



**VYSOKÁ ŠKOLA
CHEMICKO-TECHNOLOGICKÁ
V PRAZE**

Konferenční sborník ENBIK2016

Editoři:

Petr Čech, Daniel Svozil

Praha 2016

Publikace neprošla jazykovou ani odbornou úpravou.
Za obsah příspěvků odpovídají jejich autoři.

© Petr Čech, Daniel Svozil, 2016
Cover Design © Petr Čech, 2016

ISBN 978-80-7080-960-0

Obsah	3
Organizační a vědecký výbor	4
Abstrakty	7
Sekce 1	7
<i>Sekvence – nukleové kyseliny</i>	
Sekce 2	15
<i>Metody</i>	
Sekce 3	23
<i>Chemoinformatika</i>	
Sekce 4	33
<i>Struktury</i>	
Sekce 5	47
<i>Aplikace</i>	
Poster Session	55
<i>CZ-OPENSREEN</i>	
ELIXIR CZ	105
Seznam přednášek	107
Seznam posterů	111
Autorský index	117
Seznam účastníků	121

Organizační výbor

doc. Daniel Svozil, Ph.D.

Ing. Petr Čech, Ph.D.

Vědecký výbor

doc. Daniel Svozil, Ph.D.

RNDr. Petr Bartůněk, CSc.

prof. Jan Holub, Ph.D.

RNDr. Jiří Vondrášek, CSc.



SEKCE 1

Sekvence – nukleové kyseliny

L1-01

Metódy pre spracovanie sekvenačných dát z moderných sekvenátorov

Boža V.¹, Brejová B.¹, Vinař T.¹

¹ Fakulta matematiky, fyziky a informatiky Univerzity Komenského v Bratislave, Mlynská dolina, 842 48 Bratislava, Slovenská republika

V oblasti sekvenovania zažívame neustále inovácie, ktoré menia nielen cenu dát a rýchlosť, akou ich dokážeme získavať, ale aj ich charakter. Sekvenovanie druhej generácie (ako napríklad Illumina) prinieslo obrovský nárast priepustnosti sekvenovania, no bolo potrebné sa vysporiadať s oveľa kratšími čítaniami. Naproti tomu najnovšia generácia sekvenátorov (Oxford Nanopore a Pacbio) priniesla veľmi dlhé čítania, ale s vysokou chybovosťou.

Takýto rýchly vývoj prináša potrebu vysporiadať sa s kombináciou heterogénnych dát z rôznych sekvenovacích behov. Za týmto účelom sme vyvinuli novú metódu na skladanie genómov GAML. Metóda je založená na pravdepodobnostnom modelovaní, pričom očakávané vlastnosti dát, ako napríklad dĺžku čítaní, vzdialenosť párových čítaní, či chybovosť, možno do modelu zakódovať prirodzeným spôsobom a rôzne typy dát tak možno transparentne kombinovať. V príspevku tiež spomeniem problematiku dekódovania signálov zo sekvenátorov MinION (Oxford Nanopore). Vyvinuli sme nový softvér DeepNano, ktorý za pomoci rekurentných hlbokých neurónových sietí dokáže dekódovať sekvencie rýchlejšie a s vyššou presnosťou, ako softvér poskytovaný výrobcom.

L1-02

Charakterizace satelitních sekvencí z NGS dat

Novák P.¹, Macas J.¹

¹ *Biologické Centrum AV ČR, v. v. i.*

Sekvenování pomocí nových technologií (NGS) se stává vysoce dostupnou a levnou technologií v biologickém výzkumu. Satelitní sekvence jsou důležitým markerem v molekulární cytogenetice a k jejich charakterizaci se dosud používaly nástroje založené na analýze asemblovaných sekvencích, které však nejsou vždy dostupné. Bioinformatický nástroj, který by charakterizoval satelitní sekvence přímo z NGS dat, však dosud nebyl k dispozici. Proto jsme vyvinuli nástroj nazvaný TAndem REpeat ANalyzer - TAREAN, který dokáže identifikovat a charakterizovat satelitní sekvence přímo z shotgunových NGS dat a usnadní cytogenetický výzkum na dosud necharakterizovaných biologických druzích.

L1-03

Role retrotransposonů při expanzích genových rodin v genomu člověka a myši

Janoušek V.^{1,2}, Laukaitis Ch. M.³, Yanchukov A.^{2,4}, Karn R. C.³

¹ *Department of Zoology, Faculty of Science, Charles University in Prague*

² *Institute of Vertebrate Biology, ASCR, Brno, Czech Republic*

³ *Department of Medicine, University of Arizona, Tucson, AZ, USA*

⁴ *Department of Biology, Faculty of Arts and Sciences, Bülent Ecevit Üniversitesi, Zonguldak, Turkey*

Retrotransposony tvoří značnou část savčího genomu. Bylo prokázáno, že významně přispívají ke strukturálním přestavbám a k regulaci genů. Expanze a diverzifikace genových rodin je považována za zdroj evolučních novinek. Bližší prozkoumání významu retrotransposonů při expanzích genových rodin však doposud nebylo prozkoumáno. V prezentované studii jsme zjistili statisticky signifikantní vztah mezi dvěma hlavními třídami retrotransposonů (LINEs, LTRs) a liniově-specifickými expanzemi v genomu člověka a myši. Při bližším prozkoumání jsme zjistili, že LINEs a LTRs se liší distribucí a diverzitou, což naznačuje, že by tyto dvě třídy mohly přispívat odlišně při expanzích genových rodin. V okolí genových rodin jsou LTRs často asociovány s „open chromatin sites“, tj. oblastmi, kde je DNA přístupná pro vazbu transkripčních faktorů. Naopak LINEs mohou hrát spíše strukturální roli a usnadňovat genovou duplikaci. Naše výsledky také naznačují, že by expanze genových rodin – obzvláště v myším genomu – mohly procházet dvěma fázemi. První fáze by byla charakterizována zvýšenou akumulací LTR elementů a jejich využitím v „přesítování“ genových regulačních sítí. Druhá fáze by pak mohla být charakterizována rychlou expanzí z důvodu pokračující akumulace LINEs a zdá se, že v některých případech toto může vést ke zrychlení expanze – tzv. „runaway process“. Na příkladu konkrétní genové rodiny popisujeme možnou interakci LTRs a LINEs během expanze. Proces „runaway“ expanze dokládáme počítačovými simulacemi. Naše studie poskytuje důkaz o přispění retrotransposonů během expanze a evoluce genových rodin a podtrhuje tak významnost těchto elementů k diverzifikaci a adaptaci u člověka a myši.

L1-04

PredictSNP2: Konsenzuální prediktor vlivu nukleotidových mutací na rozvoj monogenních chorob

Bendl J.^{1,2,3}, Musil M.^{1,3}, Štourač J.^{1,2}, Zendulka J.³, Damborský J.^{1,2}, Brezovský J.^{1,2}

¹ *Loschmidtovy laboratoře, Přírodovědecká fakulta, Masarykova univerzita, Kamenice 5/A13, 625 00 Brno*

² *Centrum biomolekulárního a buněčného inženýrství, Fakultní nemocnice u sv. Anny v Brně, Pekařská 53, 656 91 Brno*

³ *Fakulta informačních technologií, VUT v Brně, Božetěchova 1, 612 66 Brno*

Projekt encyklopedie genomových elementů (ENCODE) změnil pohled na dlouhodobě zažité dogma, že většina lidského genomu mimo protein-kódující oblasti nezastává žádnou funkci. Vliv nalezených mutací byl v těchto oblastech však donedávna obtížně předvídatelný, jelikož na úrovni nukleotidových substitucí chybí silné fyzikálně-chemické deskriptory. K průlom došlo až v roce 2014 po zveřejnění výsledků genomových projektů ENCODE a Epigenomics Roadmap, které detekovaly určitou biologickou aktivitu, např. vazbu transkripčního faktoru, pro přibližně 80% genomu. Na základě těchto projektů byla v následujících dvou letech publikována řada nových prediktorů využívajících nově vzniklé anotace genomických regionů k prvotnímu ohodnocení efektu mutací a jejich následnou prioritizaci pro experimentální charakterizaci. Ačkoliv je výběr výpočetní metody zcela zásadní, jejich přehledné a věrohodné srovnání dosud chybělo. V rámci této studie byly zkonstruovány tři datové sady kombinující neutrální a s nemocí asociované mutace, které se zaměřují na odlišné druhy chorob: (i) monogenní, (ii) komplexní a (iii) nádorová onemocnění. Mutace nacházející se v těchto datových sadách byly posléze rozděleny do pěti specifických kategorií zahrnující regulační a sestřihové mutace v protein-nekódujících regionech a synonymní, nesynonymní a nesmyslné mutace v protein-kódujících regionech. Tyto datové sady byly použity k objektivní evaluaci nástrojů. Nejúspěšnější nástroje v každé kategorii (CADD, DANN, FATHMM, FunSeq2 a GWAVA) byly následně integrovány do podoby konsenzuálního klasifikátoru, který významně zvyšuje úspěšnost binární predikce škodlivosti mutací i přesnost odhadu predikční spolehlivosti. Kromě váhovaného konsenzuálního výsledku nástroj poskytuje uživatelské rozhraní s přístupem k výsledkům všech integrovaných nástrojů a k anotacím z osmi databází. Webový server a datové sady jsou volně dostupné na adrese <http://loschmidt.chemi.muni.cz/predictsnp2>.

L1-05

Rychlá deterministická vizualizace shotgun metagenomických dat

Sedlář K.¹, Kupková K.¹, Provazník I.¹

¹ *Ústav biomedicínského inženýrství, Fakulta elektrotechniky a komunikačních technologií, VUT v Brně, Technická 12, 616 00 Brno*

Velmi rychlý vývoj technik sekvenace DNA, především nástup třetí generace sekvenátorů, výrazně mění způsob, jakým je prováděn metagenomický výzkum, tj. výzkum genomů různých mikroorganismů nacházejících se ve společném prostředí. Stále častěji je místo cílené amplikonové sekvenace reprezentativního genu využívána kompletní shotgun sekvenace celého metagomu. Tento postup však klade zcela nové nároky na bioinformatické nástroje pro zpracování takových dat. Kompletní identifikace všech získaných sekvencí pomocí jejich srovnání s referenčními sekvencemi není možná z důvodu časové náročnosti a chybějící referenční databáze, která by musela obsahovat kompletní genomy všech mikroorganismů. Alternativní přístup však může být nalezen ve vhodné vizualizaci metagenomů, v níž jsou sekvence patřící ke stejným genomům organizovány do snadno rozlišitelných shluků, které je následně jednodušší identifikovat. Současné bioinformatické nástroje plní tuto funkci, jako například samoorganizační mapy (SOM), však obvykle strádají kvadratickou a vyšší časovou složitostí, která komplikuje zpracování velkých datových sad, nebo při nejlépe logaritmické složitosti nejsou deterministické, příkladem může být Barnes-Hut stochastic neighbor embedding (SNE).

Námi navržený postup spočívá v konverzi znakového zápisu jednotlivých čtení do numerického formátu fázového signálu, který reflektuje chemickou strukturu nukleotidů a který je specifický pro různé genomy. Řadí se tak mezi stále více využívané techniky číslcového zpracování genomických signálů. Fázový signál má podobné vlastnosti jako jiné biologické signály, např. EEG, a je tedy možné na něj aplikovat podobné algoritmy. Redukce dimenzionality do výsledné 3D reprezentace je následně provedena výpočtem Hjorthových parametrů (aktivita, mobilita a složitost). Ve finální vizualizaci jsou pak jasně patrné shluky sekvencí patřící ke stejným genomům. Mezi hlavní výhody námi navržené metody přitom patří lineární časová složitost a deterministický charakter výpočtu. Ačkoli metoda poskytuje nejlepší výsledky pro dlouhá čtení (~5000 bp) typická pro třetí generaci sekvenátorů, je aplikovatelná i na data ze starší „next generation“ sekvenace. S využitím dalších technik pro redukci dimenzionality, jako analýzy hlavních komponent (PCA), je navíc možné dosáhnout redukce i do úrovně 2D.



SEKCE 2

Metody

L2-01

Sémantická dvojšluková analýza dat genové exprese

Malinka F.¹, Kléma J.¹, Železný F.¹

¹ *České vysoké učení technické v Praze, Fakulta elektrotechnická, Karlovo náměstí 13, 121 35 Praha*

Dvojšluková analýza dat genové exprese je používaná bioinformatická metoda pro vyhledávání významných lokálních vzorů v genové matici. Matice genových expresí je uvedenou metodou analyzována, na rozdíl od klasických jednodimenzionálních shlukovacích metod, současně ve dvou dimenzích: v dimenzi obsahující geny (řádky) a v dimenzi obsahující jednotlivé vzorky (sloupce). Algoritmus pro dvojšlukování tedy identifikuje podskupiny genů v podskupině vzorků. Typickým příkladem sémantického shlukování při analýze genových dat je shlukování genů na základě hodnoty exprese, kde jednotlivé shluky mohou být charakterizovány pomocí termů z Gene ontology.

Cílem sémantického dvojšlukování je nalézt homogenní dvojšluky charakterizované společným výskytem termů náležícím znalostním bázím, jenž přísluší daným řádkům (genům) a sloupcům (vzorkům) matice. Sémantické dvojšlukování je tedy metoda kombinující dvojšlukování a sémantické shlukování.

Námi navrhovaná metoda byla motivována volně dostupným datasetem Dresden ovary table, jehož jednotlivé geny (řádky) mohou být popsány pomocí termů z Gene ontology a termy příslušné lokace (sloupce) mohou být popsány speciálně vytvořenou location ontology. Metoda je založena na technice symbolického strojového učení (rozhodovací pravidla a rozhodovací stromy), kde jsou homogenní dvojšluky popsány konjunkcemi termů z příslušných ontologií.

Tento projekt byl podpořen grantem GAČR 14-21421S. František Malinka je podporován grantem Studentské grantové soutěže ČVUT SGS16/093/OHK3/1T/13.

L2-02

Hammock: nástroj pro shlukování velkého množství krátkých peptidových sekvencí

Krejčí A.¹, Hupp T.², Lexa M.³, Vojtěšek B.¹, Muller P.¹

¹ *Regionální centrum aplikované molekulární onkologie, Masarykův onkologický ústav, Žlutý kopec 7, 656 53 Brno*

² *University of Edinburgh, Institute of Genetecs and Molecular Medicine, Cancer Research Centre, Edinburgh EH4, 2XR, UK*

³ *Fakulta informatiky Masarykovy univerzity, Botanická 68a, 602 00 Brno*

Proteiny často rozeznávají interakční partnery na základě krátkých lineárních motivů v nestrukturovaných regionech na jejich povrchu. Experimentální metody studující tyto motivy využívají krátké peptidy napodobující vlastnosti interagujících proteinů. Rozvoj takových metod dnes dovoluje využití rozsáhlých knihoven náhodných peptidů, výsledkem experimentů tohoto druhu bývá velké množství dat potenciálně obsahujících informace o několika různých interakcích. Zpracování takovýchto datových souborů je komplexní úkol nezbytný pro charakterizaci zkoumaných protein-proteinových interakcí.

V rámci přednášky bude představen softwarový nástroj pro shlukování krátkých peptidových sekvencí Hammock. Program je schopen identifikovat shluky peptidů sdílejících sekvenční motiv v datech o řádově až milionech unikátních sekvencí a vytvořit vícenásobná zarovnání nalezených shluků. Pro demonstraci výsledků byla metoda aplikována na menší soubor dat obsahující různé třídy ligandů SH3 domén a větší data reprezentující sekvence napodobující epitopy několika monoklonálních protilátek. Software úspěšně identifikuje známé motivy a odhalí i sekundární shluky, jejichž existence napovídá, že protilátky jsou schopny interagovat i s epitopy s mírně odlišnou sekvencí.

Práce byla podpořena Evropským fondem pro regionální rozvoj a státním rozpočtem České republiky (OP VaVpI - RECAMO CZ.1.05/2.1.00/03.0101).

Práce byla podpořena státním rozpočtem České republiky (LO1413).

L2-03

CeCe - simulátor mezibuněčné komunikace v dynamickém prostředí

Fatka J.¹¹ *Západočeská univerzita v Plzni, Univerzitní 8, 306 14 Plzeň*

CeCe je nástroj pro simulaci buněk a mezibuněčné komunikace ve 2D dynamickém prostředí. Jeho součástí je systém stochastických chemických reakcí, systém napojování a oddělování buněk, difúzní systém, hydrodynamika, fyzikální engine, model růstu buněk i skriptovací jazyk pro dynamické změny v chování simulace. Tyto všechny vlastnosti mu umožňují simulovat chování buněk v různých prostředích, které mohou výrazně ovlivnit chování individuálních buněk i jejich vzájemné komunikace. CeCe byl původně vytvořen jako součást projektu českého týmu v rámci mezinárodní soutěže iGEM 2015 se záměrem simulace experimentů prováděných na mikrofluidických čípech. Uvedené schopnosti však umožňují vytvářet simulace nad rámec původního účelu a to např. simulace buněčných kultur, krevních řečišť, buněčných tkání a dalších biomedicínsky zaměřených problémů.

