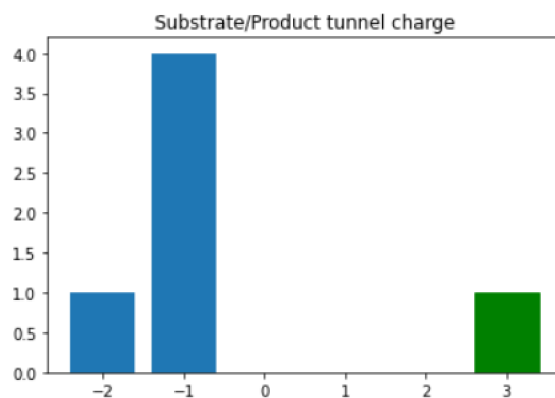
Skupina struktur uložených pod názvem *Póry*

- Průměrné úzké hrdlo 1,2 Å
- průměrná délka 91,0 Å
- nejnižší hodnota úzkého hrdla 0,5 Å
- prochází nejčastěji látka vyhovující náboji kanálu -4 a velikosti maximálně 1,8 Å

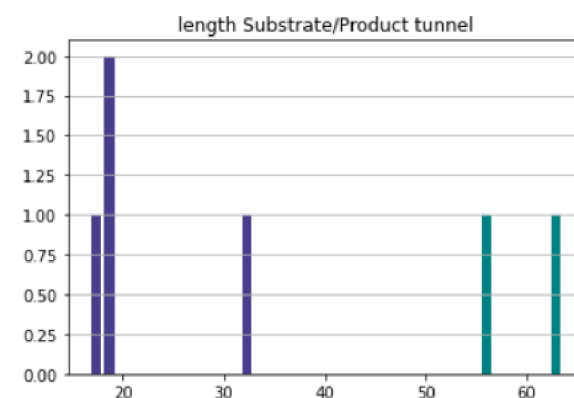
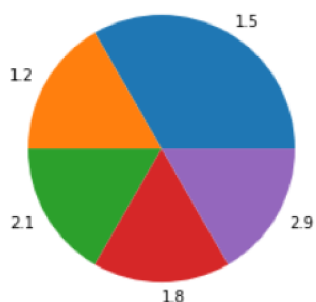


Skupina struktur uložených pod názvem *Substrát/produkt tunel*

- průměrné úzké hrdlo 1,9 Å
- průměrná délka 37,4 Å
- nejnižší hodnota úzkého hrdla 1,2 Å
- prochází nejčastěji látka vyhovující náboji kanálu -1 a velikosti maximálně 2,9 Å

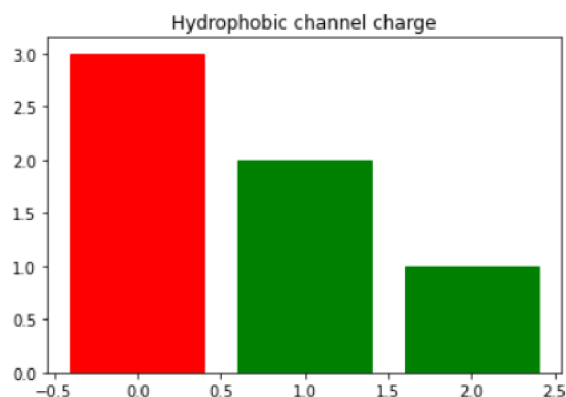


bottleneck Substrate/Product tunnel

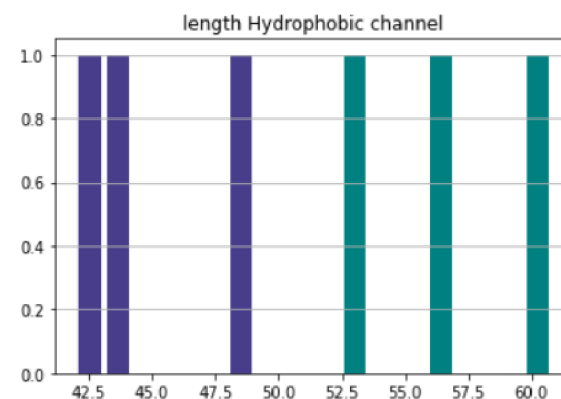
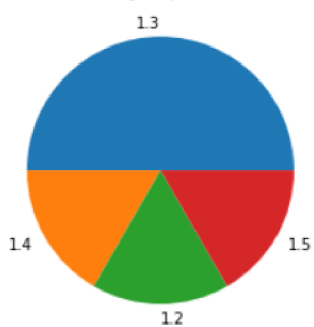


Skupina struktur uložených pod názvem *Hydrofobní kanál*

- průměrné úzké hrdlo 1,4 Å
- průměrná délka 50,7 Å
- nejnižší hodnota úzkého hrdla 1,2 Å
- prochází nejčastěji látka vyhovující náboji kanálu 0 a velikosti maximálně 1,5 Å

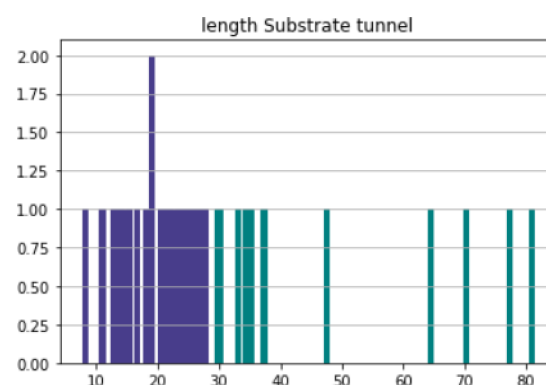
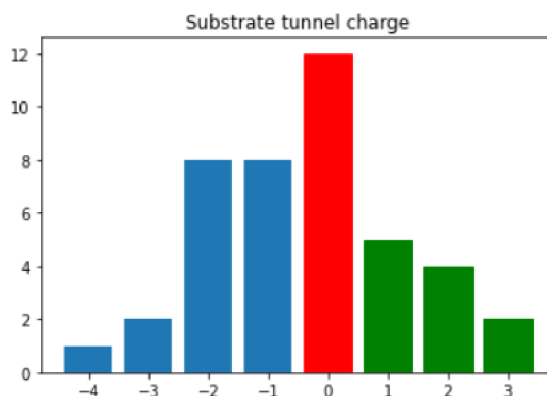
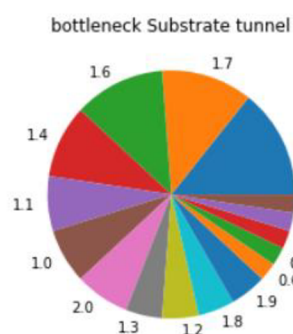


bottleneck Hydrophobic channel



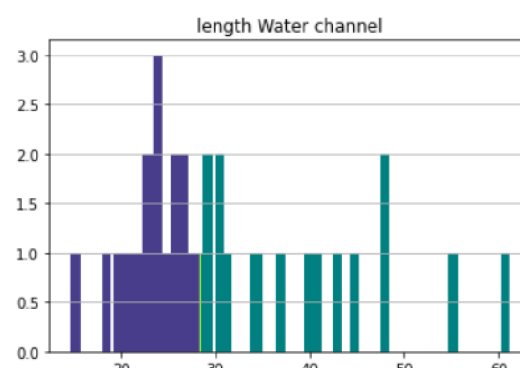
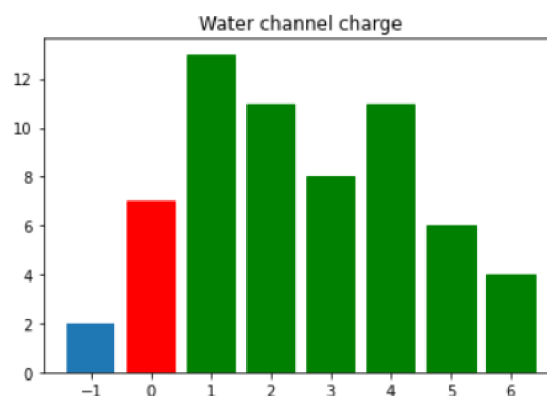
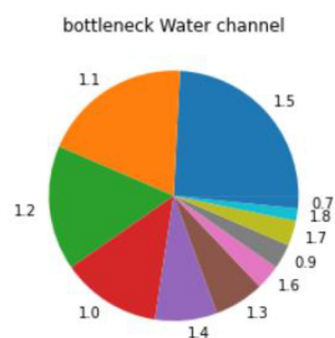
Skupina struktur uložených pod názvem *Substrátový tunel*

- průměrné úzké hrdlo 1,4 Å
- průměrná délka 28,3 Å
- nejnižší hodnota úzkého hrdla 0,5 Å
- prochází nejčastěji látka vyhovující náboji kanálu 0 a velikosti maximálně 2,3 Å

