

Pro modelování struktury jsem si vybrala lidský protein anoctamin-1 (Q5XXA6). Jde o transmembránový protein z rodiny TMEM16 proteinů, který funguje jako vápníkem aktivovaný chloridový kanál. Sekvence tohoto proteinu je známá u několika organismů včetně člověka, avšak jeho struktura byla dosud experimentálně určena pouze pro *mus musculus*. Motivací pro výběr tohoto proteinu je porovnání struktury lidského proteinu s myším (Q8BHY3). Sekvenční identita lidského a myšího proteinu je přibližně 90%, jedinou větší odlišností je 26 bází dlouhá inzerce na pozici 475 (lidský protein je o 26 bází delší). Zajímá mne, jestli tato inzerce nějak významněji ovlivní strukturu proteinu.

Sekvence proteinu, se kterou budeme dále pracovat, je získána z UniProtKB databáze. Jsou zde uváděny 3 izoformy vznikající alternativním sestřihem. Budeme pracovat s izoformou č. 1 (Q5XXA6-1), která je uváděna jako kanonická. Jedná se o homodimer, predikován bude pouze monomer.

Metody:

Jelikož existuje homolog proteinu s velkou sekvenční identitou, dává smysl použít homologní modelování. Jako první jsem si vybrala program ROSETTA. Jde o plně automatizovanou implementaci metody Rosetta. Tuto metodu jsem si vybrala zejména proto, že dosahuje velmi dobrých výsledků v soutěžích CASP. Vstupní sekvence se parsuje na 'domény' metodou 'Ginzu', tím se detekují regiony homologní k nějakému templátu. Regiony, ke kterým nebyl přiřazen templát, mohou být dále predikovány *de novo*. V celém procesu se kombinuje velké množství metod a přístupů - BLAST, PSI-BLAST, HMMER, MSA, Mammoth algoritmus, ... To pokládám za velkou výhodu. Také je výhodou to, že díky dělení na domény může být pro každou část proteinu použit jiný templát nebo i jiná metoda.

Predikce ROSETTOU trvala poměrně dlouho – 5 dní se parsovaly domény a dalších cca 30 hodin zabralo stavění modelu. Proto jsem pro srovnání použila další dva automatické servery, RaptorX a Phyre2, u nichž predikce trvala podstatně kratší dobu – přibližně půl dne.

Nevýhodou všech těchto metod je to, že nelze zadat vlastní alignment. V tomto případě ale myslím, že to takový problém není, protože sekvenční identita je velká a alignment by neměl být moc komplikovaný.

Výsledky:

ROSETTA predikovala 1 doménu pokrývající celou sekvenci. Pro tuto doménu bylo vytvořeno 5 různých modelů. Nebylo poskytnuto žádné hodnocení kvality modelů, pro porovnání jsem tedy zvolila všech 5.

Phyre2 vytvořil 20 modelů, z nichž 7 bylo vyhodnoceno jako spolehlivých a 3 z nich měly vysokou sekvenční identitu (kolem 90%) s templátem. Pokrytí těchto tří modelů bylo kolem 70%, nebyla tedy jako u ROSETTY vymodelována celá sekvence. Pro první model sloužil jako templát 5OC9, což je lidský anoctamin-10. Druhý model byl modelován podle myšího anoctaminu-1 (6BGI). 5NL2, další struktura myšího anoctaminu-1, byla použita jako templát pro třetí model; tato struktura má ale velmi špatné rozlišení (6,6 Å). Pro porovnání jsem zvolila pouze první a druhý model.

Výstupem programu RaptorX byl pouze jeden model. Jako templát byl použit myší anoctamin-1 6BGI.

Modely jsem porovnávala s experimentálně určenou strukturou myšího anoctaminu-1. Pro porovnání jsem zvolila strukturu 5OYB, která měla nejlepší rozlišení, bohužel i tak poměrně špatné (3.75 Å). V PDB souboru chybí prvních 116 bází a hodně smyček, ty jsem tedy porovnat nemohla. Zaměřila jsem se na místo, kde je v lidském proteinu inzerce, a na místo, kde je vázán vápník.

Tři helixy, mezi kterými je vazebné místo pro vápník, byly u všech modelů kromě jednoho stejné a shodné s myším proteinem. U prvního modelu z Phyre2 byl jeden z těchto helixů vymodelován jinak.

Zdá se, že inzerce v lidském proteinu nezpůsobí významnou odlišnost obou proteinů. U většiny modelů je zde navíc smyčka, například u druhého modelu z ROSETTY je zde smyčka a malý alfa helix.

Modely ROSETTY se nejvíce liší prvních cca 200 aminokyselin, ale hlavní část s transmembránovými helixy je velice podobná myšímu proteinu.

Model vytvořený programem RaptorX se odlišuje nejvíce. Dlouhá smyčka v oblasti 454-514 nejspíš bude chyba v modelu. Nesedí dobře ani některé transmembránové helixy, které u ostatních modelů byly správně.

Podle validace modelů pomocí programu WhatCheck se zdá, že modely získané programy Phyre2 a RaptorX mají podstatně horší kvalitu. Jejich Z-skóre se odchylují od normálu ve více než polovině hodnocených kritérií. Modely z ROSETTY mají některé vlastnosti neobvyklé, ale celkově je hodnocení výrazně lepší. Nejlepší se zdá hodnocení třetího modelu z ROSETTY. Zdá se mi ale, že tento model bude mít chybu od pozice cca 950 do konce. Je zde dlouhá smyčka s alfa helixem na konci, která „vyčnívá“ ze struktury.

Pokud bych měla shrnout výsledek predikce, tak na základě modelů lze usoudit, že lidský anoctamin-1 bude mít nejspíš stejnou funkci jako u myši. Vazebné místo pro vápník a počet i přibližné pozice transmembránových alfa helixů budou stejné a ve struktuře by neměly být zásadní odlišnosti. Co se týče kvality modelů, je zde velký prostor pro zlepšení. Samotný templát neměl dobré rozlišení, a modely tak nemůžou být velmi spolehlivé. Na druhou stranu, sekvenční identita byla nadprůměrně vysoká a pro to, abychom si udělali alespoň přibližný obrázek o tom, jak bude protein vypadat, by tyto modely měly postačovat. Pokud bych měla vybrat jeden model, který se mi zdá nejlepší, zvolila bych model z ROSETTY, nejspíše ten první.