L2-04

Virtuální paralelní infrastruktury pro velká data v metabolomice

Fesl J.^{1,2,3}, Doležalová M.^{1,2}, Cehák J.^{1,2}, Moos M.², Janeček J.⁴, Šimek P.²

¹ Jihočeská univerzita v Českých Budějovicích, Přírodovědecká fakulta, Ústav aplikované informatiky, Branišovská 1645/31A, 370 05 České Budějovice

² Biologické centrum AV ČR, v. v. i., Entomologický ústav, Laboratoř analytické biochemie a metaboliky, Branišovská 1160/31, 370 05 České Budějovice

³ ČVUT FEL, Katedra počítačů, Karlovo nám. 13, 121 35 Praha 2

⁴ ČVUT FIT, Katedra počítačových systémů, Thákurova 9, 160 00 Praha 6

Metabolomická data, jejich analýza a interpretace jsou žhavým fenoménem posledních let. Stále objemnější a komplikovanější data vyžadují pro své zpracování dostatečně výkonné výpočetní infrastruktury v podobě dedikovaných paralelních počítačů či distribuovaných počítačových clusterů, jejichž vytváření a správa vyžaduje nemalé finanční prostředky. Náš tým navrhl a sestavil prototyp univerzálního elastického distribuovaného výpočetního systému (Centrální Mozek Univerzity, CMU) založeného na virtualizační technologii, který lze využít k řešení náročných bioinformatických problémů. Elasticita systému spočívá v tom, že během několika minut lze změnit jednu konkrétní výpočetní infrastrukturu na jinou bez nutnosti použití jiného hardwaru. Prototyp systému obsahuje sofistikovaný modul pro alokaci výpočetních zdrojů, vyvažování zátěže a monitorování stavu systému. Další předností našeho řešení je možnost dynamické rozšiřitelnosti výpočetní kapacity přímo za běhu, včetně možnosti hardwarových úprav či údržby.

Systém CMU je využíván například k analýze chemometrických dat naměřených pomocí hmotnostních spektrometrů (LR/HR LC i GC MS) a jejich následné interpretaci. Pro analýzu chemometrických dat vyvíjíme od roku 2006 softwarovou platformu Metabolite Mapper (M2), která je dnes již komplexním maturovaným systémem. M2 umožňuje rychlé paralelní zpracování souborů v nativním formátu firmy Thermo, vyhledání píků, časovou synchronizaci a identifikaci látek, srovnávání různých skupin měření navzájem atd. Při vývoji systému byl kladen velký důraz na možnost jeho využití v paralelních či distribuovaných systémech, které umožňují zpracování velkých dat v relativně krátkém čase.

L2-05

Chameleon 2: algoritmus pro hierarchické shlukování dat

Bartoň T.^{1,2}, Kordík P.²

¹ *Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4*

² *Fakulta informačních technologií, Tháškova 9, 160 00 Praha 6*

Tradiční shlukovací algoritmy se často nedokáží přiblížit výsledkům, které byly ručně anotované člověkem. Problémy působí především odlehle hodnoty, velmi rozdílná distribuce dat či šum v pozorovaných datech, které člověk dokáže velmi rychle rozpoznat.

Navrhujeme nový shlukovací algoritmus založený hierarchickém shlukování dat, který vychází z původní verze grafového algoritmu Chameleon. Náš přístup vykazuje vysokou kvalitu výsledného shlukování, které se v mnoha případech blíží lidskému úsudku. V benchmarku založeném na umělých datech (především FCPS) vykazuje v průměru zlepšení o 15% oproti běžně používaným algoritmům.

Algoritmus v první fázi používá metodu nejbližších sousedů, na základě které sestaví graf. Tento graf je následně pomocí odebrání hran rozdělován na menší části přičemž ověřujeme zda-li odebráním hrany dosáhneme zlepšení funkce, používané pro interní ohodnocení kvality shlukování. V další fázi se snažíme eliminovat izolované body, které jinak negativně ovlivňují výslednou kvalitu shlukování. Posledním krokem je aglomerativní fáze, kdy spojujeme nejbližší body na základě lineární kombinace dvou kritérií.

Navrhovaný algoritmus je efektivnější než hierarchické shlukování a zároveň poskytuje kvalitnější výsledky i v základní konfiguraci bez nutnosti náročné optimalizace parametrů.



SEKCE 3

Chemoinformatika

L3-01

Predikce toxicity nanočástic

Anděl M.^{1,2}, Kléma J.^{1,2}

¹ Oddělení genetické ekotoxikologie, Ústav experimentální medicíny AV ČR, Vídenská 1083, 142 20 Praha 4

² Katedra počítačů, Fakulta elektrotechnická ČVUT v Praze, Karlovo náměstí 13, 121 35 Praha 2

Zatímco predikce vlastností chemických sloučenin, QSAR, se stala zavedenou disciplínou, její zobecnění pro nanočástice, tzv. nano-QSAR, je stále netriviální. Metody QSAR odvozují aktivitu chemické sloučeniny na základě struktury jejich molekul. Tyto přístupy těžší zejména ze strukturní pestrosti organických molekul a invariance jejich vlastností v libovolném množství. Jinak je tomu v případě predikce aktivity, zejména toxicity, v případě částic obecně. Nanočástice mají často jednoduché anorganické složení, avšak kombinatoricky složitý návrh (tj. různé typy potahu, obohacení, povrchového ošetření). Vedle vnitřních vlastností základního materiálu, mají také vlastnosti, které jsou funkcí jejich velikosti, tvaru nebo povrchové struktury. Na těchto vnějších vlastnostech pak závisí jevy jako aglomerace částic, tvorba proteinového obalu, potažmo příjem částice buňkou. Modelovat tyto vlastnosti je obtížné vzhledem k náročnosti výpočtu kvantově chemických deskriptorů pro větší krystaly nanočástic, respektive špatné dostupnosti fyzikálně chemických měření po popis vnějších jevů jako aglomerace nebo příjem buňkou. Tyto vnější jevy, stejně jako toxická odpověď, jsou dále velmi citlivé na experimentální podmínky (koncentrace, doba expozice, složení rozpouštědla), které jsou v případě většiny současných toxikologických studií velmi heterogenní, ba dokonce nedostupné.

Ze zmíněných důvodů mají současné modely toxicity nanočástic velmi omezenou platnost. Existují relativně přesné modely, založené na kvantově chemických deskriptorech, které dokonce vysvětlují mechanistickou podstatu toxicity, avšak platné pouze pro několik málo dobře charakterizovaných částic, hodnocených za homogenních experimentálních podmínek [1]. Tyto modely však nejsou vhodné pro velké projekty jako MODENA COST, zaměřené na sběr dat z evropských laboratoří. Abychom se vyrovnali s nastíněnými problémy, zejména odstínili vliv experimentálních podmínek na toxickou odpověď, a odhalili skutečnou toxicitu částic, vyvíjíme nové metody, založené na relačním strojovém učení [2], zejména na bayesovských sítích a induktivním logickém programování. Cílem relačního přístupu je modelovat nejen vztah k toxicitě, ale i ostatní relace v doméně; např. modelovat experimentální podmínky, a ty nakonec od výsledku „odečíst“.

Náš výzkum probíhá v rámci výzkumné infrastruktury NanoEnviCz, podporované Ministerstvem školství, mládeže a tělovýchovy (granty LM2015073 a COST.CZ, LD14002 NANOSTOX).

Reference

- [1] Puzyn, Tomasz, et al. "Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles." *Nature nanotechnology* 6.3 (2011): 175-178.
- [2] Getoor, Lise. *Introduction to statistical relational learning*. MIT press, 2007.

L3-02

Léčiva na biologických membránách pohledem výpočetní chemie

Berka K.¹

¹ *Katedra fyzikální chemie, RCPTM, Přírodovědecká fakulta, Univerzita Palackého v Olomouci, 17. listopadu 1192/12, 771 46 Olomouc*

Bariéra, rezervoár, nebo molekulární cíl – to jsou jen některé funkce, které hrají biologické membrány v komplexní výměně látek v organismu. Detailní znalost interakcí mezi molekulami a membránami je tak podmínkou i pro lepší pochopení chování léčiv v organismu. Nízkomolekulární látky - léčiva - s biologickými membránami interagují mnoha různými způsoby v závislosti na svých fyzikálně chemických vlastnostech a koncentraci. Mohou se na membrány adsorbovat, nebo do nich mohou vstupovat, akumulovat se v nich nebo jimi prostupovat na druhou stranu. A přesto je nejčastějším a často jediným parametrem používaným k popisu interakce molekul s membránou ve většině databází rozdělovací koeficient oktanol/voda (logP). Protože ale biologické membrány nejsou tvořeny homogenní oktanolovou fází, tak se v přednášce zaměřím na popis toho, jak pozici látek na membráně [1] jejich afinitu k ní [2-4] či penetrační vlastnosti [5] počítat a tedy predikovat. Cílem tohoto snažení je do budoucna zpřesnit a poskytnout detailní experimentální i predikované informace o afinitě, pozicování a penetraci látek membránami v otevřeně přístupné databázi.

Reference

- [1] Berka, K., Hendrychová, T., Anzenbacher, P. & Otyepka, M. Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site. *J. Phys. Chem. A* 115, 11248–55 (2011).
- [2] Paloncýová, M., Berka, K. & Otyepka, M. Molecular Insight into Affinities of Drugs and Their Metabolites to Lipid Bilayers. *J. Phys. Chem. B* 117, 2403–10 (2013).
- [3] Paloncýová, M., DeVane, R., Murch, B., Berka, K. & Otyepka, M. Amphiphilic drug-like molecules accumulate in a membrane below the head group region. *J. Phys. Chem. B* 118, 1030–9 (2014).
- [4] Paloncýová, M. et al. Benchmarking of Force Fields for Molecule–Membrane Interactions. *J. Chem. Theory Comput.* 10, 4143–4151 (2014).
- [5] Paloncýová, M., DeVane, R. H., Murch, B. P., Berka, K. & Otyepka, M. Rationalization of reduced penetration of drugs through ceramide gel phase membrane. *Langmuir* 30, 13942–8 (2014).

L3-03

QSAR modelování jako nástroj pro zkoumání chemogenomického prostoru

Škuta C.^{1,2}, Svozil D.^{1,2}

¹ CZ-OPENSOURCE, Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

² Laboratoř informatiky a chemie, Fakulta chemické technologie, VŠCHT Praha, Technická 5, 166 28 Praha 6

Chemogenomika je výzkumná oblast s cílem systematicky identifikovat účinek všech možných ligandů na všechny možné biologické cíle. Chemogenomická data jsou uspořádána ve formě tzv. chemogenomické matice, interakční matice biologický cíl-ligand obsahující odpovídající vazebné afinity. Z praktických důvodů není možné ke tvorbě této matice použít pouze experimentálních metod. *In silico* přístupy proto hrají důležitou roli při vývoji prediktivních chemogenomických aplikací. Tyto metody jsou typicky založeny na porovnávání ligandů anotovaných biologickými vlastnostmi jejich vazebných míst nebo vazebných míst anotovaných vlastnostmi jejich ligandů. QSAR modelování (Quantitative Structure-to-Activity Relationship) je typickým zástupcem *in silico* metod běžně používaným pro predikci afinit na jednotlivých biologických cílech. S použitím QSAR modelování a databáze bioaktivit ChEMBL jsme vytvořili workflow pro predikci chemogenomického prostoru ve velkém měřítku (stovky až tisíce cílů). V rámci tohoto workflow byly vytvořeny a validovány QSAR modely, které predikují afinity s velkou měrou věrohodnosti a přesnosti. Výsledná množina modelů definuje z hlediska biologických cílů hranice chemogenomického podprostoru, který je dále možné zkoumat. Zhodnocení relevance prostoru jsme otestovali na základě popisné síly predikovaných afinit, uspořádaných do tzv. afinitních fingerprintů, v rámci běžných klasifikačních a regresních úloh. V obou případech vykazovaly afinitní fingerprinty srovnatelné nebo lepší výsledky než běžně používané strukturní deskriptory (ECFP či MACCS strukturní klíče).

L3-04

Jak najít jehlu v kupce sena aneb orientace v dnešních databázích

Jarkovská K.¹, Svoboda O.¹, Pech D.¹, Šauer M.¹

¹ Chemaxon Ltd. 1031 Budapest, Zahony U. 7, Hungary

Jedním z nejvýznamnějších trendů v medicínské chemii v posledních letech je obrovské rozšíření chemických, biologických a klinických dat. Odhaduje se, že celý chemický prostor obsahuje na 10^{60} molekul. Největší kombinatorické databáze dnes pracují asi s 10^{11} molekul. Interní databáze firem a institucí jsou sice menší, ale také nezadržitelně rostou. Při hledání nových léků tak analýza dat čím dál častěji nahrazuje zkušenosti organických chemiků. Dramatické rozšíření použití velkých databází má ale i své stinné stránky. Tou nejzásadnější je patrně to, jak se v tak obrovských sadách chemických dat vyznat.

Moderním nástrojem, jak hledat aktivní látky je dnes takzvané testování s vysokou propustností (*high-throughput screening*). Jedná se o techniku, kdy se najednou automaticky testují desítky až stovky tisíc látek, které jsou potenciálně biologicky aktivní a mohly by být použity jako nové léky. Přestože je toto testování nesmírně užitečné, má i své slabiny. Tou hlavní je to, že drtivá většina testovaných látek biologicky aktivní není. Naším hlavním cílem je ukázat, jak vystavět lepší databázi, ze které vyřadíme co nejvíc neaktivních látek, a zároveň zachováme většinu látek aktivních. Představme si, že aplikujeme testování s vysokou propustností pro malou skupinu látek. Poté analyzujeme data a na základě této analýzy vytvoříme databázi novou, zacílenou na aktivní látky.

Svůj postup ukážeme na příkladu volně dostupných dat z testování s vysokou propustností zaměřeného na malárii. Testování bylo provedeno v dětské nemocnici v St. Jude v Memphisu [1] a zahrnovalo 309 474 chemických látek. Z těchto látek budeme vycházet. Pokusíme se ukázat, jestli je možné pomocí datové analýzy počáteční databázi výrazným způsobem zredukovat, ale tak abychom zachovali vše důležité. Předvedeme si využitelnost hierarchického klastrování (použijeme nástroj JKlustor od firmy ChemAxon). Budeme také uvažovat vliv počátečního výběru dat. Odpovíme tím na otázku, jak může datová analýza pomoci při návrhu testování s vysokou propustností.

Reference

- [1] Guiguemde, W. A.; et al. Chemical genetics of *Plasmodium falciparum*. *Nature* 2010, 465, 311–315.

L3-05

Pokročilé dotazování jazykem SPARQL v databázích malých molekul

Galgonek J.¹

¹ *Ústav organické chemie a biochemie AV ČR, v. v. i., Flemingovo nám. 2, 166 10 Praha 6*

V oblasti bioinformatických a chemoinformatických databází se v posledních letech stále častěji uplatňují technologie Sémantického webu, jako jsou Resource Description Framework (RDF) a dotazovací jazyk SPARQL. Tyto technologie umožňují lepší propojení různých zdrojů dat a snadnější vyhledávání v těchto datech. Z jejich použití ovšem plyne také několik nedostatků, které mohou pro běžné uživatele představovat jistá omezení. Z jazyka SPARQL nelze standardním způsobem volat speciální procedury, které by poskytly data, která nelze běžným způsobem předpočítat a uložit do databáze. Příkladem mohou být procedury, které umožňují vyhledat chemické sloučeniny na základě strukturní podobnosti. Rovněž jazyk SPARQL nemá příliš velké požadavky na sémantickou správnost dotazu. Prostým překlepem tak lze například snadno vytvořit dotaz, který je formálně zcela správně, avšak nevrátí žádný výsledek, což uživateli znesnadňuje psaní a ladění dotazů. Posledním nedostatkem je způsob prezentace výsledků SPARQL dotazu. Obecně se výsledky dotazu zobrazují jako tabulka obsahující pouze databázové identifikátory či literály. Taková prezentace výsledků je však pro uživatele velmi nepohodlná. V námi představeném systému jsme se pokusili vyřešit všechny tyto nedostatky. Naše rozšíření jazyka SPARQL umožňuje volat procedury. Dotazy jsou kontrolovány proti ontologii uložených dat a uživatelé jsou v případě objevených nesrovnalostí varováni. Výsledky dotazů jsou prezentovány pomocí šablon definovaných pro jednotlivé entity uložené v databázi. Pro ukázkou funkčnosti našeho systému jsme zvolili data z databází ChEBI a PubChem. Procedury umožňující vyhledávání na základě strukturní podobnosti jsou založeny na projektu OrChem. Systém je možné vyzkoušet na adrese <https://bioinfo.uochb.cas.cz/projects/chemRDF>.

L3-06

Vývoj nového farmakoforového modelu

Mareška V.¹, Spiwok V.¹¹ *VŠCHT Praha, Ústav biochemie a mikrobiologie, Technická 3, 166 28 Praha 6*

Použití farmakoforů se stává stále více populární způsobem hledání nových léčiv. V porovnání s tradičními metodami, farmakoforové modely dovolují být rychlým a účinným nástrojem pro virtuální screening velkých databází sloučenin. Obecně se farmakoforové modely snaží charakterizovat sloučeniny transformací jejich strukturních vlastností na kolektivní proměnné, čímž vytváří “otisky prstů” sloučenin, které mohou být snadno porovnány. Pokusili jsme se navrhnout a implementovat nový farmakoforový model založený na: CATS, SQUID a LIQUID modelech. Náš model jsme využili pro screening více než 39 milionu molekul z databáze ZINC. Nyní model testujeme na hledání nových inhibitorů cyklooxygenasy-2 (COX-2) se stejnou nebo dokonce lepší biologickou aktivitou než mají dosud známé inhibitory. Společně s dockingem budeme testovat model na screening dalších databází, hledání nových aktivních molekul a budeme dále vylepšovat výpočetní rychlost a výkonnost našeho modelu.