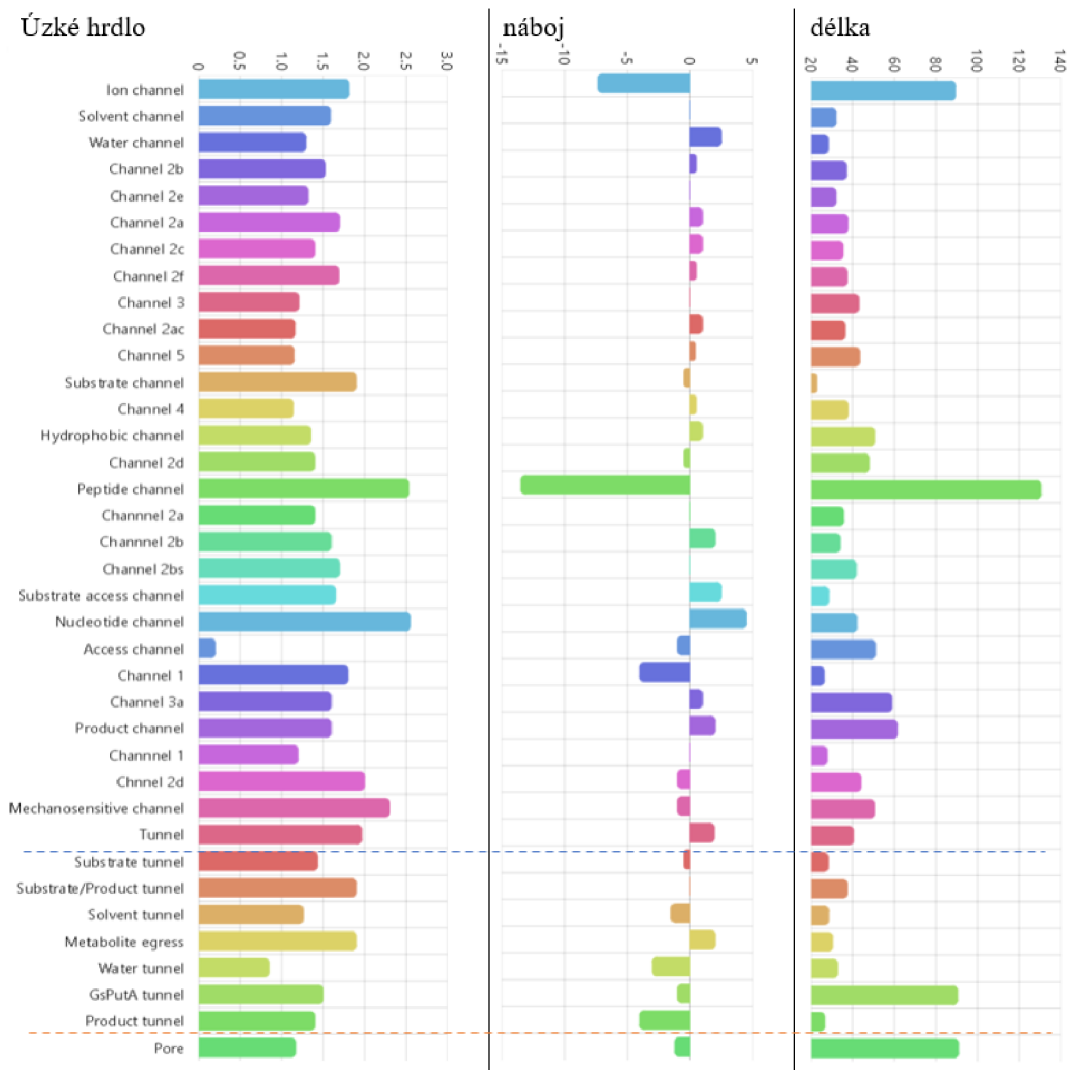


Skupina struktur uložených pod názvem *Kanál pro vstup molekul vody*

- průměrné úzké hrdlo 1,3 Å
- průměrná délka 28,3 Å
- nejnižší hodnota úzkého hrdla 0,7 Å
- prochází nejčastěji látka vyhovující náboji kanálu 1 a velikosti maximálně 1,8 Å



Obr. 27: Sada grafů pro jednotlivé kategorie kanálů získaných z ChannelsDB. Jednotlivé skupiny se od sebe liší jak v délce kanálů či nejčastějším nábojem kanálu, tak v průměru úzkého hrdla.



Obr. 25: Porovnání průměrných hodnot délky, náboje a úzkého hrdla selektivity u kanálů dostupných v databázi ChannelsDB. Jednotlivé skupiny, představují struktury se stejným pojmenováním v databázi, v řádku jsou pak uvedeny stejné skupiny pro odlišné vlastnosti. Pořadí skupin je určeno na tři podtřídy (podle názvu) a to kanály, tunely a póry, ty jsou oddělené horizontální přerušovanou čarou. Jednotlivé názvy v podskupinách jsou řazeny podle počtu výskytu v ChannelsDB a tedy představují komplexnější výsledek, než skupiny s jednou strukturou. První graf zobrazuje průměrné hodnoty úzkého hrdla v jednotlivých skupinách. V případě náboje a druhého grafu se výrazně liší skupiny peptidových kanálů a iontových kanálů. Poslední graf ukazující průměrnou délku kanálů ve skupinách zobrazuje čtyři skupiny zásadně odlišné od zbytku, a to s větší průměrnou délkou u iontových kanálů, pórů, peptidových kanálů a GsPutA tunelů. Grafy jsou vytvořeny pomocí AMCHARTS (<https://www.amcharts.com/demos/column-with-rotated-series/>). Z výsledků lze vidět, že iontové kanály, společně s póry, které představují hojně zastoupené skupiny, jsou delší než nejčetnější skupina z tunelů.

4.4 Predikce procházejících látek vzhledem k vlastnostem tunelu

4.4.1 Výpočet tvaru molekuly

Pro predikci, jaké látky by mohly procházet daným tunelem je potřeba znát vlastnosti molekul. Vlastnosti tunelu jsou známy z výstupového souboru z MOLEonline. Prvotním krokem bylo stažení *.csv souboru, obsahující informace o molekulách. K tomuto kroku došlo u ligandů a iontů v databázi PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). Stažené soubory obsahují následující informace o molekulách *cid* (pubchem cid), *compdname* (jméno molekuly), *compdsynonym* (synonyma molekuly), *mw* (molekulární váha), *mf* (vzorec molekuly), *polararea* (polární povrch), *complexity*, *xlogp*, *heavycnt* (počet těžkých atomů), *hbonddonor* (počet donorů vodíkové vazby), *hbondacc* (počet akceptorů vodíkové vazby), *rotbonds* (počet rotačních vazeb), *inchi*, *isosmiles* (isomerický smiles), *inchikey*, *iupacname* (jméno IUPAC), *meshheadings*, *annothits*, *annothitcnt*, *aids*, *cidcdate*, *sidsrname*, *depcatg*, *annotation*.

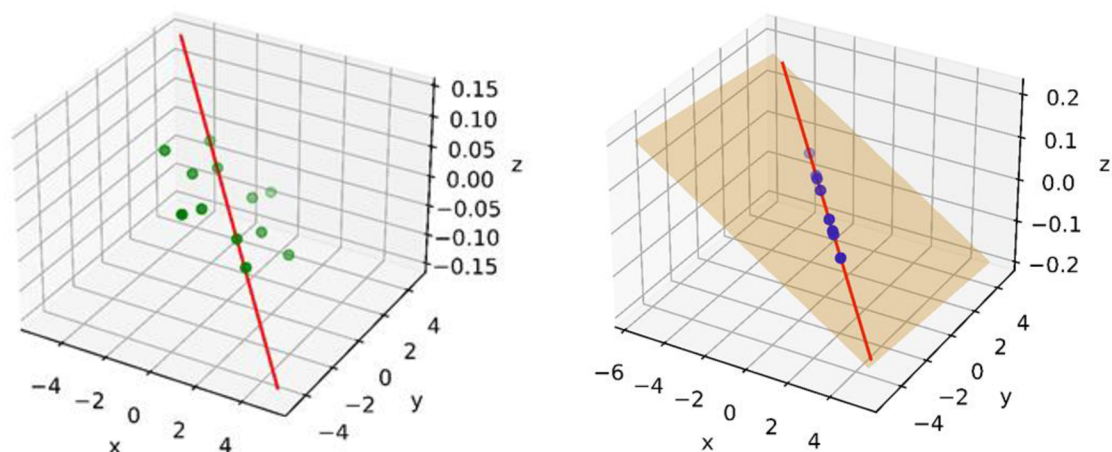
U všech dostupných molekul z ligandů a iontů byl vypočítán jejich tvar spolu s jejich nábojem pomocí napsaného kódu v programovacím jazyce Python, podle následujícího postupu. Nejprve se daný soubor *.csv načel do programu s použitím knihovny Pandas, tak vznikl *DataFrame*, který byl následně zpracováván. Pro ionty je tento *DataFrame* 81řádkový, u ligandů je to 21709 molekul.

Program po získání *DataFrame* prochází jednotlivé řádky, které obsahují molekuly a jejich informace. Pomocí isomerického SMILES a RDKit knihovny v Pythonu je vypočítán náboj dané molekuly. Díky té samé knihovně lze získat informace o umístění atomů molekuly v prostoru (ať už s vodíky nebo bez vodíků). Tyto informace jsou klíčové pro výpočet tvaru molekuly.

