

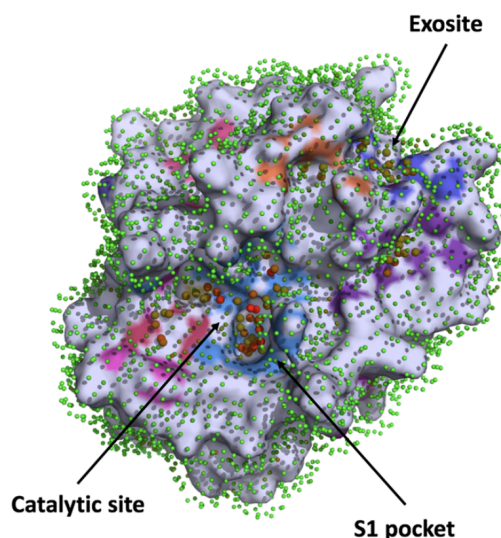
Jak software z Matfyzu napomáhá v boji (nejen) s onemocněním COVID-19

Počítačové metody jsou dnes nedílnou součástí procesu hledání nových léčiv, zvláště pak v tzv. racionálním vývoji léčiv. Zde je snaha identifikovat funkční sloučeniny s ohledem na znalost struktur a funkcí, molekulárních cílů, které jsou v systému narušovány, nebo které naopak zdravé procesy narušují. Libovolný proces lze na molekulární úrovni chápat jako provázaný systém interakcí, kdy jedna molekulární interakce umožní další, tvořící biochemickou kaskádu. Chápeme-li smysl jednotlivých interakcí, můžeme navrhovat molekuly, které by kaskádu přerušili a inhibovali tak daný proces. V případě virů může jít o proces vniknutí viru do buňky, nebo proces replikace v buňce. Zde je cílem identifikace malých molekul, tzv. ligandů, které se navážou na konkrétní molekulu a tím buď znemožní samotný proces replikace viru nebo znemožní interakci s hostitelským organismem, což ve výsledku vede k deaktivaci viru v organismu.

Ligandy lze nejlépe identifikovat, když známe strukturu cílové molekuly, typicky proteinu, a optimálně přímo strukturu aktivního místa, kam by se daný ligand mohl navázat. Právě v tomto procesu pomáhá nástroj [P2Rank](#) [1] vyvinutý na Matematicko-fyzikální fakultě UK ve spolupráci s Přírodovědeckou fakultou UK. Tento nástroj je schopen pro vstupní proteinovou strukturu identifikovat místa na povrchu proteinu, které jsou s vysokou pravděpodobností schopny vázat nespécifikovanou malou molekulu. Znalost těchto tzv. aktivních míst, pak umožňuje procházet knihovny malých molekul a počítačově testovat, zda konkrétní molekula je schopná se v identifikovaném aktivním místě udržet, a tedy hrát roli ligandu, který by mohl modulovat funkci daného proteinu. P2Rank je nástroj založen na strojovém učení, který se dokáže z velké trénovací sady známých protein-ligand komplexů naučit, jaké vlastnosti typicky mají místa na povrchu proteinů, které jsou schopny vázat ligand a tuto znalost pak aplikuje v procesu identifikace aktivních míst. Nástroj je k dispozici jako webový server [PrankWeb](#) [2], kdy uživatel může zadat cílový protein a za několik okamžiků vidět predikované aktivní místa přímo na zadané struktuře, nebo jako samostatná aplikace umožňující predikci ve velkém měřítku, např. na celém proteomu příslušného organismu. P2Rank je volně přístupný, ve srovnání s dalšími open access nástroji tohoto typu vede k výrazně lepším predikcím a je lehce použitelný. To umožňuje jeho okamžité využití akademickými týmy z celého světa, jak lze pozorovat na příkladu hledání látek aktivních proti viru SARS-CoV-2.