L3-07

Využití nábojových deskriptorů v chemoinformaticce

Geidl S.¹, Hejret V.¹, Svobodová Vařeková R.¹, Koča J.¹

¹ *Masaryk University, National Centre for Biomolecular Research, Faculty of Science and CEITEC - Central European Institute of Technology, Kamenice 5, 625 00 Brno*

Molekulové deskriptory [1] jsou jedním z nejdůležitějších nástrojů současné chemoinformatiky. Pomocí těchto deskriptorů můžeme efektivně popsat a kvantifikovat různé vlastnosti a aspekty molekul. Deskriptory jsou z definice čísla, která lze algoritmicky vypočítat na základě struktury molekul. Nejčastějšími aplikacemi deskriptorů je jejich využití pro predikci aktivity [2] a vlastností chemických látek (např.[3]) nebo pro zkoumání podobnosti jednotlivých molekul [4]. Specifickou skupinou deskriptorů jsou tzv. nábojové deskriptory, které jsou počítány na základě parciálních atomových nábojů. Některé z nábojových deskriptorů se prokázaly jako velmi užitečné pro predikci klíčových chemických vlastností molekul [2,3]. Na druhé straně, práce s nábojovými deskriptory je netriviální a stále zůstává výzvou, protože existuje velké množství kvantově mechanických i empirických metodik pro výpočet nábojů.

Ve svém příspěvku bych chtěl poskytnout přehled dostupných i nově navržených nábojových deskriptorů. Dále se zaměřím na využitelnost nábojových deskriptorů, kterou ukáži na tvorbě modelů pro predikci hodnot rozdělovacího koeficientu logP. Rovněž budu prezentovat informace o dostupných softwarových nástrojích pro práci s nábojovými deskriptory, včetně ukázky námi vytvořeného software.

Reference

- [1] Todeschini R, Consonni V: Handbook of Molecular Descriptors. Wiley-VCH Verlag GmbH, Weinheim, Germany; 2008.
- [2] Furuhashi A, Aoki Y, Shiraishi H: Development of ecotoxicity QSAR models based on partial charge descriptors for acrylate and related compounds. SAR QSAR Environ Res 2012, 23:731–49.
- [3] Svobodová Vařeková R, Geidl S, Ionescu C-M, Skřehota O, Kudera M, Sehnal D, Bouchal T, Abagyan R, Huber HJ, Koča J: Predicting pKa Values of Substituted Phenols From Atomic Charges: Comparison of Different Quantum Mechanical Methods and Charge Distribution Schemes. J Chem Inf Model 2011, 51:1795–1806.
- [4] Kearsley SK, Sallamack S, Fluder EM, Andose JD, Mosley RT, Sheridan RP: Chemical Similarity Using Physiochemical Property Descriptors. J Chem Inf Model 1996, 36:118–127.



SEKCE 4

Struktury

L4-01

Suboptimální struktury hledají optimální RNA

Pánek J.¹, Schwarz M.¹, Vohradský J.¹

¹ MBÚ AV ČR, Vídeňská 1083, 142 20 Praha

Nekódující RNA jsou regulační molekuly s geny v intergenových nebo intragenových oblastech genomů. V prokaryotech, bioinformatická identifikace nových homologů známých RNA v dalších druzích často selhává kvůli slabě evolučně konzervovaným sekvencím, strukturám, syntenii a genomické pozici, s výjimkou evolučně úzce příbuzných druhů. Kvůli eliminaci důsledků slabé evoluční konzervovanosti, zde jsme se zaměřili na využití struktury RNA, která je nejkonzervovanější vlastnost RNA. Analýza struktury jedné z dobře popsanych bakteriálních RNA, 6S RNA, ukázala, že na rozdíl od optimálních a konsenzuálních struktur, suboptimální struktury jsou schopné zachytit homologii RNA i v divergentních druzích. Byla vytvořena výpočetní metoda pro identifikaci homologních RNA založená na suboptimálních strukturách. Metoda byla použita na silně divergentní druhy a byla schopná identifikovat v nich nové homologní RNA.

L4-02

Homologní predikce terciární struktury RNA

Hoksza D.¹, Galvánek R.¹

¹ *Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Malostranské nám. 25, 118 00 Praha*

Přístupy k predikci terciární struktury makromolekul lze rozdělit na de novo metody a metody homologního modelování. Zatímco pro predikci 3D struktury proteinů je homologní modelování využíváno běžně, v případě RNA struktur je tento přístup aplikován zřídka. Proto zde představujeme metodologii pro predikci 3D struktury RNA molekuly (cíle) využívající znalost příbuzné struktury (šablona). V prvním kroku je proveden sekvenční zarovnání cíle a vzoru. Následně jsou identifikovány příbuzné úseky obou molekul a příslušné podstruktury jsou pak přeneseny ze šablony na cíl. Části struktury, které nejsou společné, jsou následně dopredikovány externím nástrojem. Součástí této fáze je ošetření indelů a správné napojení predikovaných a přenesených částí struktury. Experimenty ukázaly, že kvalita predikce je především závislá na sekvenční podobnosti šablony a cíle a na délce a počtu mezer. Pro struktury o velikostech stovek nukleotidů se sekvenční podobností nad 50% a poměrem mezer do 50% je metoda schopná generovat predikce typicky do deseti jednotek RMSD vzhledem k referenční struktuře.

L4-03

Lokální konformery DNA a jejich automatické přiřazování

Černý J.¹, Božíková P.¹, Nečasová I.¹, Čech P.², Svozil D.², Schneider B.¹

¹ AV ČR, v. v. i., Biotechnologický ústav, Průmyslová 595, 252 50 Vestec

² VŠCHT, Laboratoř informatiky a chemie, Technická 5, 166 28 Praha 6

V této práci prezentujeme výsledky rozšířené analýzy lokální struktury DNA. Struktura DNA je na globální úrovni obecně popisovaná pomocí tří forem, jako nejčastěji se vyskytující B forma, dále A forma a nejméně zastoupená Z forma. Lokálně, na úrovni dinukleotidů, můžeme rozlišit desítky různých konformací, zahrnující například přechodné struktury mezi A a B formami, nebo formy se syn orientací báze. Výsledky reprezentují významné doplnění dříve publikovaného klasifikačního protokolu pro popis konformací DNA dinukleotidů (Svozil et al. Nucleic Acids Research 2008). Nově definovaný soubor 57 dinukleotidových konformerů zvaných ntC, jsme určili procesem selfkonzistentní klasifikace struktur 4439 dinukleotidů pocházejících z neredundantních struktur DNA s vysokým krystalografickým rozlišením. V průběhu analýzy byly nalezeny konformery, které se objevují v různých duplexových i neduplexových strukturách, například v kvadruplexech, four-way junction a dalších jednovláknových topologiích. Dále ukazujeme stručné anotace konformerů s důrazem na konformery nově popsané a jejich užitečnost demonstrujeme na popisu některých ukázkových struktur DNA.

L4-04

Detekce biomolekulárních substruktur pomocí PatternQuery

Pravda L.¹, Sehnal D.¹, Svobodová Vařeková R.¹, Koča J.¹

¹ *Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta a CEITEC – Středoevropský technologický institut, Masarykova Univerzita Brno, Kamenice 5, 625 00 Brno-Bohunice*

Přesně definované molekulární vzory jako jsou vazebná a katalytická místa, proteinové motivy nebo specifické sekvence aminokyselin nebo nukleotidů ovlivňují celou řadu biologických funkcí buňky. Jejich definice a přesná identifikace v biomolekulárních strukturách nachází uplatnění v mnoha oblastech živých věd jako je například strukturní a funkční charakterizace a srovnání proteinů, analýza nově vyřešených struktur nebo návrh a úprava nových funkčních míst proteinů.

PatternQuery (PQ) je online služba pro optimální definici a extrakci těchto biomolekulárních strukturních vzorů pomocí jednoduchého, avšak robustního, dotazovacího jazyka. Tyto vzory lze přesně popsat pomocí unikátních výrazů definujících primární, sekundární nebo terciární strukturu a jejich libovolné kombinace. Výsledky jsou poté přehledně zobrazeny v grafické a sumární podobě. Dále je lze stáhnout a použít jako vstup pro další *in silico* analýzy. Díky široké paletě základních výrazů a možnosti jejich neomezené kombinace lze popsat pestrou paletu různých molekulárních vzorů a jednoduše realizovat analýzy dříve obtížné nebo úplně nedostupné. Mezi hlavní výhody této služby patří její univerzálnost, rychlost a možnost dalšího rozšíření.

Interaktivní real-time aplikace PatternQuery je dostupná online (<http://ncbr.muni.cz/PatternQuery>) a také jako konzolová aplikace pro identifikaci biomolekulárních substruktur v jednotlivých proteinech. Velmi rychle lze zpracovat i kompletní databáze nebo jejich libovolné pod části (Protein Data Bank). Řada příkladů použití s příslušnými PQ dotazy a kompletní dokumentace je dostupná online. Služba je dostupná zdarma.

Na základě PatternQuery je také budována ve spolupráci s PDBe online služba pro rychlé streamování biomolekulárních substruktur - CoordinateServer.

Reference

- [1] Sehnal, D., Pravda, L., Svobodová Vařeková, R., Ionescu, C.-M., & Koča, J. (2015). PatternQuery: web application for fast detection of biomacromolecular structural patterns in the entire Protein Data Bank. *Nucleic Acids Research*, 43(W1), W383–W388. <http://doi.org/10.1093/nar/gkv561>

L4-05

Výskyt sekundárních struktur v sekvenčním prostoru proteinů

Treťjačenko V.¹, Vymětal J.², Bednárová L.², Vondrášek J.², Hlouchová K.^{1,2}

¹ *Univerzita Karlova v Praze, Přírodovědecká fakulta, Katedra biochemie, Hlavova 2030, 128 00 Praha 2*

² *Ústav organické chemie a biochemie, Flemingovo náměstí 2, 166 10 Praha 6*

V proteinové databázi Uniprot je v současné době evidováno více než 50 milionů potvrzených a predikovaných sekvencí. Strukturní databáze PDB čítá nad 100 tisíc proteinových struktur reprezentujících přibližně 1300 známých proteinových uspořádání (z angl. folds). Tato čísla ovšem tvoří zanedbatelný zlomek vůči astronomickému počtu kombinací představujících sekvence z 20 proteinogenních aminokyselin (pro 100 aminokyselinové řetězce je 10^{130} možných sekvenčních variant), které nedostaly šanci v biologickém kontextu (angl. „never born proteins“ - NBP). Dřívější hypotézy předpokládají, že třidimenzionální uspořádání proteinů je založeno na přírodní selekci vzácných sekvencí s danými vlastnostmi, zatímco sekvence nevystavené evolučnímu tlaku budou pravděpodobně neuspořádané.

Za účelem potvrzení nebo vyvrácení těchto spekulací jsme analyzovali knihovnu náhodných sekvencí. Naším cílem bylo vypracovat postup pro určení frekvence a podoby sekundárních a terciárních struktur. Prvním krokem byla bioinformatická analýza knihovny pomocí prediktoru sekundárních struktur, proteinové neuspořádanosti a ostatních fyzikálně-chemických vlastností (rozpuštěnost, hydrofobicita, agregační potenciál). S použitím těchto nástrojů jsme v knihovně náhodných sekvencí odhalili neočekávaně vysokou frekvenci výskytu sekundárních struktur.

Pro experimentální charakterizaci, s cílem ověření a následující analýzy našich výsledků, jsme připravili 45 proteinů z knihovny náhodných sekvencí. Sekvence byly vybírány tak, aby zapadaly do tří skupin, reprezentujících celé spektrum výskytu sekundárních struktur (vysoký, nízký a náhodný výběr z celé). Jednotlivé rozpustné proteiny byly exprimovány *in vitro*, purifikovány a charakterizovány biofyzikálními metodami (hmotnostní spektrometrie (MS), dynamický rozptyl světla (DLS), cirkulární dichroismus (CD)). CD spektra potvrzují výskyt sekundárních struktur pro jednotlivé skupiny náhodných sekvencí a souhlasí se závěry bioinformatické analýzy o téměř totožném vysokém výskytu sekundárních struktur v náhodném sekvenčním prostoru v porovnání s jejich biologickými protějšky.

V budoucnu se zaměříme na analýzu různých aminokyselinových repertoárů podle stejného protokolu a na vyhodnocení vlivu aminokyselinového složení na tendenci proteinů pro tvorbu sekundárních struktur. Konkrétně je naším cílem pozorovat vliv

evoluce genetického kódu (tedy postupného přidávání aminokyselin k dostupným kodonům) na potenciál proteinu pro tvorbu sekundárních struktur.

L4-06

HotSpot Wizard 2.0: Automatický návrh aminokyselinových mutací a chytrých knihoven pro proteinové inženýrství

Štourač J.^{1,2}, Bendl J.^{1,2,3}, Šebestová E.¹, Vávra O.¹, Musil M.^{1,2,3}, Brezovský J.^{1,2}, Damborský J.^{1,2}

¹ *Loschmidtovy laboratoře, Ústav experimentální biologie a Centrum pro výzkum toxických látek v prostředí, Přírodovědecká fakulta, Masarykova univerzita, Kamenice 5/A13, 625 00 Brno*

² *Mezinárodní centrum klinického výzkumu, Fakultní nemocnice u sv. Anny v Brně, Pekařská 53, 656 91 Brno*

³ *Ústav informačních systémů, Fakulta informačních technologií, Vysoké učení technické v Brně, Božetěchova 1, 612 66 Brno*

HotSpot Wizard je interaktivní webový server sloužící k predikci aminokyselinových reziduí vhodných k mutagenези [1] a konstrukci knihoven proteinových mutantů se zvýšenou aktivitou, změněnou specifitou nebo stabilitou [2]. Vhodné pozice pro mutagenези jsou vyhodnocovány na základě proteinové struktury pomocí kombinace strukturních, funkčních a evolučních informací získaných ze 3 internetových databází a 18 výpočetních nástrojů. Aplikace byla navržena s důrazem na snadné používání bez nutnosti pokročilých znalostí o studovaném systému - proto byly všechny výchozí hodnoty parametrů zvoleny na základě extenzivních analýz tak, aby vhodně reprezentovaly co nejširší spektrum vstupních dat. Pro každou proteinovou strukturu jsou automaticky vyhodnoceny čtyři odlišné strategie proteinového inženýrství: (i) identifikace evolučně variabilních a tedy relativně bezpečně modifikovatelných reziduí, které jsou umístěny ve funkčně relevantních přístupových tunelech či katalytických kapsách a mají tedy dobrý potenciál modulovat funkci proteinu, (ii) detekce vysoce flexibilních regionů, jejichž mutagenези může vést ke zvýšení stability lokálním omezením pohyblivosti, (iii) identifikace pozic s častým evolučním výskytem jiné aminokyseliny, jejíž zpětná mutace často stabilizuje protein a (iv) identifikace vzájemně evolučně korelovaných reziduí, které indikují jejich význam pro funkci či stabilitu proteinu. Analýza výsledků probíhá přímo ve webovém rozhraní, které kromě propojení s vizualizačním nástrojem nabízí také pokročilé možnosti filtrace. Vybraná rezidua je možné dále využít pro návrh knihovny proteinových mutantů s podporou automatické detekce vhodných cílových aminokyselin a korespondujících degenerativních kodonů. Aplikace je volně dostupná akademickým uživatelům na adrese <http://loschmidt.chemi.muni.cz/hotspotwizard>.

Reference

- [1] Pavelka, A., Chovancová, E., Damborský, J., 2009: HotSpot Wizard: a Web Server for Identification of Hot Spots in Protein Engineering. *Nucleic Acids Research* 37: W376-W383.
- [2] Bendl, J., Štourač, J., Šebestová, E., Vávra, O., Musil, M., Brezovský, J., Damborský, J., 2016: HotSpot Wizard 2.0: Automated Design of Site-Specific Mutations and Smart Libraries in Protein Engineering. *V recenzi*.

L4-07

Extrakce biologicky relevantních informací ze 3D struktur biomakromolekul

Svobodová Vařeková R.¹, Sehnal D.¹, Pravda L.¹, Geild S.¹, Horský V.^{1,2}, Raček T.^{1,2}, Koča J.¹

¹ *Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta a CEITEC – Středoevropský technologický institut, Masarykova Univerzita Brno, Kamenice 5, 625 00 Brno-Bohunice*

² *Fakulta informatiky, Masarykova univerzita, Botanická 68a, 602 00 Brno*

Biomakromolekuly (proteiny, nukleové kyseliny, polysacharidy atd.) jsou důležitými biologickými entitami, které tvoří základní funkční a stavební jednotky buněk. Tyto biomakromolekuly jsou rozsáhlými objekty, skládají se z několika tisíců atomů, a jejich různé části (fragmenty) hrají odlišné role - např. tvoří aktivní místa, váží určité ligandy či kovy, obklopují tunely nebo póry, jsou zodpovědné za formování molekuly jako celku apod. Výzkum těchto fragmentů může vést k velmi užitečným výsledkům jako je např. získání šablon pro návrh léků, porozumění vztahu mezi strukturou a funkcí proteinů nebo určení putativních funkcí biomakromolekul [1,2]. Klíčovou vlastností biomakromolekulárních fragmentů je jejich 3D struktura. Protože v současné době máme v rámci Protein Data Bank k dispozici nadkritické množství 3D strukturních dat biomakromolekul, můžeme se zaměřit na analýzu jejich fragmentů.