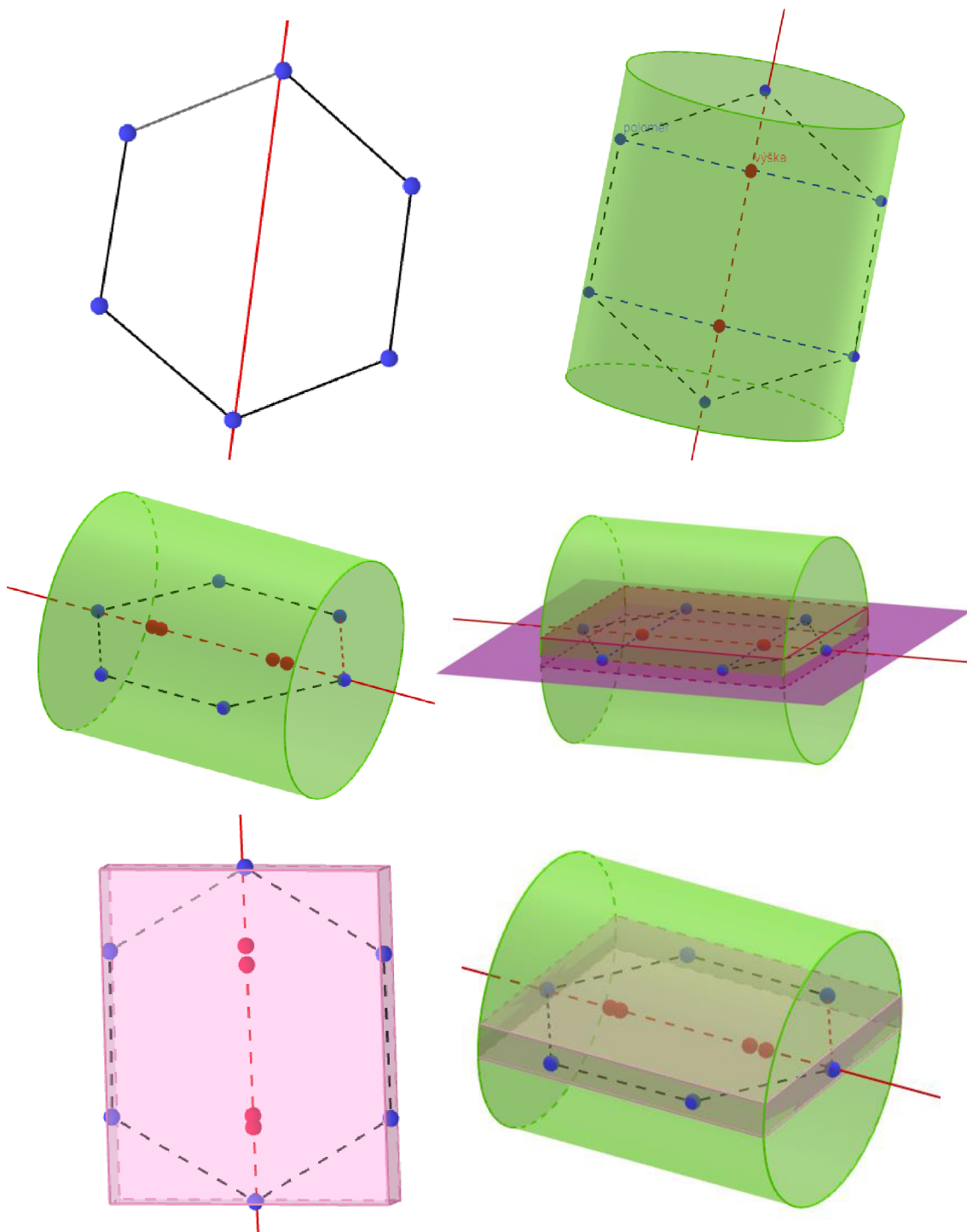
Nejprve probíhá výpočet nejmenšího poloměru molekuly (nejušší strana pro průchod tunelem), molekula je ve výsledku ohraničena válcem. Pomocí metody nejmenších čtverců je umístěna přímka mezi získané body atomů. Taková přímka pak představuje potenciální střed podstavy válce (kružnice). Následně jsou vypočítány vzdálenosti přímky od jednotlivých atomů (nejmenší možné, v pravém úhlu), vzdálenost nejdlejšího atomu pak představuje vypočítaný poloměr podstavy válce. Nejevzdálenější z toho důvodu, aby kružnice byla opsaná všem atomům ze zkoumané molekuly. (Obr. 26)

Další informace o velikosti molekuly je výška, ta se získává také z daného válce a je to vzdálenost podstav válce. Body představující atomy molekuly se promítnou na přímku, která je proložena body pomocí metody nejmenších čtverců. Body se tak vyskytují v jedné linii a dva nejvzdálenější body představují výšku válce, tedy i výšku molekuly. (Obr. 26)

Kvůli molekulám, které mají spíše plochý tvar je zařazen výpočet nejvyšší strany molekuly, to je provedeno pomocí prokládání bodů rovinou metodou nejmenších čtverců. Od této roviny jsou spočítány vzdálenosti jednotlivých atomů (nejmenší možné vzdálenosti, v pravém úhlu), nejvzdálenější atomy v obou směrech poté představují třetí rozměr molekuly. Válec je tak prořezán dvěma vodorovnými rovinami. Vzhledem k tomu, že molekula je umístěná nejdelší stranou ve směru výšky válce, může vést prokládaná rovina vodorovně s podstavou válce jen v případě symetrické molekuly, a tedy tehdy kdy jsou vzdálenosti stejné na výšku i do hloubky, což vede ke stejným výsledkům třetího rozměru. (Obr. 27)



Obr. 26: Obrázky vykreslené přímo pomocí kódu. Ukazující, jak se počítají parametry velikosti molekuly z pozic jednotlivých atomů. Levý obrázek zobrazuje benzen umístěný v prostoru, spolu s navázanými vodíky, červená přímka vede v prostoru atomů, ta se prokládá body podle metody nejmenších čtverců. Zjištěná přímka určuje střed válce a vzdálenost od přímky k nejvzdálenějšímu bodu představuje poloměr válce (aby byla kružnice opsaná všem atomům). Pravý obrázek ukazuje tu samou situaci, tedy benzen, ale jeho body jsou přečíslovány na umístění na přímce (přímka proložena podle metody nejmenších čtverců). Body na červené přímce jsou určeny pomocí druhé přímky protínající bod atomu spolu s přímkou proloženou podle metody nejmenších čtverců, zároveň přímky jsou na sebe kolmé. Z vypočítaných bodů se zjistí výška válce, jako dva nejvzdálenější body v linii. Na druhém obrázku lze také vidět proložení atomů rovinou pomocí metody nejmenších čtverců, nejvzdálenější body od roviny v obou směrech pak určují nejvyšší velikost molekuly.



Obr. 27: Na prvním obrázku lze vidět šestiúhelník představující benzen, bez atomů vodíku. Atomy jsou proloženy přímkou pomocí metody nejmenších čtverců. Druhý obrázek už ukazuje, jak je pomocí získaných informací vložen válec s co nejmenším poloměrem. Poloměr je získán jako vzdálenost nejodlehlejšího bodu od přímky. Zároveň je získána výška válce jako další rozměr molekuly. Třetí obrázek znázorňuje válec, který ohraničuje body atomů. Na čtvrtém obrázku je vykreslena rovina, která je proložena body atomů, ta je umístěná také pomocí metody nejmenších čtverců. Pátý spolu s šestým obrázkem ukazují, jak jsou vedeny dva řezy válcem, oba jsou rovnoběžné s prokládanou rovinou pomocí metody nejmenších čtverců a ve vzdálenosti nejodlehlejších atomů od roviny ve směru na horu i dolů. Řezem program zjistí hloubku molekuly, což je výhodné především pro plošší molekuly. Popsaným postupem se získají tři rozměry molekuly, vytvořené pomocí (<https://www.geogebra.org/classic?lang=cs>)

Výpočet probíhá pro každou molekulu z *datasetu*, výsledky jako jsou poloměr, výška a hloubka molekuly, spolu s jejím nábojem (získaným pomocí knihovny RDKit) jsou vepsány do *DataFramu* a pak připojeny k *.csv souboru, který byl získán z PubChem. Takový postup se aplikoval z důvodu vyšší výpočetní složitosti vzhledem k velikosti vstupu, u metody nejmenších čtverců a výpočtu nejvzdálenějšího bodu od prokládané přímky. Dále se používá už jen získaný *.csv soubor, u kterého trvá načtení a procházení krátkou dobu.

4.4.2 Průchod molekul tunelem podle vypočítaných vlastností

Do dalšího kódu se načítá soubor tunnels.xml, tento soubor byl získán z nástroje MOLEonline. Pro vyhledání souboru je nutné napsat cestu k jeho umístění. Následně probíhá převod z *.xml souboru do textového řetězce v kódu. Tento postup je až na malé odchylky stejný jako v předchozím případě, při načítání souboru pro vykreslování 2D struktury tunelu. Soubor *.cif slouží jen pro informaci, o jaký se jedná protein, tedy informace PDB identifikátoru. Ze souboru tunnels.xml se vybírají informace především o poloměru tunelu v jednotlivých vzdálenostech (vrstvách kanálu), náboj tunelu a aminokyseliny vyskytující se v jeho okolí.