Koronaviry jsou RNA viry, které nesou svoji genetickou informaci v dlouhém vlákně RNA z něž syntetizují dlouhý polyprotein, který je následně štěpen tzv. proteázami na jednotlivé proteiny. Některé z těchto proteinů pak tvoří transkripčně-replikační komplex, jiné formují virion, obal, který chrání genetickou informaci viru.

P2Rank byl využit např. ve studiích identifikujících aktivní místa na hlavní proteáze viru [3, 4] ve snaze identifikovat ligandy, které by se do těchto míst mohly vázat a tak znemožnit štěpení virového polyproteinu do jednotlivých



konstituentních proteinů. Další studie [5,6] využily P2Rank k mapování tzv. spike proteinu, který SARS-CoV-2 používá k navázání a vstupu do hostitelské buňky. V dalších studiích [7,8] byl P2Rank využit pro pochopení způsobu interakce a navržení ligandů cílící aktivní místa proteinů replikačně-transkripčního komplexu. Takové sloučeniny by pak mohly vést k zabránění replikace viru v buňce. Všechny zmíněné cíle pak byly s pomocí P2Ranku studovány a použity firmou IBM při vývoji frameworku CogMol [9]. Tento framework si klade za cíl využít metody hlubokého strojového učení pro rychlou identifikaci potenciálních aktivních sloučenin při výskytu nových epidemií, přičemž SARS-CoV-2 je první taková epidemie.

P2Rank je ovšem obecný nástroj a jeho aplikace se neomezuje pouze na SARS-CoV-2 jak ukazují další studie, které P2Rank používají. To může podtrhnout i spolupráce s evropskou proteinovou databankou PDBe, největším repozitářem struktur na světě sídlícím v Evropském bioinformatickém institutu EBI, s kterým P2Rank sdílí predikce o zhruba 130.000 strukturách uložených v PDB [10].

1. [Krivák, Radoslav, and David Hoksza. "P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure." *Journal of cheminformatics* 10.1 \(2018\): 39.](#)
2. [Jendele, Lukas, et al. "PrankWeb: a web server for ligand binding site prediction and visualization." *Nucleic acids research* 47.W1 \(2019\): W345-W349.](#)
3. [Liang, Julia, et al. "Site mapping and small molecule blind docking reveal a possible target site on the SARS-CoV-2 main protease dimer interface." *Computational Biology and Chemistry* \(2020\): 107372](#)
4. [Singh, Natesh, et al. "Structure-based drug repositioning over the human TMPRSS2 protease domain: search for chemical probes able to repress SARS-CoV-2 Spike protein cleavages." *European Journal of Pharmaceutical Sciences* 153 \(2020\): 105495](#)
5. [Onawole, Abdulmujeeb T., et al. "COVID-19: CADD to the rescue." *Virus Research* \(2020\): 198022.](#)
6. [Shekhar, Nishant, et al. "In Silico Structure-Based Repositioning of Approved Drugs for Spike Glycoprotein S2 Domain Fusion Peptide of SARS-CoV-2: Rationale from Molecular Dynamics and Binding Free Energy Calculations." *mSystems* 5.5 \(2020\).](#)
7. [Bandyopadhyay, Suritra, et al. "Polypharmacology of some medicinal plant metabolites against SARS-CoV-2 and host targets: Molecular dynamics evaluation of NSP9 RNA binding protein." \(2020\).](#)
8. [Pandey, Preeti, et al. "Insights into the biased activity of dextromethorphan and haloperidol towards SARS-CoV-2 NSP6: in silico binding mechanistic analysis." *Journal of Molecular Medicine* \(2020\): 1-15.](#)
9. [Chenthamarakshan, Vijil, et al. "Target-specific and selective drug design for covid-19 using deep generative models." *arXiv preprint arXiv:2004.01215* \(2020\).](#)
10. ["PDBe-KB: a community-driven resource for structural and functional annotations." *Nucleic Acids Research* 48, no. D1 \(2020\): D344-D353.](#)