Ve svém příspěvku se budu věnovat jednotlivým krokům této analýzy - validaci, detekci, extrakci a charakterizaci strukturních fragmentů - a metodikám pro jejich realizaci. Dále budu prezentovat několik konkrétních aplikací vybraných analýz: porovnání kvality strukturních dat různých tříd ligandů [3], analýzu anatomie enzymatických tunelů [4] a vzhled do procesu aktivace proapoptotických proteinů na základě studia nábojového transferu [5].

Reference

- [1] Watson, J. D., Laskowski, R. A., Thornton, J. M. (2005) Predicting protein function from sequence and structural data. *Current opinion in structural biology*, 15, 275–284.
- [2] Eidhammer, I., Jonassen, I., Taylor, W. R. (2000) Structure comparison and structure patterns. *Journal of computational biology*, 7, 685–716.
- [3] Sehnal, D., Svobodová Vařeková, R., Pravda, L., Ionescu, C.-M., Geild, S., Horský, V., Jaiswal, D., Wimmerová, M., Koča, J. (2015) ValidatorDB: database of up-to-date validation results for ligands and non-standard residues from the Protein Data Bank. *Nucleic Acids Res.*, 43, D369–D375.

- [4] 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, 379.
- [5] Ionescu, C.-M., Svobodová Vařeková, R., Prehn, J.H.M., Huber, H.J., Koča, J. (2012) Charge profile analysis reveals that activation of pro-apoptotic regulators Bax and Bak relies on charge transfer mediated allosteric regulation. *PLoS Comput. Biol.*, 8, e1002565. doi:10.1186/s13321-015-0107-1.

L4-08

Altruistická metadynamika

Hošek P.¹, Toulcová D.¹, Bortolato A.², Spiwok V.¹

¹ *Ústav biochemie a mikrobiologie, Vysoká škola chemicko-technologická, Praha, Technická 3, 166 28 Praha 6*

² *Heptares Therapeutics Ltd, BioPark, Broadwater Road, Welwyn Garden City AL7 3AX, United Kingdom*

Výzkum a vývoj léčiv byl dlouhou dobu příliš náročný pro metody naložené na molekulární dynamice. V posledních letech však začaly simulační metody s vylepšeným vzorkováním, mezi něž patří například metadynamika či umbrella sampling, pronikat něj pronikat. Pro tento účel jsme vyvinuli rozšíření multiple walker metadynamiky nazvané Altruistická metadynamika, která je schopná odhadnout plochu volné energie několika systémů zároveň. Jako metadynamika v průběhu simulace kumuluje umělý potenciál ve formě Gaussových kopců, ale navíc sdílí omezené množství potenciálu s ostatními systémy, čímž je urychluje. Podle nastavení je možné buď zvýšit přesnost ploch volných energií či dříve dosáhnout konvergence. Popsaná metoda dosahuje nejvyšší efektivity při simulacích chemicky podobných systémů, kde sdílení potenciálu může být vyšší, například při screeningu potenciálních léčiv. Metoda byla testována na modelových plochách volné energie, při simulacích alanindipeptidu v různých silových polích a na konformacích monosacharidů D-hexopyranosové řady.



SEKCE 5

Aplikace

L5-01

Efektivní stabilizace proteinů a identifikace nových substrátů pro biotechnologie: hybridní experimentální a výpočetní přístupy

Brezovský J.^{1,2}, Daniel L.^{1,2}, Bednář D.^{1,2}, Bendl J.^{1,2,3}, Grulich M.⁴, Šebestová E.¹, Damborský J.^{1,2}

¹ *Loschmidtovy laboratoře, Ústav experimentální biologie a Centrum pro výzkum toxických látek v prostředí, Přírodovědecká fakulta, Masarykova univerzita, Kamenice 5/A13, 625 00 Brno*

² *Mezinárodní centrum klinického výzkumu, Fakultní nemocnice u sv. Anny v Brně, Pekařská 53, 656 91 Brno*

³ *Ústav informačních systémů, Fakulta informačních technologií, Vysoké učení technické v Brně, Božetěchova 1, 612 66 Brno*

⁴ *Laboratoř enzymových technologií, Mikrobiologický ústav AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha*

Uplatnění enzymů v biotechnologiích je často limitované jejich stabilitou v provozních podmínkách a rozsahem jejich substrátové specifity. Proteiny mohou být stabilizovány metodami proteinového inženýrství. Stávající postupy však často predikují pouze jednobodové mutace, které je dále třeba experimentálně charakterizovat a následně kombinovat. Experimentální konstrukce vícebodových mutantů je časově a finančně vysoce náročná. Obdobně náročným procesem je experimentální screening k nalezení nových substrátů. V rámci přednášky budou představeny dva bioinformatické a chemoinformatické projekty: i) stabilizace proteinů za pomoci přímého návrhu vícenásobných mutantů s využitím strukturních a sekvenčních predikcí a ii) hledání nových substrátů enzymů se známým reakčním mechanismem pomocí virtuálního screeningu, využívajícího geometrická kritéria k filtrování získaných výsledků. Oba projekty vedly k významnému zlepšení praktické použitelnosti studovaných enzymů.

L5-02

Aktivace genu NKD1 je známkou aberantní signalizace Wnt dráhy v nádorech epitelu: bioinformatická analýza velkých veřejně dostupných datasetů

Kolář M.¹, Stančíková J.¹, Krausová M.¹, Fafílek B.¹, Švec J.^{1,2}, Neroldová M.³, Dobeš J.¹, Horázná M.¹, Janečková L.¹, Vojtěchová M.¹, Oliverius M.³, Jirsa M.³, Kořínek V.¹

¹ Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

² 3. lékařská fakulta Univerzita Karlova v Praze, Šrobárova 50, 100 34 Praha 10

³ Institut klinické a experimentální medicíny, Vídeňská 1958/9, 140 21 Praha 4

Wnt signální dráha patří mezi základní signální mechanismy řídící vývoj tělesných orgánů. Její aktivita proto podléhá striktní regulaci, která probíhá na několika úrovních a zahrnuje jak pozitivní tak negativní zpětné vazby. V této studii [1] jsme se zaměřili na jednu z negativních zpětných vazeb, která působí skrz regulátor NKD1.

Expresí NKD1 je v karcinomech střeva a jater oproti normální tkáni zvýšena a, jak ukazujeme díky vlastní sbírce střevních polypů a karcinomů, je jasnou známkou tvorby novotvarů. U nádorů tlustého střeva (CRC) je u většiny pacientů Wnt signální dráha aktivována díky onkogenním mutacím v genu APC a NKD1 je exprimován v téměř všech CRC. Objevuje se již u dysplasií a míra jeho exprese souvisí s podtypem nádoru a progresí. Toto pozorování jsme potvrdili i na datasetu 366 CRC nádorů z projektu The cancer genome atlas (TCGA). U nádorů jater (HCC) je situace složitější. Bioinformatickou analýzou jsme zjistili, že i zde exprese NKD1 souvisí s aberantní aktivací Wnt signální dráhy: Na základě tří markerů aktivity této dráhy (AXIN2, EPHB2 a GS) jsme stratifikovali nádory jater do čtyř kategorií. Tři z těchto kategorií vykazovali nízkou aktivaci Wnt-signalizace, čtvrtá skupina (11 z 24 vzorků) měla silnou aktivitu signální dráhy (HCC_Wnt) a vysokou expresi NKD1. Vzorky z této skupiny měly ve většině případů mutace v třetím exonu genu kódujícího β -catenin, jeden ze základních mediátorů Wnt-signální dráhy. Z jedenácti nádorů jater, které měli mutace v tomto exonu, devět spadalo právě do této skupiny. K ověření našich pozorování jsme zpracovali HCC dataset z TCGA (197 vzorků). Shlukování na základě Wnt signatury identifikovalo tři podskupiny nádorů, z nichž jedna měla silně aktivovanou Wnt signální dráhu (HCC_Wnt, 37 vzorků). U většiny z těchto nádorů (33 vzorků) jsme pozorovali mutace v genech kritických pro aktivaci Wnt signální dráhy: β -catenin (26 vzorků), APC, AXIN1 a AXIN2.

Ačkoliv u studovaných nádorů je aberantní aktivace Wnt dráhy primárně způsobena onkogenními mutacemi, nemůžeme vyloučit ani vliv epigenetických změn genů kódujících členy Wnt signální dráhy. V rámci tohoto „multiple-hit“ konceptu jsme se

zaměřili na metylační změny ve vzorcích CRC z TCGA datasetu a studovali vliv těchto změn na transkripci NKD1. Nalezli jsme jeden lokus, jehož hypomethylace je významně asociována se zvýšenou expresí NKD1 v nádorech. V souhrnu, naše výsledky ukazují, že exprese NKD1 může být použita ke zpřesnění diagnózy CRC a HCC, především ve vztahu k aberantní aktivaci Wnt signální dráhy.

Tato práce byla podpořena z následujících grantů: GAČR: P305/11/1780, ústavní podpora poskytnutá AV ČR: RVO 68378050, projekt BIOCEV: CZ.1.05/1.1.00/02.0109 pro V.K., ústavní podpora poskytnutá MZ ČR: 00023001 pro M.O., M.N. a M.J. a MZ VES: NV16-29032A pro M.K.

Reference

- [1] Stančíková et al., *Cellular Signalling* 27 (2015) 245–256

L5-03

Identifikace a lokalizace mutací v repetitivní oblasti genu *MUC1*

Přistoupilová A.^{1,2,3}, Stránecký V.^{1,2,3}, Hodaňová K.^{1,3}, Hartmannová H.^{1,3},
Piherová L.^{1,3}, Vrbacká Čížková A.¹, Živná M.¹, Kmoch S.^{1,2,3}

¹ Ústav dědičných metabolických poruch, 1. lékařská fakulta, Univerzita Karlova
v Praze, Ke Karlovu 2, 128 08 Praha 2

² Biotechnologické a biomedicínské centrum Akademie věd a Univerzity Karlovy ve
Vestci, Průmyslová 595, 252 42 Vestec

³ Národní centrum lékařské genomiky, Ke Karlovu 2, 128 08 Praha 2

Rozvoj nových metod sekvenace genomu (Next Generation Sequencing Technologies, NGS), v posledních letech velmi zjednodušil analýzu lidského genomu. I přesto ale zůstávají některé jeho části obtížně přístupné. Jedním z takových příkladů je gen *MUC1*, kódující transmembránový glykoprotein mucin-1, jehož kódující sekvence je GC bohatá (82%) a sestává z variabilního počtu (25-120) polymorfních tandemových repetitivních (VNTR) o délce 60 bází. Posunové mutace v tomto genu vedou k syntéze abnormálního proteinu, který má začátek sekvence shodný s původní sekvencí mucinu-1 a koncovou část naprosto unikátní. Syntéza, přítomnost a toxické působení mutovaného proteinu v tubulárních buňkách vede k postupné ztrátě tubulárních funkcí a medulárnímu cystickému onemocnění ledvin (MCKD1), které ve většině případů končí renálním selháním s nutností transplantace ledvin. K včasné diagnostice MCKD1 je klíčová identifikace kauzálních mutací a jejich lokalizace ve VNTR. Jedním z faktorů ovlivňujících dobu nástupu a tíži onemocnění je totiž s největší pravděpodobností délka mutovaného proteinu, která závisí na celkovém počtu tandemových repetitivních a lokalizaci kauzální mutace v rámci VNTR. Identifikace a lokalizace mutací v *MUC1* je obtížná. Vzhledem k délce a povaze repetitivní oblasti nelze použít klasické Sangerovo sekvenování a NGS sekvenování je limitováno obtížnou mapovatelností sekvenovaných fragmentů.

Naše metoda, založená na amplifikaci VNTR oblasti pomocí Long Range PCR, jejím sekvenování na sekvenátoru Illumina HiSeq a specifické bioinformatické analýze získaných dat, umožňuje identifikaci mutací v genu *MUC1*. Tuto metodu jsme úspěšně otestovali na vzorcích s již známými kauzálními mutacemi a dále se nám pomocí této metody podařilo identifikovat další dvě nové mutace. Pro lokalizaci mutací ve VNTR oblasti v současné době testujeme metodu nanopórového sekvenování na přístroji MinION od firmy Oxford Nanopore, který umožňuje sekvenování jednotlivých molekul až o délce desítek kilobází.

L5-04

Evoluční a speciální genomika na modelu dvou druhů slavíků

Reifová R.¹, Mořkovský L.¹, Janoušek V.¹, Reif J.², Rídl J.³, Pačes J.³,
Choleva L.⁴, Janko K.⁴, Nachman M. W.⁵

¹ *Katedra zoologie, Přírodovědecká fakulta, Univerzita Karlova v Praze, Viničná 7, 128 00 Praha 2*

² *Ústav pro ochranu životního prostředí, Přírodovědecká fakulta, Univerzita Karlova v Praze, Benátská 2, 128 01 Praha 2*

³ *Ústav molekulární genetiky, Akademie věd České republiky, Vědeňská 1083, 142 20 Praha 4*

⁴ *Ústav živočišné fyziologie a genetiky, Akademie věd České republiky, Rumburská 89, 277 21 Liběchov*

⁵ *Museum of Comparative Zoology and Integrative Biology, University of California, 3101 Valley Life Science Building, Berkeley, California 947 20, USA*

Speciace čili vznik nových druhů je klíčový evoluční proces, jehož porozumění je důležitým cílem evoluční biologie. Nové genomické přístupy, které lze aplikovat i na divoce žijící organismy, představují unikátní možnost nahlédnout do genetické podstaty vzniku druhů. Náš tým se dlouhodobě zabývá speciací na modelu dvou nedávno divergovaných druhů pěvců, slavíka obecného a slavíka tmavého. Oba druhy v oblasti sekundárního kontaktu příležitostně hybridizují a vytvářejí sterilní F1 hybridní samice, avšak fertilní F1 hybridní samce. Mezi oběma druhy dochází ke genovému toku, který je významně nižší na pohlavním chromosomu Z než na autosomech. Porovnáním transkriptomů několika jedinců obou druhů slavíků a jejich mapováním na genom zebřičky, jsme identifikovali genomické ostrovy zvýšené diferenciace, o kterých se předpokládá, že nesou geny odpovědné za reprodukční izolaci. Diferenciace byla obecně vyšší na chromosomu Z než na autosomech. Funkční analýza genů v ostrovech diferenciace ukázala, že jsou mezi nimi častější geny s funkcí v samičí oogenezi. Naše studie je jednou z prvních, která odhalila kandidátní geny hybridní sterility samic.



Poster Session

Poster session je sponzorována
Národní infrastrukturou chemické biologie
CZ-OPENSREEN.



P-01

Identifikácia a štúdium domén ľudského ryanodínového receptora 2 zodpovedných za ochorenia srdca

Bauerová V.¹, Borko L.¹, Benko M.¹, Bauer J.¹, Zahradníková A.^{1,2}, Ševčík J.¹, Kutejová E.¹

¹ Oddelenie biochémie a štruktúrnej biológie, Ústav molekulárnej biológie SAV, Dúbravská cesta 21, 845 51 Bratislava

² Oddelenie pre výskum svalových buniek, Ústav molekulárnej fyziológie a genetiky SAV, Dúbravská cesta 9, 840 05 Bratislava

Ryanodínové receptory sú doteraz najväčšie známe iónové kanály [1]. Nachádzajú sa v membráne sarkoplazmatického retikula a sprostredkovávajú prenos vápnikových katiónov do cytoplazmy, čím spúšťajú kaskádu reakcií podmieňujúcich kontrakciu svalu [2]. U cicavcov boli identifikované tri izoformy tohto iónového kanála, ktoré sa prednostne exprimujú v kostrovom svalstve (RyR1), v myokarde (RyR2) a rôznych orgánoch (RyR3). Dysfunkcia RyR kanála spôsobuje závažné svalové ochorenia – myopatie (dysfunkcia RyR1) [3], tachykardie a arytmie (dysfunkcia RyR2) [4].

V našom laboratóriu sa zaoberáme predovšetkým štruktúrnymi štúdiami izoformy RyR2 z človeka a využívame prepojenie „*in silico*“ prístupu s laboratórnymi experimentmi. Pomocou bioinformatických analýz sme v géne RyR2 identifikovali prítomnosť 14 domén [5,6], ktoré sme následne exprimovali a purifikovali v bakteriálnom systéme *E.coli*. Najdlhší fragment, zahŕňajúci prvé tri N-terminálne domény, sme úspešne kryštalizovali [7] a určili jeho terciárnu štruktúru pomocou RTG- a SAXS štruktúrnej analýzy [8]. „*In silico*“ analýza vybraných aminokyselín, mutácie ktorých podmieňujú závažné srdcové ochorenia nám pomáha porozumieť ich úlohe v štruktúre a funkcii RyR2 kanála.

Táto práca bola finančne podporená projektmi VEGA č. 2/0140/16 a APVV-0628-10.