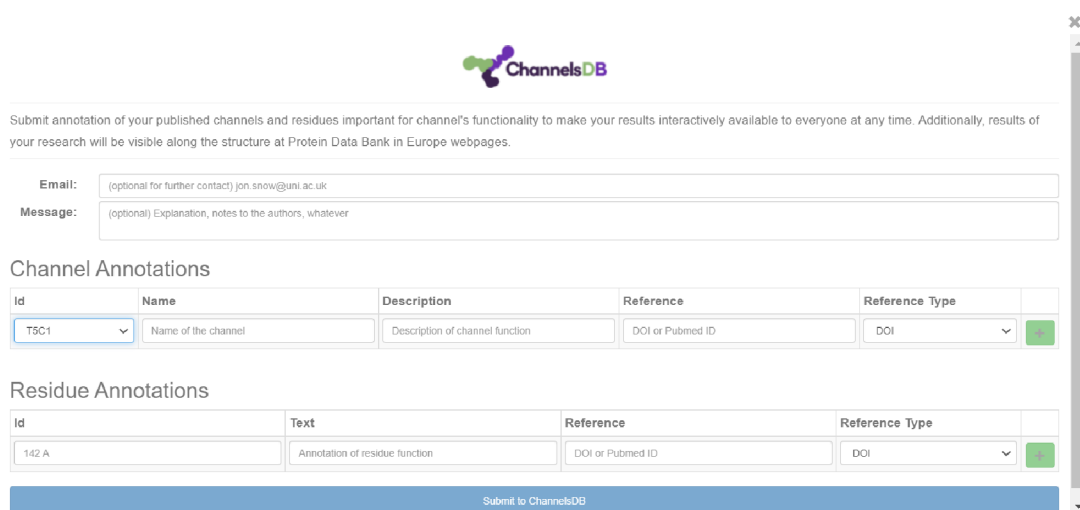
Jako zkoumaná data poslouží získaný *.csv soubor z ligandů a iontů. Ten představuje spojení souborů ligandů a iontů. K filtrování molekul, které by mohly procházet tunelem se používá kritérium, kterým je poloměr molekuly, ten musí být menší než nejužší místo tunelu. A náboj molekuly musí být opačně nabitý než náboj tunelu ve stejné hodnotě, uživatel však může změnit toleranci náboje, kdy při vysoké toleranci, může být náboj zanedbán. Přesněji se prochází pro daný tunel všechny hodnoty *Distances*, dokud jsou menší než daná *End distance*. Pokud je hodnota *Distance* větší než *End distance*, vezmou se všechny hodnoty *Radius*, poloměry doposud zkoumané části tunelu a z nich se vybere nejmenší *radius*, jako požadavek pro průchod tunelem. To samé se odehrává pro hodnotu *FreeRadius*. Molekuly, které splňují tyto kritéria jsou vloženy do *DataFramu*, ten se následně používá pro další vrstvy tunelu, aby se dále vybíralo jen z molekul, co mohou procházet předchozími vrstvami.

4.4 Doplnování informací o tunelech do ChannelsDB

Byl napsán kód v programovacím jazyce Python, který využívá Selenium Webdriver. Pomocí rozhraní API se Python může připojit k prohlížeči prostřednictvím Selenium. Selenium odesílá standardní příkazy Pythonu do různých prohlížečů, a to navzdory rozdílům v designu jejich prohlížeče. Selenium se může spouštět pomocí skriptů Python pro Firefox, Chrome, IE na různých operačních systémech. Tak se spouští jednotlivé stránky ve webovém prohlížeči, pro přidávání informací do databáze.

Po spuštění kód načte stránku MOLEonline (<https://mole.upol.cz/>). Následně čtyřmístný PDB identifikátor vloží do předem načteného MOLEonline do kolonky pro přidávání PDB identifikátoru. Jednotlivé políčka pro psaní textu, nebo tlačítka se vyhledávají podle XPath. Jako další program zmáčkne tlačítko next, tím se přejde do vizualizace struktury proteinu, pro identifikaci kanálů v biomakromolekule následuje zmáčknutí tlačítka `Submit`. Po následném zmáčknutí tlačítka `ChannelsDB Annotate`, jsou nalezené kanály dostupné k jejich anotaci do databáze ChannelsDB.

Po výběru daného kanálu dojde k automatickému vyplnění anotace (Obr. 28), jako je e-mail zadávajícího, Id tunelu, předpokládaný název spolu s popisem, jaké ligandy či ionty by mohly kanálem procházet (výpočet již popsany v předchozím textu). Jako reference je zvolen MOLEonline (nevychází se z popsanych kanálů v člancích). Automaticky jsou také vyplněny informace o aminokyselinách vyskytujících se v okolí tunelu.



Submit annotation of your published channels and residues important for channel's functionality to make your results interactively available to everyone at any time. Additionally, results of your research will be visible along the structure at Protein Data Bank in Europe webpages.

Email: (optional for further contact) jon.snow@uni.ac.uk

Message: (optional) Explanation, notes to the authors, whatever

Channel Annotations

Id	Name	Description	Reference	Reference Type	
T5C1	Name of the channel	Description of channel function	DOI or Pubmed ID	DOI	<input type="checkbox"/>

Residue Annotations

Id	Text	Reference	Reference Type	
142 A	Annotation of residue function	DOI or Pubmed ID	DOI	<input type="checkbox"/>

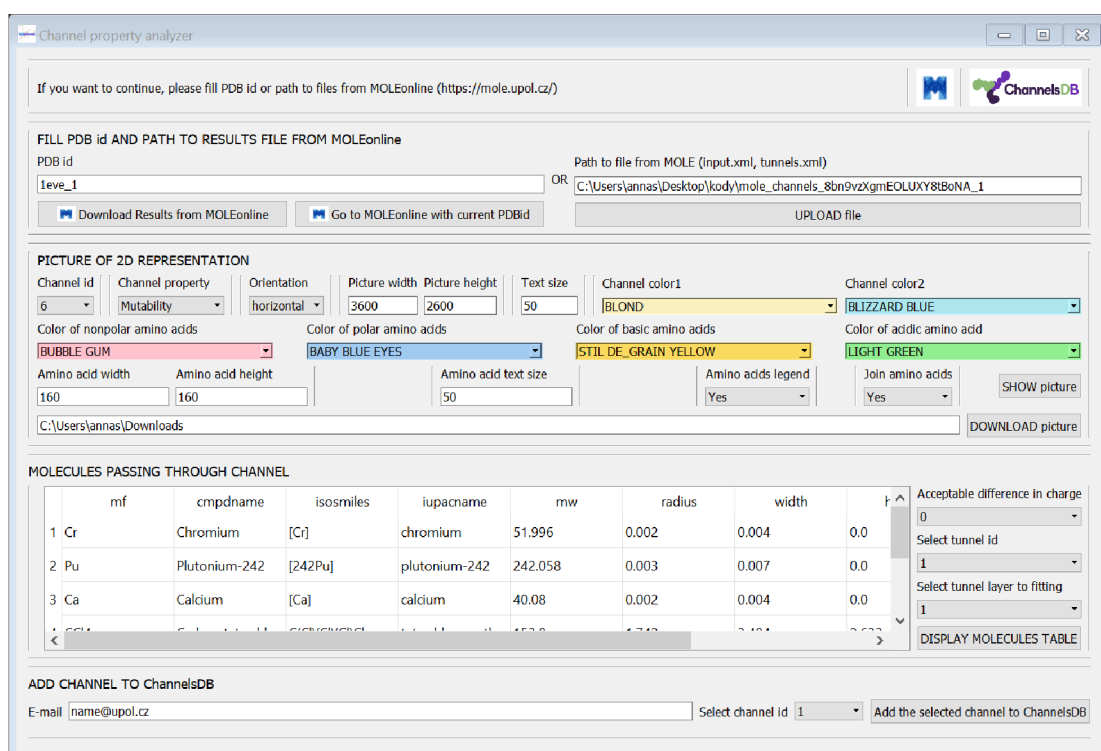
Submit to ChannelsDB

Obr. 28: Stránka MOLEonline po načtení struktury proteinu a zmáčknutí tlačítka ChannelsDB Annotation. Dojde k zobrazení formuláře sloužící k anotaci kanálu do databáze.

4.5 GUI sloužící k ovládání programů

Pro snadnější ovládání naprogramovaných kódů bylo vytvořeno dialogové okno DP.ui (Obr.29) díky nástroji Qt Designer. To umožňuje jednoduché ovládání pomocí klikáním na tlačítka, vybíráním z možností, či vepisování kritérií pro běh programů. Pomocí hlavního souboru main.py za použití knihovny PyQt5 jsou nastavovány funkce jednotlivým položkám z dialogového okna (propojovány s funkcemi v kódu).