Reference

- [1] Takeshima H, Nishimura S, Matsumoto T, et al., 1989. Nature 339, p. 439-445.
- [2] Meissner G, 2004. Cell Calcium 35, p. 621-628.
- [3] Shaaban S, Ramos-Platt L, Gilles FH, Chan WM, Andrews C, De Girolami U, Demer J, Engle EC. JAMA Ophthalmol. 2013 Dec;131(12):1532-40.
- [4] Medeiros-Domingo A., Bhuiyan Z.A., Tester D.J., Hofman N., Bikker H., van Tintelen J.P., Mannens MM, Wilde AA, Ackerman MJ. Journal of American College of Cardiology, vol: 54(22). 2009. p.2065-2074.
- [5] Bauerová-Hlinková V., Hostinová E., Gašperík J., Beck K., Borko L., Lai F.A., Zahradníková A., Ševčík J. Protein Expr. Purif. 71, 2010 p.33–41.

- [6] V. Bauerová-Hlinková, J. Bauer, E. Hostinová, et al. 2011. In: Bioinformatics – Trends and Methodologies. Edited by M.A. Mahdavi. Printed in Croatia. Intech, Rijeka, Croatia, p. 325-352
- [7] Borko L., Košťan J., Zahradníková A., Pevala V., Gašperík J., Hostinová E., Urbáníková L., Djinovic-Carugo K., Bauerová-Hlinková V., Ševčík J. Protein & Peptide Letters 20 (11). 2013. p. 1211-6.
- [8] Borko L., Bauerová-Hlinková V., Hostinová E., Gašperík J., Beck K., Lai F.A., Zahradníková A., Ševčík J. 2014. Acta Cryst. D70, p.2897-2921.

P-02

Molekulární dokování růstových regulátorů

Bazgier V.¹, Berka K.²

¹ Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra fyzikální chemie, CR Haná, 17. listopadu 12, 771 46 Olomouc

² Univerzita Palackého v Olomouci, Přírodovědecká fakulta, Katedra fyzikální chemie, RCPTM, 17. listopadu 12, 771 46 Olomouc

Vzhledem k současnému omezení a tlaku proti výzkumu a využívání geneticky modifikovaných (GMO) rostlin v rámci EU se evropský výzkum v agronomické oblasti zaměřil na možnosti regulace růstu rostlin a tedy i zvyšování výnosů biomasy pomocí "chytré zelené chemie". Regulátory růstu, také nazývané fytohormony, v tomto konceptu hrají rozhodující roli, neboť ovlivňují životní cyklus každé rostliny i při daleko nižších množstvích než klasická hnojiva. Mezi nejdůležitější fytohormony můžeme řadit brasinosteroidy, cytokininy a další. Díky nedávným pokrokům ve strukturní biologii rostlin je dnes možné zjednodušit hledání nových kandidátních látek pomocí metod racionálního designu. Metody molekulového dokování, běžně využívané v případě návrhu léčiv, nám tak otevírají nové možnosti jak relativně levnou cestou vyhledávat nové, potenciálně zajímavé chemické látky, které interagují s vybranými rostlinnými receptory a molekulárními cíli. Tento poster si klade za cíl seznámit čtenáře s možnostmi drug-design metod při vývoji nových růstových regulátorů na příkladech brasinosteroidů [1,2] a cytokininů [3,4].

Tato práce je podpořena granty GAČR 15-08202Y (V.B.) a 15-19266S (K.B.). V.B. byl podpořen studentským grantem UP (IGA_PrF_2016_028).

Reference

- [1] Kvasnica M, Oklestkova J, Bazgier V, Rarova L, Berka K, Strnad M: Biological activities of new monohydroxylated brassinosteroid analogues with a carboxylic group in the side chain. *Steroids*, 85, 58-64, 2014.
- [2] Kvasnica M, Oklestkova J, Bazgier V, Berka K, Budesinsky M, Rárová L, Kořínková P, Strnad M: Design, synthesis and biological activity of new brassinosteroid analogues with p-substituted phenyl group. *submitted*
- [3] Nisler J, Končítiková R, Kopečný D, Zatloukal M, Bazgier V, Berka K, Zalabák D, Briozzo P, Strnad M, Spíchal L: Novel thidiazuron-derived inhibitors of cytokinin oxidase/dehydrogenase. *submitted*.
- [4] Gallová L, Gandalla Z, Bazgier V, Berka K, Spíchal L.: 9-deaza Cytokinin analogues as a tool to study Arabidopsis Histidine Kinase binding properties. *in preparation*.

P-03

Studium molekulárních základů evoluce feromonové komunikace hmyzu prostřednictvím RNA-seq

Buček A.¹, Vogel H.², Brabcová J.¹, Prchalová D.¹, Kindl J.¹, Valterová I.¹, Pichová I.¹

¹ Ústav organické chemie a biochemie, AV ČR, Flemingovo náměstí 2, 166 10 Praha 6

² Max Planck Institute for Chemical Ecology, Hans-Knoell-Strasse 8, Jena, Německo

Výzkum genomů a transkriptomů, který býval doménou konsorcií vědeckých týmů zaměřených na studium modelových organismů, byl s rozvojem sekvenačních technik nové generace (NGS) zpřístupněn i pro studium nemodelových druhů. Hmyz, který představuje abundatní, různorodou a ekonomicky relevantní skupinu čítající nejméně 1 000 000 popsaných druhů, byl před rozšířením NGS studován na transkriptomické úrovni pouze omezeně. Pro pochopení molekulárních mechanismů evoluce feromonové komunikace, jenž představuje druhově specifický komunikační kanál koordinující klíčové aspekty života hmyzu, např. reprodukci, jsme studovali geny zapojené v biosyntéze feromonů u třech evropských druhů čmeláků, čmeláka zemního (*Bombus terrestris*), ekonomicky významného opylovače skleníkových rostlin, čmeláka hájového (*B. lucorum*) a čmeláka skalního (*B. lapidarius*), kteří se navzájem zásadně liší chemickým složením samčího feromonu.

Prostřednictvím RNA-seq sekvenování a *de novo* sestavení transkriptomů feromonových žláz a referenčních tkání těchto tří druhů čmeláků bylo získáno více než 30 000 kontigů pro každý druh z nichž 12 000 – 13 000 bylo anotováno výrazy Gene Ontology (GO) popisující jejich predikované molekulární funkce a biologické procesy, ve kterých jsou geny zapojeny. V této skupině predikovaných genů bylo u jednotlivých studovaných druhů čmeláků identifikováno 400 – 1 200 genů abundatně a přednostně exprimovaných ve feromonové žláze. Vizuální inspekci těchto kandidátů byla popsána komplexní síť genů zapojených v biosyntéze derivátů mastných kyselin a terpenových látek, které představují hlavní feromonové složky, a byly identifikovány geny, jejichž rozdílná míra exprese s největší pravděpodobností přispívá k druhově specifickému složení samčích feromonů u studovaných druhů [1, 2, a nepublikované výsledky].

Tato práce byla podpořena projektem LO1302 (MŠMT) a grantem GAČR č. 15-06569S.

Reference

- [1] Prchalová D, Buček A, Brabcová J, Žáček P, Kindl J, Valterová I, Pichová I: *ChemBioChem* (2016), 17:260–267
- [2] Buček A, Brabcová J, Vogel H, Prchalová D, Kindl J, Valterová I, Pichová I: *Insect Mol Biol* (2016, in press).

P-04

Dolbico: Databáze lokálních konformací biomolekulČech P.¹, Svozil D.¹, Černý J.², Schneider B.²¹ *VŠCHT Praha, Laboratoř informatiky a chemie, Technická 5, 166 28 Praha 6*² *AV ČR, v. v. i., Biotechnologický ústav, Průmyslová 595, 252 50 Vestec*

Tímto příspěvkem představíme Dolbico (Database of Local Biomolecular Conformations) – databázi, která uchovává a poskytuje výsledky klasifikace lokálních konformací struktur deoxyribonukleových kyselin (DNA). Vývoj této databáze bezprostředně navazuje na náš předchozí projekt [1], ve kterém jsme pomocí metod strojového učení vyvinuli postup pro klasifikaci dinukleotidů v molekulách DNA [2]. Jako zdroj strukturních dat využíváme databázi Protein Data Bank (PDB) [3]. Databáze Dolbico se s databází PDB pravidelně synchronizuje a zcela automaticky provádí třídění dinukleotidových konformací molekul DNA podle našeho systému klasifikace. Tímto významně obohacuje základní strukturní data o další informace o uspořádání molekul DNA na lokální úrovni. K databázi jsme vytvořili uživatelské rozhraní formou webové aplikace, které nabízí několik nástrojů, jejichž pomocí lze strukturní data a výsledky klasifikace jednoduše analyzovat a pokročilými dotazy systematicky prohledávat. Veškeré výsledky dotazů jsou vizualizovány přímo v prostředí webové aplikace nebo mohou být uloženy do souboru pro další zpracování. V budoucnu plánujeme do databáze zahrnout též klasifikaci lokálních konformací RNA a proteinů, čímž se Dolbico stane ultimátním zdrojem informací o strukturách biomolekul na lokální úrovni. Databáze Dolbico je volně dostupná na adrese <http://www.dolbico.org>.

Reference

- [1] Čech P., et al., Automatic workflow for the classification of local DNA conformations. *BMC Bioinformatics*, 2013, 14(1).
- [2] Svozil D., et al., DNA conformations and their sequence preferences. *Nucleic Acids Research*, 2008, 36 (11): 3690-3706.
- [3] Berman H. M., et al., The Protein Data Bank. *Nucleic Acids Research*, 2000, 28(1): 235-242.

P-05

Návrh chemické knihovny potenciálních agonistů glukokortikoidního receptoru pomocí algoritmu Molpher

Čmelo I.^{1,2}, Svozil D.^{1,2}, Bartůněk P.², Sedlák D.²

¹ CZ-OPENSSCREEN & Laboratoř informatiky a chemie, Fakulta chemické technologie, VŠCHT Praha, Technická 5, 166 28 Praha 6

² CZ-OPENSSCREEN, Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

Molpher je algoritmus umožňující mapovat chemický prostor mezi zadanou dvojicí vstupních struktur. Zakládá se na postupném převádění počáteční struktury na strukturu cílovou pomocí jednoduchých strukturních obměn, jako je např. mazání či záměna atomů nebo vazeb. Mezi startovací a cílovou strukturou tak vzniká „cesta“ chemickým prostorem sestávající z tzv. morfů, sloučenin, které v různých poměrech kombinují vlastnosti obou výchozích struktur. Takto vzniklé „cesty“ je možno použít k mapování zajímavých oblastí chemického prostoru a v nich obsažené morfy lze pak použít pro návrh chemických knihoven.

Přednáška pojednává o pilotním nasazení algoritmu Molpher při návrhu chemické knihovny agonistů glukokortikoidního receptoru (GR). Výchozím bodem byly struktury již známých agonistů GR, získané z databáze ChEMBL a z primárního screenu provedeného s využitím prostředků infrastruktury CZ-OPENSSCREEN. Struktury agonistů byly použity jako vstupní dvojice do algoritmu Molpher, který tak zmapoval jejich mezilehlý chemický prostor. Mapováním vytvořené morfy byly poté podobnostně vyhledány v databázi komerčně dostupných látek ZINC. Výsledkem je chemická knihovna komerčně dostupných látek pocházejících z oblastí chemického prostoru vytyčených známými ligandy GR.

P-06

Uživatelsky orientované webové nástroje Loschmidtových laboratoří

Štourač J.^{1,2}, Bendl J.^{1,2,3}, Musil M.^{1,2,3}, Pavelka A.¹, Šebestová E.¹, Vávra O.¹, Brezovský J.^{1,2}, Damborský J.^{1,2}

¹ *Loschmidtovy laboratoře, Ústav experimentální biologie a Centrum pro výzkum toxických látek v prostředí, Přírodovědecká fakulta, Masarykova univerzita, Kamenice 5/A13, 625 00 Brno*

² *Mezinárodní centrum klinického výzkumu, Fakultní nemocnice u sv. Anny v Brně, Pekařská 53, 656 91 Brno*

³ *Ústav informačních systémů, Fakulta informačních technologií, Vysoké učení technické v Brně, Božetěchova 1, 612 66 Brno*

Pro vývoj nových biotechnologií jsou nezbytné laboratorní experimenty, které jsou finančně a časově náročné. Rozvoj moderních výpočetních metod a narůstající výkon počítačů umožňuje nahrazovat některé experimenty počítačovými analýzami a tím redukovat počet laboratorních pokusů na nezbytné minimum. V příspěvku představíme sadu webových nástrojů pro proteinové inženýrství, predikci vlivu mutací na lidské zdraví a identifikaci transportních cest v proteinech. **HotSpot Wizard** je webová aplikace implementující čtyři různé strategie proteinového inženýrství sloužící pro predikci aminokyselinových reziduí vhodných k mutagenезi a konstrukci knihoven proteinových mutantů se zvýšenou aktivitou, změněnou specifitou nebo stabilitou. **PredictSNP** a **PredictSNP2** jsou meta-nástroje pro binární rozhodování škodlivosti aminokyselinových a nukleotidových substitucí v lidském genomu. Nástroje jsou doplněny konsenzuální predikcí, uniformními odhady důvěryhodnosti předpovězeného efektu a anotací z relevantních databází. **Caver** představuje rodinu nástrojů pro analýzu a vizualizaci tunelů a kanálů ve statických i dynamických proteinových strukturách. Caver je poskytován ve stand-alone verzi, jako plug-in pro PyMOL, vizualizační software s grafickým rozhraním a nově jako webová aplikace. Všechny prezentované nástroje jsou dostupné jako webová služba s intuitivním uživatelským rozhraním, které je činí snadno přístupnými široké uživatelské komunitě:

Reference

HotSpot Wizard – <http://loschmidt.chemi.muni.cz/hotspotwizard/>
PredictSNP – <http://loschmidt.chemi.muni.cz/predictsnp/>
PredictSNP2 – <http://loschmidt.chemi.muni.cz/predictsnp2/>
Caver – <http://loschmidt.chemi.muni.cz/caver/>

P-07

Computer aided drug design towards new inhibitors of Kynurenine Aminotransferase

Počítačový návrh inhibitorů kynurenin aminotransferázy

Dehaen W.¹, Maryška M.¹, Kuchař M.¹, Svozil D.¹

¹ *Vysoká škola chemicko-technická v Praze, Technická 5, 166 28 Praha 6*

Kynurenine aminotransferase (KAT) is a promising target for the development of future central nervous system(CNS) medication. KAT is an enzyme catalyzing the transformation of kynurenine into kynurenic acid, a metabolite implicated in CNS disorders including schizophrenia. A limited amount of inhibitors of KAT are known but they suffer from drawbacks such as low chemical diversity, potential toxicity and difficulty of preparation. In this study, in silico methods were used to predict new distinct chemotypes that inhibit KAT. Candidate inhibitors were selected, refined and scored using pharmacophore modeling of known inhibitors, bulk virtual screening and rational drug design. Molecular docking of the candidate inhibitors and comparison with binding modes of known inhibitors were used to cross-validate the novel compounds. This resulted in a list of highly promising drug candidates. The obtained molecules have superior qualities such as chemical diversity and synthetic and commercial availability and thus facilitate the development of superior, modern therapies for CNS disorders.

P-08

Nástroj pro generování elementárního složení z dat hmotnostních spektrometrů s vysokým rozlišením

Doležalová M.^{1,2}, Cehák J.^{1,2}, Fesl J.^{1,2}

¹ Jihočeská univerzita v Českých Budějovicích, Přírodovědecká fakulta, Ústav aplikované informatiky, Branišovská 1645/31A, 370 05 České Budějovice

² Biologické centrum AV ČR, v. v. i., Entomologický ústav, Laboratoř analytické biochemie a metabolomiky, Branišovská 1160/31, 370 05 České Budějovice

Identifikace neznámých sloučenin je jedním z nejdůležitějších úkolů vědeckého výzkumu pomocí kapalinové chromatografie spojené s hmotnostní spektrometrií s vysokým rozlišením (Liquid chromatography–High resolution mass spectrometry, LC-HRMS). Výstupem hmotnostního spektrometru jsou chromatogramy popisující jednotlivé sloučeniny v měřeném vzorku, zejména jejich retenční čas na chromatografické koloně a molekulovou hmotnost. Pokud je molekulová hmotnost měřena s nízkou přesností, nelze spolehlivě určit bez dalších znalostí, o kterou sloučeninu, tj. který sumární vzorec, se jedná. HRMS spektrometry mají rozlišovací schopnost tak vysokou, že je možné rozlišit i jednotlivé izotopické příspěvky prvků sloučeniny, což velmi významně přispěje k její identifikaci. Nástroj pro generování elementárního složení z dat hmotnostních spektrometrů s vysokým rozlišením (IoM, Interpreter of Mass) slouží k predikci sumárních vzorců odpovídajících přesně naměřené molekulové hmotnosti. Používá k tomu zejména přesnou molekulovou hmotnost, izotopickou obálku, vzájemné poměry prvků a další heuristická pravidla. Architektura IoM byla vyvinuta v programovacím jazyce C++ s využitím paralelních procesorů. Podporuje nativní import dat z různých formátů prodejců (Thermo, Bruker) a je součástí softwaru Metabolite Mapper, který byl vyvinut pro potřeby laboratoře analytické chemie a metabolomiky Biologického centra AV ČR.