Vzhledem k zakládání výsledků na nástroji MOLEonline (<https://mole.upol.cz/>) a databázi ChannelsDB (<https://channelsdb.ncbr.muni.cz/>) jsou do dialogového okna umístěny tlačítka s jejich logy, které slouží k přesměrování na dané webové stránky. Načítání webových stránek je uskutečněno díky Selenium je open-source nástroj, který automatizuje webové prohlížeče. Poskytuje jediné rozhraní, které umožní psát testovací skripty v programovacích jazycích. (<https://www.browserstack.com/selenium>)



Obr. 29: Dialogové okno vytvořené v nástroji Qt Designer, sloužící k ovládání programů.

Uživatelé poskytne možnost měnit vstupní parametry pro funkce, aniž by musel ovládat programovací jazyk. Uživatelé umožní načtení webových stránek ChannelsDB, nebo MOLEonline pomocí kliknutí na tlačítko. Po načtení výsledného souboru z MOLEonline může probíhat vizualizace vybraného kanálu ve 2D reprezentaci s možností změny barev, velikosti obrázku, velikosti textu a to podle uvážení samotného uživatele. Jako další je možnost vypsát do tabulky molekuly, které byly vybrány, že by mohly procházet daným kanálem do dané vrstvy s ohledem na jejich fyzikálně-chemické vlastnosti.

Dialogové okno obsahuje *Line Edit* pro přidávání textu, které slouží k zapsání PDB identifikátoru a následně se může využít pro načtení webové stránky MOLEonline s vizualizací zadané proteinové struktury a jejích kanálů. Další tlačítko slouží ke stažení výsledků výpočtů kanálů ve struktuře, především soubor tunnels.xml je důležitý k dalším možnostem ovládání GUI, soubor *.cif určí PDB identifikátor. Program, po kliknutí na tlačítko, soubor s výsledky automaticky stáhne, extrahuje a přepíše cestu k uloženým souborům v jiném *Line Edit* GUI. Takový soubor je vždy výsledkem výpočtu MOLEonline s defaultními hodnotami. Současně se také načtou identifikátory zjištěných kanálů do *comboboxů* pro možnost výběru. Tento proces probíhá pomocí kódu `load_tunnels_id.py`, který se načítá jako modul v `main.py`. třída `_load_tunnels_id` obsahuje funkce, sloužící k vyhledání a načtení souboru tunnels.xml. Následně v textu souboru vyhledává identifikátory kanálů v proteinu.

Načtení molekuly může probíhat i jiným způsobem a to tak, že uživatel stáhne soubor s výsledky sám a zadá do *Line Edit* cestu ke staženému souboru, se zmáčknutím tlačítka se načítají identifikátory kanálů a současně se mění také vypsaný PDB identifikátor na současně zkoumanou strukturu (informace se získává ze souboru *.cif) ve stažené složce.

Z kódu `main.py` se volají další dva moduly, které obsahují slovníky. Jeden z nich `color_dic.py` obsahuje slovník barev pro knihovnu Arcade, která se používá pro vykreslování kanálu ve 2D reprezentaci, spolu s podbarvením pozadí u jednotlivých položek GUI v *comboboxech*. Další `dictionaries.py` obsahuje slovníky získané pomocí analýzy kanálů uložených v ChannelsDB, ty dále slouží k predikci názvu procházeného kanálu podle jeho vlastností.

Grafickou reprezentaci umožňuje modul jménem `tunnel_arcade.py`, díky němu se vykresluje kanál ve 2D reprezentaci. Pomocí GUI lze měnit kritéria pro výchozí obrázek, jako je velikost obrázku, barva, orientace kanálu (horizontální, vertikální), jaký kanál se vykreslí, velikost vykreslovaných aminokyselin a velikost písma, lze také zobrazit s legendou pro barvy aminokyselin, nebo s propojením aminokyselin v řetězci za sebou. výsledný obrázek se může zobrazit, nebo stáhnout. Výška obrázku (u vertikálního šířka), by měla být ve velikosti, aby se aminokyseliny co jsou nad sebou (vedle sebe) vlezly do vykreslení, jinak dochází k překrytí tunelu vykreslovanou aminokyselinou.

Část `fit_molecule.py` načítá soubor `*.csv`, který obsahuje informace o předem vypočítaných rozměrech ligandů a iontů (zdroj PubChem). Pro vybraný kanál a jeho vrstvu zkoumá, které z molekul by mohly procházet, porovnávají se dva rozměry molekuly s úzkým hrdlem kanálu pro danou vrstvu a s nejmenším *FreeRadius*. Společně s tím lze zvolit kritérium pro akceptovatelný rozdíl náboje molekuly a vrstev kanálu. Při průchodu jednotlivými vrstvami se zkoumají vlastnosti kanálu a porovnávají se s vlastnostmi kanálů z ChannelsDB s cílem najít pro daný kanál nejpřesnější pojmenování. Pokud se takové pojmenování nenajde, pojmenuje se jako ligand channel.

Kód `submit_to_channelsdb.py` složí k automatickému vepsání získaných informací do databáze ChannelsDB. Webová stránka se načítá díky Selenium. K vybranému kanálu se vepisují aminokyseliny v jeho okolí. Zjištěné molekuly, které by mohly kanálem procházet se doplňují do popisu a názvu kanálu, který je předem určen podle podobností již známých kanálů. Jako reference je vepsáno MOLEonline, protože výsledky nevycházejí z publikovaných článků, ale z výzkumu výsledků MOLEonline.

Byl napsán manuál vysvětlující používání dialogového okna k dosažení tížených výsledků (viz. Přílohy `Manual_channel_property_analyzer.pdf`).

5 DISKUZE

Vytvořená vizualizace tunelů ve 2D reprezentaci vytvořená v programovacím jazyce Python za použití knihovny Arcade se principiálně shoduje s vizualizací dostupnou na stránkách MOLEonline (<https://mole.upol.cz/>). Vizualizace v MOLEonline spočívá v programu napsaném v JavaScriptu. Avšak MOLEonline v 2D grafickém zobrazení neobsahuje umístění aminokyselinových zbytků v oblasti proteinových tunelů. Oproti vizualizaci v MOLEonline není vytvořená vizualizace v rámci diplomové práce interaktivní a slouží k zobrazení nebo uložení vykresleného tunelu jako obrázku. Spolu s možnostmi měnit vlastnosti, lze pomocí dialogového okna vytvořit obrázek vyhovující určitým požadavkům. Stejně jako v MOLEonline lze měnit vlastnost, která je zobrazená pomocí přechodu dvou barev, avšak v online nástroji je zařazena ještě barva bílá mezi tyto dvě barvy určující opačné extrémy vybrané vlastnosti. Uživatel volí, který z kanálů si přeje zobrazit, barvu pro zobrazení tunelu, aminokyselin v okolí tunelu, velikost písma. Uživatel volí dvě barvy, které se prolínají podle vlastnosti kanálu, která je předem zvolena (*Mutability, Polarity, LogS, LogD, LogP, Hydropathy, Hydrophobicity, NumNegatives, NumPositives, Ionizable, či Charge*). Současně uživatel může libovolně měnit velikost vykreslovaného obrázku, ale čím větší je obrázek, tím má větší kvalitu (počet pixelů). K možnostem vizualizace slouží také změna orientace tunelu, a to jak horizontálně, tak vertikálně. Lze zvolit, zda budou aminokyseliny, které následují po sobě v řetězci spojeny čarou, sloužící ke snadnému určení na první pohled. Také lze obrázek vykreslit s legendou, či bez ní ta určuje, jaké barvy jsou zvoleny pro jednotlivé třídy aminokyselin (polární, nepolární, bazické a kyselé).

Vzhledem k vlastnostem tunelu a počítání jeho poloměru pomocí vkládání co největší koule vepsané povrchu tunelu, bylo cílem zjistit co nejmenší poloměr molekuly pro průchod tunelem. Oproti jiným metodám, kde vkládají molekulu do tvaru koule, nebo přímo do tvaru daného vazebného místa (Jiang & Kim, 1991) byl použit tvar válce. Válec má podstavu o poloměru nejmenší možné kružnice, která splňuje, že všechny atomy molekuly se nachází uvnitř válce. Výpočet probíhal pomocí metody nejmenších čtverců a molekule byl opsán válec s co nejmenším poloměrem. Následně byly především kvůli plošším molekulám vedeny dva rovinné řezy válcem tak, aby byl prostorový útvar co nejužší a zároveň obsahoval všechny atomy dané molekuly, tím byla získaná hloubka molekuly. Průměr a výška válce představují výšku a šířku. Vypočítané rozměry byly uloženy do *.csv souboru a následně sloužily jako zdroj potenciálních molekul pro

průchod tunelem. Ze souboru tunnels.xml probíhá získávání informace o daném tunelu a z předem připraveného souboru jsou následně vybrány molekuly, které odpovídají velikosti daného tunelu a které by mohly procházet. Oproti tvaru koule má válec výhodu, že vyhledává nejužší místo molekuly pro průchod, avšak nevýhodou by mohly být dlouhé molekuly, které by tunelem nemohly procházet kvůli jeho zakřivení.

Navrhované automatické doplňování anotace do databáze ChannelsDB umožňuje uživateli rychlejší přidávání spolu s ulehčením práce při manuálním zadávání aminokyselin vyskytující se v okolí tunelu z výsledků výpočtů MOLEonline.

6 ZÁVĚR

Diplomová práce spočívala ve vypracování literární rešerše v oblasti typů proteinových tunelů v různých typech proteinů. Praktická část se týkala naprogramování kódů sloužící k vizualizaci kanálů ve 2D reprezentaci společně s aminokyselinovým zastoupením na jeho povrchu. Mělo dojít k analýze látek procházejících tunely a snaha o predikci těchto látek vzhledem ke struktuře proteinového tunelu. Získané informace měly být doplňovány do ChannelsDB za použití nástroje MOLE.

Pro jednotlivé cíle práce byly napsány kódy v programovacím jazyce Python. Jeden z nich slouží pro vykreslení 2D vizualizace zadaného tunelu, společně s aminokyselinovým zastoupením na jeho povrchu. Pro vykreslení je potřeba výsledný soubor z MOLEonline (tunnels.xml). Vzhledem k současnému zobrazení MOLEonline kanálů ve 2D, které neumožňuje vykreslení aminokyselin v okolí kanálu by bylo vítané zakomponovat vytvořený kód a vizualizaci samotnou do MOLEonline.

Z databáze ChannelsDB byly získány informace o uložených kanálech. Především probíhalo zkoumání náboje, délky a poloměru úzkého hrdla kanálu. Z informací byla provedena analýza jak průměrných hodnot, tak všech hodnot u jednotlivých skupin a výsledky byly vyneseny a reprezentovány grafem. Z průměrných, maximálních a minimálních hodnot byly určeny průměrné vlastnosti molekul, které daným kanálem mohou procházet. Došlo k výpočtu rozměrů ligandů a iontů získaných z databáze PubChem. Následně byl napsán kód, který zjišťuje, které molekuly z *datasetu* by mohly procházet vybraným kanálem. Molekuly procházející kanály, hrají důležitou roli k pochopení transportních drah a mohly by představovat potenciální léčiva vstupující do těsné blízkosti ukrytých aktivních míst. Je proto důležité najít molekuly u kterých by byl průchod kanálem možný.

Získané informace o ligandech, které představují možné molekuly pro vstup do tunelu mohou být přidávány do ChannelsDB pomocí napsaného programu. Ten současně předpovídá název daného kanálu na základě již přidaných analyzovaných struktur z ChannelsDB. Automatickým vepisováním do ChannelsDB se usnadní celkový proces a společně s narůstajícím množstvím informací v databázi bude poskytnut lepší přehled o strukturách kanálů.

7 SUMMARY

The diploma thesis relates elaborating of a literature search in the field of types of protein tunnels in various types of protein. The practical part included programming of codes that are used to visualization in 2D representation together with amino acids location along channels. Analysis of molecules passing through channels should have been made together with the effort to predict these molecules considering channel properties. The obtained information should have been inserted into ChannelsDB database.

For each individual aim of this work, codes were written in Python programming language. One of them is used to draw 2D visualization of selected channel together with the amino acids, that are in channels surface. For run code is necessary to download results from MOLEonline (especially tunnels.xml). Due to the actual MOLEonline visualization of channels in 2D, which cannot be able to show amino acids, it would be welcome to incorporate written code and the visualization itself into MOLEonline.

From database ChannelsDB were obtained information about stored channels. Especially research was underway of length, radius bottleneck and charge values of channel. Data were analyzed, like as average values and all values as well. From the information, an analysis of both average values and all values for individual groups was performed and the results were plotted and represented by a graph. Considering the average, maximum and minimum values, the average properties of the molecules that can pass through a given channel were determined. With using written code in Python, shape of ligands and ions were calculated. Subsequently, a predictive code was written to determine molecules from the *dataset* that could pass through the selected channel. Molecules passing through channels are important, they can play significant role in understanding transport paths in organisms, and they could be potential drugs, that can enter close to active site. Therefore prediction of passing molecules is important for drug design.

Obtained information about ligands, which represent possible molecules for passing through channel can be submitted to ChannelsDB using written code as well. Code can predict name of channel, while the prediction is based on structures of previous analysis stored in ChannelsDB. Process is facilitated by automatic pre-filling of information to boxes. Automatic pre-filling information of annotate channels to ChannelsDB could be helpful for increasing number of database entries.

8 LITERATURA

- Agre, P. (2004). Aquaporin water channels (Nobel Lecture). *Angewandte Chemie (International ed. in English)*, 43(33), 4278–4290. <https://doi.org/10.1002/anie.200460804>
- Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Walter, P., & Roberts, K. (2015). *Molecular Biology of the Cell, Sixth Edition*. WW Norton & Co.
- AMCHARTS columns-with-rotated-series <https://www.amcharts.com/demos/column-with-rotated-series/> (29. 4. 2022)
- Batiza, A. F., Rayment, I., & Kung, C. (1999). *Structure (London, England : 1993)*, 7(5), R99–R103. [https://doi.org/10.1016/s0969-2126\(99\)80061-6](https://doi.org/10.1016/s0969-2126(99)80061-6)
- Beckstein, O., Biggin, P. C., Bond, P., Bright, J. N., Domene, C., Grottesi, A., Holyoake, J., & Sansom, M. S. (2003). Ion channel gating: insights via molecular simulations. *FEBS letters*, 555(1), 85–90. [https://doi.org/10.1016/s0014-5793\(03\)01151-7](https://doi.org/10.1016/s0014-5793(03)01151-7)
- Berka, K., Hanák, O., Sehnal, D., Banás, P., Navrátilová, V., , D. J. ., Jaiswal, D., Ionescu, C. - M., Svobodová Vareková, R., Koca, J., & Otyepka, M. (2012). MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic acids research*, 40(Web Server issue), W222–W227. <https://doi.org/10.1093/nar/gks363>
- Brezovksy, J., Babkova, P., Degtjarik, O., Fortova, A., Gora, A., Iermark, I., Rezancova, P., Dvorak, P., Kuta Smatanova, I., Prokop, *et al.* (2016). Engineering a de Novo Transport Tunnel. *ACS Catalysis*, 6(11), 7597–7610. <https://doi.org/10.1021/acscatal.6b02081>
- Brezovsky, J., Chovancova, E., Gora, A., Pavelka, A., Biedermannova, L., & Damborsky, J. (2013). Software tools for identification, visualization and analysis of protein tunnels and channels. *Biotechnology advances*, 31(1), 38–49. <https://doi.org/10.1016/j.biotechadv.2012.02.002>
- Catalytic Site Atlas <http://www.ebi.ac.uk/thornton-srv/databases/CSA> (25. 4. 2022)
- Cid, H., Bunster, M., Canales, M., & Gazitúa, F. (1992). Cid, H., Bunster, M., Canales, M. a Gazitúa, F. (1992) Hydrofobnost a strukturní třídy v proteinech. *Protein Eng. Des. Sel.*, (5), 373–375.
- Coleman, R. G., & Sharp, K. A. (2009). Finding and characterizing tunnels in macromolecules with application to ion channels and pores. *Biophysical journal*, 96(2), 632–645. <https://doi.org/10.1529/biophysj.108.135970>
- Eisenberg, B. (1998). Ionic channels in biological membranes: natural nanotubes described by the drift-diffusion equations. *Sixth International Workshop on Computational Electronics. Extended Abstracts*, (Cat. No.98EX116), 78-79. <https://doi.org/10.1109/IWCE.1998.742714>
- Geogebra <https://www.geogebra.org/classic?lang=cs> (1. 4. 2022)
- Gora, A., Brezovsky, J., & Damborsky, J. (2013). Gates of enzymes. *Chemical reviews*, 113(8), 5871–5923. <https://doi.org/10.1021/cr300384w>
- Gouaux, E., & Mackinnon, R. (2005). Principles of selective ion transport in channels and pumps. *Science (New York, N.Y.)*, 310(5753), 1461–1465. <https://doi.org/10.1126/science.1113666>
- Gullingsrud, J., & Schulten, K. (2004). Lipid bilayer pressure profiles and mechanosensitive channel gating. *Biophysical journal*, 86(6), 3496–3509. <https://doi.org/10.1529/biophysj.103.034322>
- Hedfalk, K., Tömroth-Horsefield, S., Nyblom, M., Johanson, U., Kjellbom, P., & Neutze, R. (2006). Aquaporin gating. *Current opinion in structural biology*, 16(4), 447–456. <https://doi.org/10.1016/j.sbi.2006.06.009>
- Hollander, M., Rasp, D., & Helms, V. (2021). ProPores2: Web Service and Stand-Alone Tool for Identifying, Manipulating, and Visualizing Pores in Protein Structures. *J. Chem. Inf. Model.*, 61(4), 1555–1559. <https://doi.org/10.1021/acs.jcim.1c00154>