P-09

Analýza zmesných vzoriek vo forenzej genetike pomocou metód sekenovania novej generácie

Nagyová E.¹, Žižkovičová P.¹, Sviežená B.², Choma A.², Ďuriš F.^{1,3,4}, Budiš J.^{3,4}, Minárik G.^{1,4}

¹ *Prírodovedecká fakulta Univerzity Komenského, Ilkovičova 6, 842 15 Bratislava*

² *Oddelenie biológie a genetickej analýzy, Kriminálny ústav, Sklabinska 1, 831 06 Bratislava*

³ *Fakulta matematiky, fyziky a informatiky Univerzity Komenského, Mlynská dolina F1, 842 48 Bratislava*

⁴ *Geneton s. r. o., Galvaniho 7, 821 04 Bratislava*

Analýza zmesných forenzných DNA vzoriek je aj v súčasnosti výrazne obmedzená a interpretácia získaných profilov je stále problematická. Štandardne používané metódy DNA profilizácie založené na genotypizácii STR markerov v kombinácii s fragmentovou analýzou alebo Sangerovým sekvenovaním nie sú dostačujúce, najmä v prípadoch, keď je kvalita alebo kvantita analyzovanej DNA nízka. Rovnako je tomu v prípade, že DNA vo vzorke pochádza od viac ako dvoch prispievateľov alebo keď sú tieto vo veľmi nevyváženom pomere. Cieľom našej štúdie bolo potvrdiť možnosť využitia nových metód multiparalelného sekvenovania pre forenznú analýzu zmesných vzoriek a implementovať pracovný postup pre takúto analýzu. Ako markery sme použili mikrohplotypy, ktoré môžu byť veľmi informatívnym zdrojom pri riešení mnohých forenzných otázok, vrátane DNA profilizácie dvoch alebo viacerých prispievateľov v zmesnej vzorke.

Pripravili sme arteficiálne zmesné vzorky spojením dvoch vzoriek DNA, ktoré sme vybrali na základe ich DNA profilov vo vybraných markeroch. Množstvo DNA minoritného prispievateľa vo vzorke bolo 1%; 2,5%; 5%; 10% a 25%. Pomocou PCR sme pripravili špecifické amplikóny zodpovedajúce 12 vybraným mikrohplotypovým markerom (Hap1, Hap2, HAP3, HAP4, HAP5, shortHAP1, shortHAP3, shortHAP5, mikroHAP49, mikroHAP61, mikroHAP46, mikroHAP315), ktoré boli použité na prípravu sekvenačných knižníc. Knižnice boli sekvenované na prístroji Ion Torrent™ PGM®.

S využitím bioinformatických nástrojov, ktoré v tomto príspevku prezentujeme, bol stanovený pomer majoritnej a minoritnej zložky DNA. Pri získanom a očakávanom pomere v zmesných vzorkách sme pozorovali rovnaký trend pri poklese podielu signálu minoritného prispievateľa na základe čoho sme mohli verifikovať platnosť údajov získaných sekvenačnou metódou.

P-10

Shluková analýza lidské starodávné mtDNA: 30 000 let pod maternální dědičností

Ehler E.^{1,2,3}, Juras A.¹

¹ *Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University, ul. Umultowska 89, 61-614, Poznan, Poland*

² *Forensic DNA Service, Janovského 18, 170 00 Prague 7*

³ *Department of Biology and Environmental Studies, Charles University in Prague, Faculty of Education, Magdaleny Rettigove 4, 116 39 Prague*

Stanovení příbuznosti lidských populací a jejich fylogenetických vztahů je jedním z hlavních témat molekulární antropologie v posledních letech. Zejména rozvoj technik pro zpracování starodávné DNA (ancient DNA, aDNA) se podílí na nových možnostech výzkumu historických i prehistorických lidských populací a jejich poměrem k dnešním moderním populacím. Z hlediska bioinformatiky je v této oblasti výzkumu pravděpodobně nejpoužívanějším přístupem shluková analýza (cluster analysis) získaných dat.

Materiál: V posledních třech letech byly publikovány celé mitochondriální genomy z mnoha populací, jejichž časová distribuce sahá od začátku svrchního paleolitu až po přelom letopočtu. Zde prezentujeme analýzu vybraného setu populací a jedinců, kteří byli vyselektováni zejména s ohledem na vývoj zalidnění v oblasti střední Evropy a procesy, které se podílely na vývoji moderní středoevropské populace během neolitu a doby bronzové. Jedná se o populace mesolitických lovců-sběračů, raných zemědělců, zemědělské populace ze středního neolitu, pozdně neolitické populace střední a východní Evropy, a dále populace z doby bronzové a doby železné.

Metody: Použili jsme různé metody a varianty hierarchického shlukování, jednak s přístupem bottom-up (agglomerative clustering), jednak s přístupem top-down (k-means).

Výsledky: Jelikož obecný přístup k identifikování shluků/klastrů a tím i příbuznosti pre/historických populací je postaven na znalostech zejména jejich archeologických, případně antropologických a historických, parametrů, zaměřili jsme se naopak na agnostický (unsupervised) přístup pomocí shlukovacích algoritmů. Podařilo se nám replikovat některé známé dělení historických populací (např. založené na frekvenci mtDNA haploskupin, na způsobu obživy), a rovněž objevit některé nové vztahy mezi prehistorickými populacemi Evropy (ujasnění vztahu mezi raně neolitickými a pozdně neolitickými populacemi). Naše výsledky jsou interpretovatelné pouze s ohledem na to, že pocházejí z analýzy maternálně dědičných markerů.

Tento výzkum by podpořen projekty GAČR č. 14-36938G a NCN Symfonia II UMO-2014/12/W/NZ2/00466.

P-11

Štatistický model pre anotáciu genomických mikrosatelitov

Gazdarica J.^{1,3}, Budiš J.^{2,3}, Ďuriš F.³, Radvanszky J.^{1,3}, Szemes T.^{1,3}

¹ *Prírodovedecká fakulta, Katedra molekulárnej biológie, Univerzita Komenského, Mlynská dolina, Bratislava*

² *Fakulta matematiky, fyziky a informatiky, Katedra Informatiky, Univerzita Komenského, Mlynská dolina, Bratislava*

³ *Geneton s.r.o., Galvaniho 7, Bratislava*

Mikrosatelity sú krátke, opakujúce sa úseky DNA dĺžky 2 až 6 bp, ktoré sa vyznačujú vysokou variabilitou medzi jedincami. Rozvoj sekvenátorov novej generácie (NGS) umožnil ich hromadnú genotypizáciu, výstupné čítania ale často nedokážu zachytiť reálny počet opakovaní motívu. Problémom sú chyby pri sekvenovaní, najmä bodové mutácie a preskočenie výskytu motívu. Sekvenovaním navyše nedokážeme zachytiť nadmerne dlhé motívy, ktoré môžu súvisieť s patologickými prejavmi jedinca.

V našej štúdii využívame na problém identifikácie a anotácie čítaní s mikrosatelitmi metódu profilových skrytých markovovských modelov. Navrhujeme distribúciu, ktorá modeluje chyby v počte opakovaní motívov spôsobené pridaním alebo vynechaním opakovania motívu pri sekvenačnom behu. Cieľom našej práce je odlíšenie homozygotných a heterozygotných aliel, ktoré je potrebné pri detekcii špecifických genetických chorôb, napríklad myotickej dystrofie.

P-12

ChrgDescCalc.py – nástroj na výpočet nábojových deskriptorů

Hejret V.¹, Geidl S.¹, Svobodová Vařeková R.¹, Koča J.¹

¹ *Masaryk University, National Centre for Biomolecular Research, Faculty of Science and CEITEC - Central European Institute of Technology, Kamenice 5, 625 00 Brno*

Nábojové deskriptory jsou podskupinou molekulových deskriptorů [1,2], které v chemoinformatice používáme především k predikci aktivit a fyzikálně chemických vlastností molekul. Nábojové deskriptory mají tu výhodu, že využívají pro svůj výpočet parciální atomové náboje, které popisují rozložení elektronů v molekule a můžeme díky nim získat velký přínos při popisu některých důležitých vlastností. Přestože existuje několik kvantově mechanických i empirických metodik pro výpočet nábojů, většina současných softwarových nástrojů pro výpočet deskriptorů využívá pouze jednoduché Gasteiger-Marsili náboje. Navíc uvedené softwarové nástroje nabízí pouze zlomek z dostupných nábojových deskriptorů. Z tohoto důvodu jsme se rozhodli vytvořit software pro výpočet nábojových deskriptorů, založených na libovolných atomových nábojích. Tento nástroj umožňuje výpočet 68 různých nábojových deskriptorů, jejichž definice byly převzaty především z literatury [1]. Software je napsán v jazyce Python a umožňuje zpracovat tisíce běžných organických molekul během několika sekund. Rád bych představil výsledky porovnání deskriptorů využívající kvantově mechanické a empirické náboje. Dále se nám pomocí vybraných deskriptorů podařilo vytvořit kvalitní QSPR modely pro predikci významných fyzikálně chemických vlastností – rozdělovacího koeficientu logP a disociační konstanty pK_a pro molekuly fenolů.

Reference

- [1] Todeschini R, Consonni V: Handbook of Molecular Descriptors. Wiley-VCH Verlag GmbH, Weinheim, Germany; 2008.
- [2] Todeschini R, Consonni V: Molecular Descriptors for Chemoinformatics, Volume 41 (2 Volume Set). John Wiley & Sons; 2009.

P-13

Analýza genové exprese u novorozenců narozených v prostředí s různou úrovní znečištění ovzduší

Hoňková K.¹, Rössnerová A.¹, Rössner P. Jr.¹, Pavlíková J.¹, Gmuender H.², Švecová V.¹, Pulkrabová J.³, Hajšlová J.³, Velemínský M.⁴, Šrám R. J.¹

¹ Ústav experimentální medicíny AV ČR, v. v. i., v Praze

² Genedata, Basel, Švýcarsko

³ Vysoká škola chemicko-technologická v Praze

⁴ Jihočeská univerzita v Českých Budějovicích

Severomoravský kraj patří k částem České republiky s největší mírou znečištění ovzduší prachovými částicemi a benzo[a]pyrenem způsobené převážně těžkým průmyslem. Těto specifické situace se využívá pro studium dopadu znečištění ovzduší na genom novorozenců narozených ve dvou expozičně odlišných lokalitách – Karvině (expozičně zatížené) a v Českých Budějovicích (kontrolní oblast). Cílem studie je určit biochemické dráhy a transkripty, které se exprimují odlišně v obou lokalitách a ve dvou odběrových obdobích (léto a zima).

Ke studii bylo využito celkem 231 vzorků pupečníkové krve a z leukocytů byla izolována RNA, která byla dále upravena *in vitro* transkripcí, hybridizována na čípech HumanHT-12 v3.0 Gene Expression BeadChip (Illumina) a čipy byly skenovány na přístroji iScan (Illumina). Redukovaná microarray data ze softwaru GenomStudio (Illumina) byla zpracována v prostředí R a na data byl aplikován lineární model pro identifikaci signifikantních transkriptů. Devět vybraných transkriptů bylo ověřeno qPCR metodou.

K dispozici byla také naměřená expoziční data (pro prachové částice a B[a]P), která byla korelována (Pearsonův korelační koeficient < -0.5 a > 0.5) s celkem 3865 transkripty. Analýza drah byla provedena z odlišně exprimovaných transkriptů ($FC < 0.67$ a > 1.51) v modulu Functional classification analysis v online nástroji DAVID v6.7. Vyšší počet odlišně exprimovaných transkriptů byl nalezen u novorozenců narozených v kontrolní oblasti Českých Budějovic. Mezi některými z afektních drah pro zimní období v obou lokalitách byly dráhy primární imunodeficiency a neurotrophinové signální dráhy. Specifickými drahami pro subjekty z Karvině byly nalezeny signální dráhy na receptoru T-buněk a pro subjekty z Českých Budějovic převážně dráha pro nemalobuněčný plicní karcinom a signální dráha na receptoru B-buněk. Podrobnější výsledky z microarray analýzy budou diskutovány a předběžné výsledky z verifikace qPCR budou také představeny.

Studie je podpořena Grantovou agenturou České republiky č. 13-13485S.

P-14

Evoluční historie nejstaršího lentiviru: charakterizace integrací ELVgv a pozitivní selekce *TRIM5* v řádu *Dermoptera*

Hron T.¹, Farkašová H.¹, Padhi A.², Pačes J.¹, Elleder D.¹

¹ Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

² Department of Animal and Avian Sciences, University of Maryland, College Park, MD, USA

Endogenní retroviry jsou genetické elementy, které vznikají při integraci exogenního viru do genomu zárodečné buněčné linie. To vede k vertikálnímu přenosu virové DNA do dalších generací hostitele. Tyto "virové fosilie" představují unikátní zdroj informací o evoluci retrovirů. Lentiviry jsou významnou skupinou retrovirů infikující široké spektrum savců. Na rozdíl od jejich důkladně prozkoumané moderní evoluci vznik lentivirů zůstává téměř neprozkoumán. To je dáno především tím, že k tvorbě jejich endogenní formy dochází jen velmi zřídka. Nedávný objev evolučně starého endogenního lentiviru, ELVgv, v genomu letuchy malajské (řád *Dermoptera*) slibuje významný vhled do dávné historie těchto virů.

V naší práci jsme analyzovali několik provirových sekvencí ELVgv v genomech čtyřech jedinců pokrývajících všechny současné druhy dermopter. Přítomnost integrované virové DNA byla potvrzena u všech vzorků, což ukazuje, že ke vzniku ELVgv došlo před diverzifikací řádu *Dermoptera*. Díky výpočtu substituční rychlosti pro jednotlivé virové integrace jsme odhadli, že k jejich vzniku došlo před 21-40 miliony let. Výsledky fylogenetické analýzy dále pomohly upřesnit čas endogenizace ELVgv u předka dermopter, ke které došlo pravděpodobně před více než 60 miliony let. Současně jsme také ukázali přítomnost silné pozitivní selekce, která působila na hostitelský antivirový faktor *TRIM5* výlučně u předka současných dermopter - v období předpokládané aktivity ELVgv. Tyto výsledky naznačují vzájemnou koevoluci nejstarší známé lentivirové linie s hostitelským antivirovým faktorem a posouvají předpokládaný vznik lentivirů za hranice geologického období třetihor.

P-15

Analýza výskytu k-merov aminokyselín na rozhraní komplexov proteín–DNA: hľadanie kovariácií s rozpoznávanými nukleotidovými sekvenciami

Jakubec D.^{1,2}, Sarker M.¹, Vondrášek, J.¹

¹ *Ústav organickej chemie a biochemie, AV ČR, Flemingovo náměstí 542/2, 166 10 Praha 6*

² *Katedra fyzikální a makromolekulární chemie, Přírodovědecká fakulta, Univerzita Karlova v Praze, Albertov 6, 128 43 Praha 2*

Štúdium princípov sekvenčne-špecifického rozpoznávania nukleových kyselín proteínmi predstavuje jednu z klasických oblastí aplikácií bioinformatických nástrojov a postupov. K dnešnému dňu je zrejmé, že napriek energetickej výhodnosti niektorých geometrických usporiadaní aminokyselinovo-nukleotidových kontaktov nie sú tieto „one-to-one“ interakcie vo väčšine prípadov determinantmi sekvenčne-špecifického rozpoznávania. Je však možné o rozpoznávacom procese niečo usudzovať na základe štúdia súvislostí medzi zložením dlhších reťazcov aminokyselín a nukleotidovými sekvenciami nimi rozpoznávaných väzobných miest? Túto otázku sme sa pokúsili zodpovedať analýzou interakčných rozhraní v neredundantnej sade stoviek štruktúr proteín–DNA komplexov. Táto sada bola ďalej rozdelená na podskupiny na základe úrovne DNA-väzobnej špecificity zúčastnených proteínov. Na interakčných rozhraniach týchto komplexov sme vypočítali frekvencie všetkých možných k-merov ($k = 2, 3, 4, 5$) aminokyselín a porovnali ich s hodnotami očakávanými na základe frekvencií výskytu jednotlivých reziduí. Identifikované nadpopulované k-mery boli porovnané medzi sadami sekvenčne-špecificky a nešpecificky sa viažúcich proteínov. Pri významných motívoch využívaných sekvenčne-špecificky sa viažúcimi proteínmi boli následne študované kovariácie so sekvenciou rozpoznávaného úseku DNA.

P-16

Na cestě k plně automatické identifikaci molekul

Szepe P.¹, Jarkovská K.¹, Svoboda O.¹, Pech D.¹, Šauer M.¹

¹ *Chemaxon Ltd. 1031 Budapest, Zahony U. 7, Hungary*

Ve společnosti ChemAxon se již od roku 1998 snažíme propojovat chemii a informatiku. Naším cílem vždy bylo umožnit chemikům, aby mohli snadno a efektivně pracovat se svými daty a mohli se tak plně věnovat hlavní náplni své práce namísto třídění a upravování dat. V současnosti nabízí ChemAxon například nástroje pro chemickou vizualizaci (Marvin Suite), vyhledávání (JChem) nebo databázový management (Instant JChem). Novinkami v portfoliu společnosti jsou webové nástroje souhrnně označované jako Plexus Suite a produkt zaměřený na práci s biomolekulami (Biomolecule Toolkit).