- Holliday, G. L., Mitchell, J. B. O., & Thornton, J. M. (2009). Understanding the functional roles of amino acid residues in enzyme catalysis. *Journal of molecular biology*, 390(3), 560–577. <https://doi.org/10.1016/j.jmb.2009.05.015>
- Chaloupková, R., Sýkorová, J., Prokop, Z., Jesenská, A., Monincová, M., Pavlová, M., Tsuda, M., Nagata, Y., & Damborský, J. (2003). Modification of activity and specificity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 by engineering of its entrance tunnel. *The Journal of biological chemistry*, 278(52), 52622–52628. <https://doi.org/10.1074/jbc.M306762200>
- ChannelsDB <https://webchemdev.ncbr.muni.cz/ChannelsDB/> (8. 2. 2022)
- ChannelsDB methods <https://webchemdev.ncbr.muni.cz/ChannelsDB/methods.html> (27. 3. 2022)
- Chemicalize www.chemicalize.org (3. 3. 2022)
- Jiang, F., & Kim, S. H. (1991). "Soft docking": matching of molecular surface cubes. *Journal of molecular biology*, 79–102. [https://doi.org/10.1016/0022-2836\(91\)90859-5](https://doi.org/10.1016/0022-2836(91)90859-5)
- Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T., & MacKinnon, R. (2003). X-ray structure of a voltage-dependent K⁺ channel. *Nature*, 423(6935), 33–41. <https://doi.org/10.1038/nature01580>
- Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences : CABIOS*, 8(3), 275–282. <https://doi.org/10.1093/bioinformatics/8.3.275>
- Kingsley, L. J., & Lill, M. A. (2015). Substrate tunnels in enzymes: structure-function relationships and computational methodology. *Proteins*, 83(4), 599–611. <https://doi.org/10.1002/prot.24772>
- Klvana, M., Pavlova, M., Koudelakova, T., Chaloupkova, R., Dvorak, P., Prokop, Z., Stsiapanava, A., Kutý, M., Kuta-Smatanova, I., Dohnalek, J., Kulhanek, P., Wade, R. C., & Damborsky, J. (2009). Pathways and Mechanisms for Product Release in the Engineered Haloalkane Dehalogenases Explored Using Classical and Random Acceleration Molecular Dynamics Simulations. *Journal of Molecular Biology*, 392(5), 1339-1356. <https://doi.org/10.1016/j.jmb.2009.06.076>
- Konc, J., Lešnik, S., & Janežič, D. (2015). Modeling enzyme-ligand binding in drug discovery. *J Cheminform*, 7, 48. <https://doi.org/10.1186/s13321-015-0096-0>
- Kozlíková, B., Andres, F., & Sochor, J. (2007). Visualization of tunnels in protein molecules. *Conference AFRIGRAPH AFRIGRAPH: International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, (5th), 111–118. <https://doi.org/10.1145/1294685.1294704>
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1), 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- Libraries in Python <https://www.geeksforgeeks.org/libraries-in-python/> (24. 3. 2022)
- LiteMol <https://www.litemol.org/> (25. 4. 2022)
- Mariano, D. C. B., Santos, L. H., Machado, K. D. S., Werhli, A. V., de Lima, L. H. F., & de Melo-Minardi, R. C. (2019). A Computational Method to Propose Mutations in Enzymes Based on Structural Signature Variation (SSV). *International journal of molecular sciences*, 20(2), 333. <https://doi.org/10.3390/ijms20020333>
- Marques, S. M., Brezovsky, J., & Damborsky, J. (2016). Role of tunnels, channels and gates in enzymatic catalysis. In A. Svedsen (Ed.), *Understanding Enzymes: Function, Design, Engineering and Analysis* (p. 40). PanStanford Publishing.

- https://loschmidt.chemi.muni.cz/peg/wp-content/uploads/2016/05/2016_Marques_Role-of-tunnels-channels-and-gates-in-enzymatic-catalysis.pdf
- Marques, S. M., Daniel, L., Buryška, T., Prokop, Z., Brezovsky, J., & Damborsky, J. (2017). Enzyme Tunnels and Gates As Relevant Targets in Drug Design. *Medicinal research reviews*, 37(5), 1095–1139. <https://doi.org/10.1002/med.21430>
- Miles, E. W., Rhee, S., & Davies, D. R. (1999). The molecular basis of substrate channeling. *Journal of biological chemistry*, 274(18), 12193–12196.
- MOLEonline <https://mole.upol.cz> (3. 5. 2021)
- MOLEonline documentation <https://mole.upol.cz/documentation/> (8. 2. 2022)
- MOLEonline FAQ <https://mole.upol.cz/FAQ> (9. 2. 2022)
- MOLEonline method <https://mole.upol.cz/method> (28. 3. 2022)
- Moreau, C. J., Dupuis, J. P., Revilloud, J., Arumugam, K., & Vivaudou, M. (2008). Coupling ion channels to receptors for biomolecule sensing. *Nature nanotechnology*, 3(10), 620–625. <https://doi.org/10.1038/nnano.2008.242>
- Niemeyer, B. A., Mery, L., Zaward, C., Suckow, A., Monje, F., Padro, L. A., Strüher, W., Flockerzi, V., & Hoth, M. (2001). Ion channels in health and disease: 83rd Boehringer Ingelheim Fonds International Titisee Conference. *EMBO reports*, 2(7), 568–573. <https://doi.org/10.1093/embo-reports/kve145>
- PDB <https://www.wwpdb.org/> (1.12.2021)
- Petrek, M., Kosinová, P., Koca, J., & Otyepka, M. (2007). MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure (London, England : 1993)*, 15(11), 1357–1363. <https://doi.org/10.1016/j.str.2007.10.007>
- Petrek, M., Otyepka, M., Banáš, P., Kosinová, P., Koca, J., & Damborský, J. (2006). CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC bioinformatics*, 7, 316. <https://doi.org/10.1186/1471-2105-7-316>
- Pilgrim, M. (2012). *Kniha: Ponořme se do Python(u) 3*. CZ.NIC.
- Pivetti, C. D., Yen, M. -R., Miller, S., Busch, W., Tseng, Y. -H., Booth, I. R., & H Saier Jr, M. (2003). Two families of mechanosensitive channel proteins. *Microbiology and molecular biology reviews: MMBR*, 67(1), 66–85. <https://doi.org/10.1128/MMBR.67.1.66-85.2003>
- Pravda, L., Berka, K., Svobodová Vařeková, R., Sehnal, D., Banáš, P., Laskowski, R. A., Koča, J., & Otyepka, M. (2014). Anatomy of enzyme channels. *BMC bioinformatics*, 15(1), 379. <https://doi.org/10.1186/s12859-014-0379-x>
- Pravda, L., Sehnal, D., Svobodová Vařeková, R., Navrátilová, V., Toušek, D., Berka, K., Otyepka, M., & Koča, J. (2017). ChannelsDB: database of biomacromolecular tunnels and pores. *Nucleic Acids Research*, 46, D399–D405. <https://doi.org/10.1093/nar/gkx868>
- Pravda, L., Sehnal, D., Svobodová Vařeková, R., Navrátilová, V., Toušek, D., Berka, K., Otyepka, M., & Koca, J. (2018a). ChannelsDB: database of biomacromolecular tunnels and pores. *Nucleic acids research*, 46(D1), D399–D405. <https://doi.org/10.1093/nar/gkx868>
- Pravda, L., Sehnal, D., Toušek, D., Navrátilová, V., Bazinger, V., Berka, K., Svobodová Vařeková, R., Koca, J., & Otyepka, M. (2018b). MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic acids research*, 46(W1), W368–W373. <https://doi.org/10.1093/nar/gky309>
- Prokop, Z., Gora, A., Brezovsky, J., Chaloupková, R., Stepankova, V., & Damborsky, J. (2012). Engineering of Protein Tunnels: Keyhole-Lock-Key Model for Catalysis by Enzymes with Buried Active Sites. In S. Lutz & U. T. Bornscheuer (Eds.), *Protein Engineering Handbook* (Volume 3, p. 42). Wiley-VCH Verlag.
- PubChem <https://pubchem.ncbi.nlm.nih.gov/> (10. 11. 2021)
- PyQt5 tutorial <https://build-system.fman.io/pyqt5-tutorial> (21. 2. 2022)

Python Arcade <https://realpython.com/arcade-python-game-framework/> (12. 2. 2022)

Python Matplotlib <https://matplotlib.org/3.5.1/tutorials/introductory/pyplot.html> ml (25. 3. 2022)

Python os module <https://www.geeksforgeeks.org/os-module-python-examples/> (26. 2. 2022)

Python Pandas <https://naucse.python.cz/lessons/intro/pandas/> (24. 3. 2022)

Python xml.dom.minidom — Minimal DOM implementation <https://docs.python.org/3/library/xml.dom.minidom.html> (26. 2. 2022)

Python xml.etree.ElementTree — The ElementTree XML API <https://docs.python.org/3/library/xml.etree.elementtree.html> (26. 2. 2022)

Qt Designer <https://doc.qt.io/qt-5/qtdesigner-manual.html> (10.3.2022)

Raushel, F. M., Thoden, J. B., & Holden, H. M. (2003). Enzymes with molecular tunnels. *Accounts of chemical research*, 36(7), 539–548. <https://doi.org/10.1021/ar020047k>

RSCB PDB <https://www.rcsb.org/> (8. 2. 2022)

Sanson, B., Colletier, J. -P., Xu, Y., Lang, P. T., , H. J. ., Jiang, H., Silman, I., Sussman, J. L., & Weik, M. (2011). Backdoor opening mechanism in acetylcholinesterase based on X-ray crystallography and molecular dynamics simulations. *Protein science : a publication of the Protein Society*, 20(7), 1114–1118. <https://doi.org/10.1002/pro.661>

Sasaki, S. (2012). Aquaporin 2: from its discovery to molecular structure and medical implications. *Molecular aspects of medicine*, 33(5-6), 535–546. <https://doi.org/10.1016/j.mam.2012.03.004>

Sehnal, D., Svobodová Vařeková, R., Berka, K., Pravda, L., Navrátilová, V., Banáš, P., , C. -M. I. ., Ionescu, C. -M., , J. K., Otyepka, M., & Koča, J. (2013). MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *Journal of cheminformatics*, 5(1), 39. <https://doi.org/10.1186/1758-2946-5-39>

Selenium <https://www.browserstack.com/selenium> (2. 1. 2022)

Singh, S., & Anand, R. (2021). Tunnel Architectures in Enzyme Systems that Transport Gaseous Substrates. *ACS omega*, 6(49), 33274–33283. <https://doi.org/10.1021/acsomega.1c05430>

Stourac, J., Vavra, O., Kokkonen, P., Filipovic, J., Pinto, G., Brezovsky, J., Damborsky, J., & Bednar, D. (2019). Caver Web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport. *Nucleic Acids Research*, (Volume 47), W414–W422. <https://doi.org/10.1093/nar/gkz378>

Summerfield, M. (2013). *Python 3: Výukový kurz* (Dotisk prvního vydání). v. The Python Standard Library <https://docs.python.org/3/library/> (16. 2. 2022)

UniProt <https://www.uniprot.org/blast/> (25. 4. 2022)

Walz, T., Hirai, T., Murata, K., Heymann, J. B., Mitsuoka, K., Fujiyoshi, Y., Smith, B. L., Agre, P., & Engel, A. (1997). The three-dimensional structure of aquaporin-1. *Nature*, 387(6633), 624–627. <https://doi.org/10.1038/42512>

Wente, S. R., & Rout, M. P. (2010). The nuclear pore complex and nuclear transport. *Cold Spring Harbor perspectives in biology*, 2(10), a000562. <https://doi.org/10.1101/cshperspect.a000562>

What is Python? Executive Summary <https://www.python.org/doc/essays/blurb/> (31. 3. 2022)

Zhang, X., , Van de Moortele, P. -F., & Schmitter, S. (2013). Complex B1 mapping and electrical properties imaging of the human brain using a 16-channel transceiver coil at 7T. *Magnetic resonance in medicine*, 69(5), 1285–1296. <https://doi.org/10.1002/mrm.24358>

Zimmerman, J. M., Eliezer, N., & Simha, R. (1968). The characterization of amino acid sequences in proteins by statistical methods. *Journal of theoretical biology*, 21(2), 170–201. [https://doi.org/10.1016/0022-5193\(68\)90069-6](https://doi.org/10.1016/0022-5193(68)90069-6)

9 SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

Å	Ångström (angstrom) jednotka délky, $1 \text{ Å} = 1 \times 10^{-10} \text{ m}$, používají se v samotném kódu, a textovém souboru.
AQP0	Lidský akvaporin, byl úspěšně demonstrován jako marker erytroidní diference.
AQP1	AQP1 je široce exprimovaný vodní kanál.
AQP2	Aquaporin-2 se nachází v apikálních buněčných membránách hlavních buněk sběrného kanálku ledvin.
AQP4	Protein vodního kanálu.
AqpM	Bakteriální akvaporin M.
AQPZ	Bakteriální akvaporin Z.
BLOSUM	V bioinformatice je matrice BLOSUM substituční matrice používaná pro sekvenční zarovnání proteinů. Matrice se používají k hodnocení srovnání mezi evolučně odlišnými proteinovými sekvencemi.
C1C	Chloridový kanál kosterního svalstva.
CSA	Catalytic Site Atlas.
CSV	Jednoduchý souborový formát určený pro výměnu tabulkových dat.
GlpF	Aquaglyceroporin katalyzuje transmembránovou difúzi glycerolu a určitých lineárních vícemocných alkoholů v <i>Escherichia coli</i> .
GsPutA	Substrátový kanál spojující dvě aktivní místa prostřednictvím hydrolyzační dutiny.
GUI	Grafické uživatelské rozhraní (Graphic User Interface).
I/O	Input/Output.
ID	Identifikátor.
JSON	(JavaScript Object Notation) je způsob zápisu dat nezávislý na počítačové platformě, určený pro přenos dat, která mohou být organizována v polích nebo agregována v objektech.
M2	Proteinový pór.
mmCIF	Flexibilní a rozšiřitelný formát tag-value pro reprezentaci makromolekulárních strukturních dat.
MS	Mechanosenzitivní.
MscL	Rodina mechanosenzitivních kanálů.
MscS	Rodina mechanosenzitivních kanálů.

NaN	Not a Number.
OS	Operační systém.
PAM	Bodově akceptovaná mutace (PAM) je nahrazení jedné aminokyseliny v primární struktuře proteinu jinou jedinou aminokyselinou, která je přijímána procesy přirozeného výběru.
PDB ID	Identifikátor struktury z PDB.
PDF	(Portable Document Format), zajišťuje stejné zobrazení na všech zařízeních.
RCSB PDB	Protein Data Bank je primární databáze 3-D struktur makromolekul jako např. proteinů, nukleových kyselin.
RGB	Barevný model, který používá zobrazení barvy podle zastoupení poměru tří hlavních (červená, zelená a modrá).
RMSF	(Root mean square fluctuation), střední kvadratická fluktuace.
S6	Proteinový pór K ⁺ .
SoPIP2;1	Rostlinné akvaporiny SoPIP2;1 nacházející se např. v plazmatické membráně listů špenátu.
TMS	Transmembránový segment.
VMD	(Visual Molecular Dynamics) je počítačový program pro molekulární modelování a vizualizaci.
XML	(Extensible Markup Language) jde o obecný značkovací jazyk.
ZIP	Formát pro kompresi a archivaci dat.

10 PŘÍLOHY

Elektronické přílohy:

Anna_Špačková_příloha.zip – obsahuje:

- složka Channel_property_analyzer:
 - main.py – Hlavní kód importující ostatní moduly a přiřazující vlastnosti položkám z GUI.
 - color_dict.py – Program obsahující slovník barev z knihovny Arcade.
 - fit_molecule.py – Kód sloužící k selekci molekul, které by podle vlastností mohly procházet vybraným kanálem.
 - load_tunnels_id.py – Načítá vlastnosti tunelu pro výběr v dialogovém okně.
 - tunnels_arcade.py – Vykresluje kanály pomocí informací z výsledného souboru tunnel.xml získaného na MOLEonline.
 - DP.ui – Dialogové okno, vytvořené v nástroji Qt Designer, sloužící k ovládání prvků.
 - dictionaries.py – Soubor obsahující slovníky vlastností získaných analýzou informací dostupných na ChannelsDB, podle kterých se určuje vlastnost zkoumaného kanálu.
 - ALL_LigandsIons.csv – Soubor obsahující tabulku ligandů a iontů z PubMed společně s vypočítanou velikostí.
 - submit_to_channelsdb.py – Program, který umožňuje automaticky vypsát získané informace do databáze ChannelsDB.
 - MOLEonline_logo.png, logo.png, ChannelsDB_logo.png, downloads.png – Obrázky, které slouží k zobrazení na tlačítkách v dialogovém okně, nebo jako ikona.
 - Count_MOLECULE_SHAPE.ipynb – Program sloužící k výpočtu rozměrů molekul (výsledkem *.csv).
 - Manual_channel_property_analyzer.pdf – Manuál k ovládání dialogového okna.
 - Složka mole_channels_8bn9vzXgmEOLUXY8tBoNA_1 – obsahující stažené výsledky (*Results*) z MOLEonline pro strukturu 1EVE.

- requirements.txt – Textový soubor obsahující používané knihovny, pomocí, kterého lze jednoduše knihovny načíst, před používáním nástroje.
- chromedriver.exe – Selenium ChromeDriver.
- Složka Analyza_ChannelsDB
 - info_from_channelsDB.ipynb – Kód sloužící k získání uložených informací z Channels DB.
 - Analysis_channelsDB.ipynb – Program, který zpracovává získané informace z ChannelsDB.
- Složka Analyza_zkoumanych_MOLEonline
 - Složka csv_data – Výsledná data analýzy zkoumaných struktur v MOLEonline.
 - Složka inputs
 - složka codes – Kódy sloužící k analýze souborů input.xml.
 - Složka data – Výsledné datové soubory získané pomocí programů.
 - Složka pictures – Vytvořené grafy pomocí programů.
 - Složka tunnels
 - Složka codes – Kódy, sloužící k získávání informací z tunnels.xml souborů.
 - Složka data - Výsledné datové soubory získané pomocí programů.
 - Složka pictures – Získané grafy pomocí programů.