Jedním z úkolů, se kterými se mnoho chemiků setkává na denní bázi, je identifikace molekul. Experimentálně můžeme změřit hmotnostní, NMR, EPR nebo IČ spektrum. Chemik ale bohužel často musí určit strukturu molekuly na základě těchto informací manuálně. V ChemAxonu bychom rádi udělali krok směrem k automatické identifikaci sloučenin. Přestože je vývoj nového produktu teprve na začátku, již nyní jsme schopni porovnat hmotnostní spektra s databází a velmi přesně určit podobnost. Rozšíření a zobecnění na další typy spekter je pouze otázkou času.

P-17

Využití báze znalostí pro predikci protein–protein interakčních míst

Jelínek J.¹, Hoksza D.¹

¹ *Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Malostranské nám. 25, 118 00 Praha*

Vzhledem k náročnosti zkoumání protein-protein interakcí je snaha predikovat interagující oblasti pomocí počítačových metod. Stávající metody však nemají příliš vysokou úspěšnost – na širěji používaných datasetech se hodnota MCC (Matthewův korelační koeficient) obvykle pohybuje kolem 0,2-0,4. Na druhou stranu se, s ohledem na počet již známých proteinových komplexů, otevírá prostor pro metody využívající báze znalostí. Proto zde uvedeme metodu, která pro každou aminokyselinu zkoumaného proteinu hledá v bázi znalostí podobné záznamy s ohledem na sousední aminokyseliny. Do úvahy se bere jak struktura okolí, tak i vybrané fyzikálněchemické vlastnosti jednotlivých sousedních aminokyselin (typ aminokyseliny; relativní velikost povrchu dostupná solventu). Získané informace se využívají pro inicializaci metody strojového učení zvané CRF (Conditional Random Field). Bázi znalostí jsme vytvořili na základě cca. 60 000 komplexů z PDB (Protein Data Bank), čímž jsme získali informace o cca. 54 milionech aminokyselin a jejich okolí. Použitelnost přístupu závisí na zvoleném způsobu dobývání znalostí, od kterého se vyžaduje dostatečně rychlé porovnávání jednotlivých struktur nebo efektivní odfiltrování nepodobných struktur. Proto se zaměříme na porovnání výhod a nevýhod dvou rozdílných přístupů: kombinaci relační databáze a specializované grafové knihovny, která je rychlá, ale schopná efektivně hledat pouze přesné shody; a metody fingerprintů, která je pomalejší, ale schopná nalézt i mírně odlišné struktury.

P-18

Komplexní bioinformatická analýza expresních a mutačních profilů dlaždicobuněčných karcinomů hlavy a krku

Kolář M.¹, Strnad H.¹, Hroudová M.¹, Štefl M.^{1,3}, Szabo P.², Plzák J.³, Smetana K. Jr.²

¹ Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

² Anatomický ústav, 1. lékařská fakulta, Univerzita Karlova v Praze, U Nemocnice 3, 128 00 Praha 2

³ Klinika ORL a chirurgie hlavy a krku, 1. lékařská fakulta, Univerzita Karlova v Praze a Fakultní nemocnice v Motole, V Úvalu 84, 150 06 Praha 5

Dlaždicobuněčné karcinomy hlavy a krku (HNSCC) jsou charakteristické vysokou morbiditou a mortalitou. Jimi způsobená mortalita zůstává kvůli vysoké různorodosti těchto nádorů v posledních 40 letech na stejné úrovni. Důvodem je i jejich etiologie: Jsou spojeny s nadměrnou konzumací alkoholu a tabákových výrobků. Rizikovým faktorem je také infekce virem HPV. Popsanou heterogenitou, společně s nedostatkem znalostí o signálních drahách, které vedou ke vzniku nádorového bujení u HNSCC, je brzděn vývoj nových léčebných postupů. Popis onkogenních drah a alterací jejich mediátorů, by mohl umožnit přesnější klasifikaci HNSCC, vývoj nových terapií a personalizaci léčby. Mezi alterace vedoucí k rakovinnému bujení patří mimo jiné somatické mutace v genech kódujících komponenty signálních drah a v jejich cis-regulačních oblastech. V této studii jsme se na vzorku devatenácti pacientů zaměřili na detekci krátkých somatických variant vzniklých v průběhu onkogeneze a na jejich vliv na signální dráhy, které mohou ovlivnit vznik nádorů a jejich progresi.

Tato práce byla podpořena grantem MZ VES: NV15-28933A.

P-19

Parametrizace ekvalizace elektronegativity (EEM) metodou diferenciální evoluce

Pazúriková J.^{1,3}, Raček T.^{1,2}, Svobodová Vařeková R.², Křenek A.³, Geidl S.², Koča J.²

¹ *Fakulta informatiky, Masarykova univerzita, Botanická 68a, 602 00 Brno*

² *Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta a CEITEC – Středoevropský technologický institut, Masarykova Univerzita Brno, Kamenice 5, 625 00 Brno-Bohunice*

³ *Ústav výpočetní techniky, Masarykova univerzita, Botanická 68a, 602 00 Brno*

Teoretický koncept parciálních atomových nábojů, který zjednodušeně popisuje rozložení elektronové hustoty na molekule, je používán v mnoha metodách výpočetní chemie. Samotné parciální náboje mohou být určeny ab-initio výpočty, ty jsou ale pro běžné použití příliš náročné. Alternativou jsou metody empirické jako je ekvalizace elektronegativity EEM [1]. Nutností EEM metody je ale jejich prvotní parametrizace na základě tréninkové sady molekul, pro něž byly parciální náboje spočteny pomocí kvantové mechaniky.

Obvyklým přístupem stanovení parametrů je jednoduchá regrese metodou nejmenších čtverců pro parametry popisující atomové typy (A,B) opakovaná pro parametr charakterizující dosah elektrostatické interakce (κ) [2]. To ale v řadě případů vede k ne zcela uspokojivým výsledkům. Přesnější alternativou je globální optimalizace prostoru parametrů, popsaná řešením [3] ale zpravidla velmi pomalu konvergují k přijatelnému výsledku.

V této práci navržený algoritmus vhodně kombinuje globální optimalizaci metodou diferenciální evoluce pro nalezení přibližných slibných řešení a optimalizaci lokální metodou přímých kvadratických aproximací [4] k jejich finálnímu zpřesnění. Druhou hlavní inovací je rozšíření možností tzv. scorovací funkce, která hodnotí kvalitu parametrů, tak, že kombinuje celkový korelační koeficient R^2 , vážený průměr korelací jednotlivých atomových typů, RMSE apod.

Algoritmus byl vyhodnocen na třech výrazně odlišných datových sadách (velikost sady, velikost molekul, typy atomů). Scorovací funkce preferující RMSE se ukazuje jako nejvhodnější, při velikosti populace algoritmu DE 500 metoda uspokojivě konverguje po pouhých 1000 krocích. Na všech datových sadách dostáváme typické hodnoty RMSE cca. 0.1, celkovou korelaci $R^2 > 0.9$ a korelace na jednotlivých typech atomů obvykle alespoň 0.7. Tyto výsledky jsou na malých datových sadách srovnatelné s uvedenou jednoduchou metodou, na větších sadách jsou výrazně lepší.

Reference

- [1] W. J. Mortier, S. K. Ghosh, and S. Shankar, "Electronegativity-equalization method for the calculation of atomic charges in molecules," *Journal of the American Chemical Society*, vol. 108, no. 15, pp. 4315–4320, 1986.
- [2] S. Geidl, T. Bouchal, T. Raček, R. Svobodová Vařeková, V. Hejret, A. Křenek, R. Abagyan and J. Koča. High-quality and universal empirical atomic charges for chemoinformatics applications. *Journal of Cheminformatics*, London: BioMed Central Ltd, 2015.
- [3] R. Storm, K. Price, Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, *Journal of Global Optimization*, Volume 11, Issue 4, pp. 341-359, 1997
- [4] Powell, M. J. D. (2004). "Least Frobenius norm updating of quadratic models that satisfy interpolation conditions". *Mathematical Programming (Springer)* 100: 183–215.

P-20

Dynamika skladania RNA pomocou „Basin Hopping“ grafov

Kucharík M.^{1,2}, Qin J.³, Stadler P. F.^{3,4}, Hofacker I. L.²

¹ *Medirex, a.s., Medirex Group, Galvaniho 7, Bratislava*

² *Institute for Theoretical Chemistry, Faculty of Computer Science, Währinger Straße 17, University of Vienna, Vienna, Austria*

³ *RTH, University of Copenhagen, Frederiksberg, Denmark*

⁴ *Department of Computer Science & IZBI & iDiv & LIFE, Leipzig University*

Funkcia danej RNA nie je vždy určená iba jej najstabilnejšou konformáciou, ale mnohokrát závisí aj od jej schopnosti zaujať rôzne konformácie v reakcii na zmenu prostredia alebo času. Počet rôznych konformácií rastie exponenciálne od dĺžky danej RNA sekvencie, preto je nevyhnutné sa zaoberať iba zjednodušeniami – abstrakciami celého priestoru konformácií. Mnohé doteraz používané abstrakcie dostatočne nereprezentujú topológiu priestoru, ktorá je veľmi dôležitá pre predpovedanie dynamiky.

V práci definujeme „Basin Hopping“ graf, určený lokálnymi minimami a energeticky efektívnymi spojeniami medzi nimi. „Basin Hopping“ graf je riedky a efektívne reprezentuje topológiu priestoru konformácií, kvôli čomu je možné pomocou neho efektívne predpovedať dynamiku sekvencií s dĺžkou do 200nt.

Tento model bol navyše rozšírený o netriviálnu množinu pseudouzlov, čo nám dovolilo študovať ich formovanie a ukázať dôležitosť pseudouzlov pri dynamike skladania RNA.

P-21

Nové ligandy G-kvadruplexu c-myc

Kužmová E.¹, Kozák J.¹, Kolář M.², Hájek M.¹, Teplý F.¹

¹ Ústav organické chemie a biochemie AV ČR, v. v. i., Flemingovo náměstí 542/2, 166 10 Praha 6

² Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

G-kvadruplexy jsou alternativní sekundární struktury DNA. V posledních letech se ukázalo, že mají důležitou regulační funkci v savčím genomu, a že jsou významně nabohacené v promotorových oblastech onkogenů jako jsou například c-Myc, VEGF, c-Kit, KRAS, BRAF. Vývoj ligandů schopných stabilizovat specifické G-kvadruplexy zmíněných onkogenů a tím ovlivnit jejich expresi je proto z terapeutického hlediska velmi žádoucí.

V této studii jsme se zabývali screeningem více než 1500 unikátních sloučenin a hledáním vhodných ligandů stabilizujících c-Myc G-kvadruplex a downregulujících c-Myc onkogen, který je významně overexprimovaný ve většině lidských nádorů a přispívá k jejich vzniku minimálně ve 40%.

Primární screeningová strategie zahrnovala *in vitro* bezbuněčné experimenty – FRET melting esej a buněčné esej – stanovení viability pomocí XTT na dvou buněčných liniích Burkittova lymfomu Raji a CA-46. U dvou nejzajímavějších látek byly dále změřeny CD spektra, která prokázala interakci s G-kvadruplexem. Jejich stabilizační účinek byl potvrzen pomocí polymerázové stop esej. Dále byl testován jejich vliv na buněčný cyklus a schopnost inhibovat expresi *c-myc* genu na úrovni mRNA a proteinu ve vybraných nádorových liniích. V neposlední řadě byl pomocí microarray experimentu studován jejich účinek na celogenomové úrovni na Raji buňkách.

Podpořeno projektem InterBioMed LO1302 od Ministerstva školství, mládeže a tělovýchovy a Akademie věd České republiky (RVO: 61388963).

P-22

Jak úspěšná může být identifikace inzerčních míst přímo ze sekvenačních dat?

Dvořáková Z.¹, Lexa M.¹

¹ *Masarykova Univerzita, Fakulta Informatiky, Botanická 68a, 602 00 Brno*

Jedním z problémů při rekonstrukci genomů ze sekvenačních dat může být přítomnost repetitivních sekvencí, buď tandemových repetit (zvaných někdy satelity), nebo různých rodin transpozonů. Záleží hlavně na délce sekvencí určených sekvenátorem (tzv. “ready”) a hloubkou sekvenování. V lepším případě se to projeví mezerami v poskládaném genomu, v horším třeba přítomností nesprávně poskládaných úseků.

Zabývali jsme se možnostmi, jak zjistit vybrané informace o transpozomech přímo ze sekvenačních dat bez potřeby přistupovat ke skládání genomu. Jako jednu z informací, která je biologicky zajímavá a nemusí být dostatečně pokryta existujícími referenčními genomy, jsme si vybrali informaci o vzájemném vnoření transpozonů do sebe. Kromě znalosti o tom, jak jsou transpozony do sebe vnořeny v genomu, by bylo jistě zajímavé i zjistit, zda informace získané přímo ze sekvencí ze sekvenátoru budou ve shodě s informacemi zjištěnými z referenčních genomů. Taky je možné využít nespočet sekvenačních experimentů dostupných například v databázi SRA a možnost vyhnout se tak výpočetně náročné a problematické operaci skládání sekvencí.

Budeme prezentovat postup založen na porovnávání sekvencí ze sekvenátoru s databází celých (tzv. “full-length”) transpozonů. Ready, které obsahují extrémní pozice některého s transpozonů v databázi, jsou dále analyzovány a je u nich určena identita zbylé sekvence. V současné aplikaci se zejména zajímáme o příslušnost sekvencí k rodinám transpozonů. Z výsledků analýzy je možné pak vytvořit tabulku s frekvencemi vzájemných vnoření jednotlivých rodin. Algoritmus sice není složitý, ale dle prvních testů výsledky (citlivost i přesnost) silně závisí na jeho parametrizaci. Budeme proto prezentovat výsledky testů na uměle připravených sekvencích, které nám otázky přesnosti pomůžou lépe pochopit.

P-23

Automatická anotácia sekundárných štruktúr v cytochrómech P450 a hodnotenie ich štruktúrnej podobnosti

Midlik A.¹, Berka K.², Svobodová Vařeková R.¹, Koča J.¹

¹ *Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta a CEITEC – Středoevropský technologický institut, Masarykova Univerzita Brno, Kamenice 5, 625 00 Brno-Bohunice*

² *Katedra fyzikální chemie, RCPTM, Přírodovědecká fakulta, Univerzita Palackého v Olomouci, 17. listopadu 1192/12, 771 46 Olomouc*

Cytochrómy P450 (CYP) sú nadrodinou proteínov vyskytujúcich sa v organizmoch od baktérií až po človeka, plniacich najčastejšie funkciu monooxygenázy. Ide o hemoproteíny s prevažne alfa-helikálnou štruktúrou. Napriek nízkej sekvenčnej identite v rámci tejto nadrodiny je štruktúra domény CYP pomerne konzervovaná – obsahuje radu prvkov sekundárnej štruktúry, ktorá sa vyskytuje u väčšiny CYPov [1,2]. V súčasnosti je v databáze PDB dostupných 472 štruktúr zodpovedajúcich členom tejto nadrodiny (podľa klasifikácie CATH) a neustále sú získavané nové štruktúry.

Označenie charakteristických prvkov sekundárnej štruktúry je u tejto nadrodiny už empiricky ustálené [2]. Preto má anotácia sekundárnej štruktúry veľký význam pre analýzu štruktúry nových CYPov, porovnanie so známymi štruktúrami a tiež pre pochopenie rozdielov v ich funkcii či precíznejšie hľadanie tunelov pre prístup k aktívnemu miestu. Taktiež umožňuje rýchle odhalenie konformačných zmien, ktoré môžu nastať pri naviazaní ligandu.

Manuálna anotácia charakteristických prvkov sekundárnej štruktúry je vzhľadom na veľký počet dostupných štruktúr pomerne časovo náročná. Táto práca sa zameriava na možnosť ich automatickej anotácie. Algoritmus automatickej anotácie je založený na superpozícii analyzovanej štruktúry so štruktúrou už anotovaného templátu a následnom priradení vzájomne si zodpovedajúcich prvkov sekundárnej štruktúry. K samotnému priradeniu sekundárnej štruktúry je využitý štandardný program DSSP [3].

Automatická anotácia je aplikovaná na sadu proteínov z nadrodiny CYP. Na výsledkoch je diskutovaná vhodnosť použitého algoritmu. Ďalej je analyzovaná vzájomná štruktúrna podobnosť týchto proteínov, jej porovnanie so sekvenčnou identitou a súvislosť so zdrojovým organizmom. Tiež je poukázané na najvýznamnejšie rozdiely v štruktúre.

Reference

- [1] Otyepka, M.; Skopalik, J.; Anzenbacherova, E.; Anzenbacher, P., What common structural features and variations of mammalian P450s are known to date? *Biochimica Et Biophysica Acta-General Subjects* **2007**, 1770, 376-389.

- [2] Johnson, E. F.; Stout, C. D., Structural diversity of eukaryotic membrane cytochrome P450s. *Journal of Biological Chemistry* **2013**, 288, 17082-17090.
- [3] Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, 22, 2577-2637.

P-24

ChemGenDB – integrovaná platforma pro sběr a analýzu chemických sloučenin a HTS dat

Müller T.¹, Jindřich J.¹, Škuta C.^{1,2}, Sedlák D.¹, Pombinho A.¹, Svozil D.^{1,2}, Bartůněk P.¹

¹ CZ-OPENSOURCE: National Infrastructure for Chemical Biology, Institute of Molecular Genetics of the ASCR, v. v. i, Vídeňská 1083, 142 20 Prague 4

² UCT Prague, Laboratory of Informatics and Chemistry, Technická 5, 166 28 Prague 6

ChemGenDB je databázový/LIMS systém ve formě webové aplikace. Využívá převážně programovací jazyk Python a další volně dostupné open-source softwarové komponenty jako databázový systém PostgreSQL či cheminformatický framework RDKit pro manipulaci s chemickými strukturami (konverze chemických formátů, výpočet chemicko-fyzikálních vlastností, substrukturní a podobnostní vyhledávání aj.). Uživatelské rozhraní je postaveno primárně na programovacím jazyce JavaScript, asynchronních dotazech a HTML 5. ChemGenDB je tedy multiplatformní a funguje ve všech běžně používaných internetových prohlížečích.

ChemGenDB v současné době nabízí následující funkce: ukládání informací o chemických vzorcích, jejich vlastnostech a strukturách, substrukturní a podobnostní vyhledávání, správa fyzického umístění vzorků na chemických destičkách, práce s 1D a 2D čárovými kódy pro identifikaci vzorků a destiček, reformáty destiček (klonování, sériové ředění, Z-reformát), vytváření destiček pro biologické experimenty, automatický import naměřených dat, jejich vizualizace a analýza (výpočet křivek koncentrace-odezva, fragmentální analýza, shluková analýza aj.).

P-25

FireProt: výpočetní platforma pro návrh termostabilních vícebodových mutantů

Musil M.^{1,2,3}, Bednar D.^{1,2}, Beerens K.¹, Sebestova E.¹, Bendl J.^{1,2,3}, Khare S.⁴, Chaloupkova R.¹, Prokop Z.^{1,2}, Brezovsky J.¹, Baker D.⁵, Damborsky J.^{1,2}

¹ *Loschmidtovy laboratoře, Ústav experimentální biologie a Centrum pro výzkum toxických látek v prostředí, Přírodovědecká fakulta, Masarykova univerzita, Kamenice 5/A13, 625 00 Brno*

² *Mezinárodní centrum klinického výzkumu, Fakultní nemocnice u sv. Anny v Brně, Pekařská 53, 656 91 Brno*

³ *Ústav informačních systémů, Fakulta informačních technologií, Vysoké učení technické v Brně, Božetěchova 1, 612 66 Brno*

⁴ *Department of Chemistry&Chemical Biology, Rutgers University, 610 Taylor Road, Piscataway, NJ08854, USA*

⁵ *Department of Biochemistry, University of Washington, Seattle, WA98195, USA*

Proteiny nalézají v současné době stále širší uplatnění v biotechnologiích a biomateriálech. Za jednu z klíčových vlastností proteinů je považována jejich stabilita v provozních podmínkách. Proteiny mohou být stabilizovány prostřednictvím modifikací jejich terciální struktury. Stávající výpočetní metody jsou schopny predikovat pouze jednobodové mutace, které je dále třeba manuálně rekombinovat. Konstrukce vícebodových mutantů se tak stává netriviálním, experimentálně vysoce náročným problémem. Cílem tohoto příspěvku je představit robustní výpočetní strategii FireProt, sloužící pro predikci vícebodových mutantů. Spolehlivost strategie byla validována s použitím mutací z databáze ProTherm. Strategii FireProt lze získat silně stabilizované proteiny s minimálním experimentálním úsilím, což jsme demonstrovali na dvou modelových enzymech haloalkandehalogenase DhaA a γ -hexachlorocyklohexandehydrochlorinase LinA. Aplikací strategie FireProt byla s využitím pouze několika vícebodových mutantů u obou enzymů výrazně navýšena jejich stabilita ($\Delta T_m = 24^\circ\text{C}$ a 21°C) bez poškození jejich funkce. Nevýhodou stávající strategie je její netriviální aplikace a náročnost na výpočetní zdroje. V současnosti pracujeme na její další optimalizaci a automatizaci za účelem zpřístupnění široké uživatelské komunitě ve formě webové služby.

P-26

Korelace mezi Bayesovským skóre a biologickou aktivitou ligandů aktivních na estrogenních receptorech α a β

Novotný J.¹, Svozil D.^{1,2}

¹ CZ-OPENSSCREEN & Laboratoř informatiky a chemie, Fakulta chemické technologie, VŠCHT Praha, Technická 5, 166 28 Praha 6

² CZ-OPENSSCREEN, Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

Naivní Bayesův klasifikátor (NBK) se v oblasti chemoinformatiky běžně používá k predikci, zda dané organické molekuly (ligandy) jsou či nejsou biologicky aktivní na daném molekulárním cíli. Atraktivní vlastností NBK je jeho pravděpodobnostní interpretace, kdy výsledkem je skóre udávající, s jakou pravděpodobností je chemická látka aktivní na daném molekulárním cíli. Toto skóre se běžně interpretuje též tak, že čím vyšší skóre, tím má látka vyšší potenci (tedy tím je aktivnější). Cílem práce bylo ověřit, nakolik koreluje Bayesovské skóre s biologickou aktivitou ligandů pro estrogenní receptory α a β .

P-27

Evoluce genomu *Burkholderia contaminans* během chronické infekce pacienta s cystickou fibrózou

Nunvář J.¹, Degrossi J.², Bloodworth R.³, Cardona S.³, Dřevínek P.¹

¹ 2. LF UK a FN Motol, Ústav lékařské mikrobiologie, V Úvalu 84, 150 06 Praha

² University of Buenos Aires, Buenos Aires, Argentina

³ Faculty of Science, University of Manitoba, Department of Microbiology, Manitoba, Canada

Plíce pacientů trpících cystickou fibrózou (CF) obsahují velké množství viskózního hlenu, který může být chronicky kolonizován bakteriemi komplexu *Burkholderia cepacia* (Bcc komplex). Chronické infekce Bcc mohou být letální, ke smrti pacientů dochází následkem fulminantní pneumonie doprovázené přechodem Bcc do krve a následnou sepsí. V této studii jsme se zaměřili na druh *B. contaminans*, nově popsáný v rámci Bcc komplexu. Porovnávali jsme genomové sekvence tří kmenů izolovaných během šesti let chronické infekce argentinské CF pacientky, které zahrnovaly první vykultivovaný kmen (FFH2055), kmen z klinicky stabilní fáze (467) a kmen izolovaný z hemokultury při fatální sepsi (MF16). Všechny kmeny náležely do stejného multilokusového sekvenčního typu (ST872).

Fylogenetická analýza ukázala, že genomové sekvence prvního a posledního izolovaného kmene jsou téměř shodné, zatímco kmen 467 je jejich genomovým derivátem. Kmen 467 se lišil přítomností 1433 mutací, což reprezentuje 0,017 % celkové velikosti genomu. Mutacemi bylo postiženo 1002 genů, tj. 14 % z celkového počtu. 94 genů bylo pseudogenizováno posuvovými mutacemi. Distribuce mutací vzhledem k jejich lokalizaci (intra/intergenové) a efektu na translaci (synonymní/nesynonymní) odpovídala náhodnému rozložení. 70 % nukleotidových substitucí tvořily AT→GC tranzice, což indikuje defekt v mismatch-repair systému (MMR). Nesynonymní mutace byly nalezeny v obou genech kódujících proteiny MMR, MutS a MutL. Měření mutační frekvence potvrdilo hypermutátorový fenotyp kmene 467 i dalších čtyř kmenů izolovaných během posledních dvou let chronické infekce. Všechna tato pozorování svědčí o omezeném vlivu purifikující selekce při dlouhodobé evoluci populace *B. contaminans* ve stabilním a nutričně bohatém prostředí plic postižených cystickou fibrózou.

Podpořeno grantem Ministerstva zdravotnictví ČR s reg. č. 15-28017A.

P-28

Porovnání výsledků genové exprese při použití různých anotací genomu

Oppelt J.^{1,2}, Jankůvová M.², Mráz M.^{1,3,4}

¹ CEITEC-Central European Institute of Technology, Masarykova univerzita, Kamenice 5, 625 00 Brno

² Národní Centrum pro Výzkum Biomolekul, Přírodovědecká fakulta, Masarykova univerzita, Kotlářská 2, 611 37 Brno

³ Interní hematologická a onkologická klinika, Univerzitní nemocnice Brno a Lékařská fakulta, Masarykova univerzita, Jihlavská 20, 625 00 Brno

⁴ University of California-San Diego, 3350 La Jolla Village Dr, San Diego, CA 92161, Spojené státy

Při analýze dat z next-generation sequencing (NGS) musíme udělat hned několik rozhodnutí, které mohou významně ovlivnit konečné výsledky. K těmto rozhodnutím patří, mimo jiné, výběr referenční anotace genomu, která popisuje, které geny a na kterých pozicích se v genomu nachází. Toto platí zejména u resekvenačních experimentů, jako je například RNA-Seq. Pro většinu modelových organismů, včetně člověka, existuje celá řada anotací. Každá anotace je spravovaná jiným konsorciem a také je většinou rozdílná v přístupu anotování samotných genů. Přístupy jsou v zásadě dva a to anotace automatická a manuální. Důsledek výběru anotace při analyzování dat může mít velký dopad vzhledem k tomu, že rozdílné anotace obsahují různé počty genů nebo i různé počty isoform. Tento fakt pocítíme hlavně při analýze rozdílně exprimovaných genů. Díky vývoji v metodách pro analýzu genové exprese v poslední době jsme již lépe schopni rozlišit zkoumat nejen jednotlivé geny, ale i jejich isoformy. Tato skutečnost jen přidává další rozměr k důležitosti výběru genové anotace.

V této analýze jsme se zaměřili na rozdíly ve výsledcích analýzy genové exprese v závislosti na výběru genové anotace. Analyzovali jsme genovou expresi u 8 párových paired-end RNA-Seq vzorků pacientů s chronickou lymfocytární leukémií, kde 4 vzorky byly před a 4 vzorky po aplikované léčbě. Pro analýzu a ukázkou rozdílných výsledků jsme vybrali 6 anotací lidského genomu (AceView, Ensembl, GENCODE, UCSC knownGene, RefSeq a Vega). Pro přesnější analýzu jsme použili jak přístup analyzování exprese genů (STAR + featureCounts + edgeR), tak analýzu jednotlivých transkriptů (BitSeq).

Při analýze jsme narazili na několik předpokládaných problémů a to hlavně při hledání odpovídajících genů napříč anotacemi vzhledem k tomu, že anotace používají jiná názvosloví a jejich převod není přímočarý. Závěry ukázaly na předpověditelné rozdíly mezi jednotlivými analýzami, které vychází z rozdílných pokrytí genomu anotacemi a to

třeba různými počty genů, jejich isoformem a typů. Analýzy též ukázaly rozdíly mezi kvantifikacemi expresí genů, které mohou mít efekt na určování rozdílné genové exprese a tím pádem mohou i způsobit rozdílné závěry. Ačkoliv již existuje několik publikací na podobné téma, tak jejich obsah je vysoce nedostačující a neporovnávají ani polovinu dostupných anotací.

P-29

Potlačení aktivity imunitního systému v primárním nádoru u pacientů s kolorektálním karcinomem consensus molecular subtype 2 je důležité pro vznik metastázy

Ostašov O.^{1,2}, Pítule P.^{1,2}, Hošek P.^{1,2}

¹ *Univerzita Karlova v Praze, Lékařská fakulta v Plzni, Biomedicínské centrum, Husova 3, 306 05 Plzeň*

² *Univerzita Karlova v Praze, Lékařská fakulta v Plzni, Ústav histologie a embryologie, Karlovarská 48, 306 05 Plzeň*

U kolorektálního karcinomu je hlavním způsobem léčby chirurgické odstranění primárního nádoru. O následujícím nasazení chemoterapie rozhoduje především rozšíření nádorových buněk do lymfatických uzlin, což ale ne vždy odpovídá skutečné potřebě a řada pacientů tak léčbu nedostane, přestože by jim výrazně pomohla a naopak. Proto je potřeba najít faktory, které by umožnily přesnější stratifikaci pacientů a cílenější používání chemoterapeutické i biologické léčby. Z veřejně dostupných mikroarrayových dat z platform Affymetrix U133A a U133plus2 byli vybráni pacienti přiřazení na základě klasifikátoru vytvořeného Guinney et al. [1] do skupiny CMS2. Z této skupiny byly dále dle TNM klasifikace vybrány dvě populace pacientů: 1) pacienti s lokálním onemocněním (tedy bez infiltrace lymfatických uzlin a vzdálených nádorových ložisek) a alespoň desetiletým obdobím bez recidivy či vzniku metastázy; 2) pacienti s metastatickou nemocí ale bez zasažení lymfatických uzlin. Stádium nádoru nebylo při výběru bráno v potaz. Celkem bylo v metastatické skupině 24 pacientů a v nemetastatické skupině 27 pacientů. S využitím „partial least squares regression“ následované „variable importance in the projection“ analýzou byly vybrány geny, které jsou nejdůležitější pro rozdíl mezi metastatickou a nemetastatickou skupinou. Následná enrichment analýza těchto genů ukázala významné obohacení o geny související s prezentací a zpracováním antigenů, aktivací T lymfocytů nebo signalizací pomocí interleukinu 12. V téměř všech případech měly vybrané geny sníženou expresi u pacientů, u kterých byla detekována metastáza v porovnání s pacienty bez metastázy či recidivy. Výsledky ukazují na významnou úlohu imunitního systému v kontrole nádorového bujení.

Reference

- [1] Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* 21, 1350–1356 (2015)

P-30

Vytěžování znalostí ze souborů chemických dat na základě interpretace QSAR modelů

Polishchuk P.¹

¹ *Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacký University and University Hospital in Olomouc, Hněvotínská 1333/5, 779 00 Olomouc*

Nedávno vyvinutý univerzální přístup pro interpretaci QSAR modelů může být využit k extrakci vztahů mezi strukturou a účinkem zachycených jakýmkoliv modelem bez ohledu na to, jaké učící metody a deskriptory byly využity k jeho konstrukci. Díky tomu je tento přístup velice perspektivní ve vytěžování znalostí z dostupných souborů chemických dat. Oproti metodám, které extrahují informace pouze na základě chemické struktury a příslušných vlastností (často se vyskytující a emergentní vzorce, korespondující páry molekul, atd.), tvorba QSAR modelů určitou míru jistoty, že daný vztah mezi strukturou a aktivitou byl opravdu pozorován a zachycen modelem. Vyvinul jsem open-source software pro strukturní a fyzikálně-chemickou interpretaci (SPCI, structural and physico-chemical interpretation) QSAR modelů, který jsem následně úspěšně aplikoval na několik datových souborů zahrnujících různé biologické a fyzikálně chemické parametry jako jsou prostupnost přes hematoencefalickou bariéru, toxicita, mutagenicita, rozpustnost a další ADMET parametry, inhibiční aktivita vůči fibrinogenovému receptoru, atd. Nástroj prezentuje vztah mezi strukturou a účinkem způsobem srozumitelným z chemického hlediska a demonstruje výhody vyvinutého přístupu oproti metodě korespondujících párů zejména pro relativně malé soubory dat. Program má grafické uživatelské rozhraní a může být používán výzkumníky s pouze základní znalostí QSAR modelování, protože všechny kroky tvorby modelu a jeho validace jsou automatizovány. Interpretace modelů může být také prováděna automaticky pomocí vyvinutých schémat případně pomocí výpočtu příspěvků vybraných strukturních motivů k sledované aktivitě. Vizualizace výsledků je možná jak přímo v programu, tak pomocí externích nástrojů. Nástroj je k dispozici na http://qsar4u.com/pages/sirms_qsar.php.

Reference

- [1] Polishchuk, P.G.; Kuz'min, V.E.; Artemenko, A.G.; Muratov, E.N. Mol. Inf. 32 (2013) 843-853.

P-31

Stanovení efektivity sestřihu pre-mRNA z RNA-seq dat

Převorovský M.¹, Hálová M.¹, Abrahámová K.¹, Libus J.¹, Půta F.¹, Folk P.¹

¹ *Univerzita Karlova v Praze, Přírodovědecká fakulta, Viničná 5, 128 43 Praha 2*

Otevřené čtecí rámce mnoha protein-kódujících genů jsou v eukaryotních genomech přerušeny nekódujícími sekvencemi – introny. V rámci maturace mRNA v jádře jsou introny z primárního transkriptu (pre-mRNA) odstraněny procesem sestřihu, který je katalyzován spliceosomem, obřím ribonukleoproteinovým komplexem. Maturovaná mRNA je následně transportována do cytoplazmy, kde je translatována. Sestřih představuje důležitou regulační vrstvu v rámci procesu genové exprese a byla identifikována široká škála proteinů ovlivňujících efektivitu sestřihu. Vliv některých sestřihových faktorů je omezený na malou skupinu funkčně či strukturně příbuzných mRNA, jiné proteiny mají globální vliv na transkriptom. Techniky sekvenování nové generace (RNA-seq) umožňují studovat transkriptom v celé jeho šíři a mimo jiné také stanovit účinnost sestřihu všech genů (resp. všech jednotlivých intronů) za různých podmínek. V prezentaci představíme kompletní analytický postup, který umožňuje z RNA-seq dat stanovit efektivitu sestřihu všech genů u kvasinek *Saccharomyces cerevisiae* a *Schizosaccharomyces pombe*, a ukážeme vliv vybraných mutací na sestřih.

P-32

NEEMP – nástroj pro parametrizaci EEMRaček T.^{1,2}, Svobodová Vařeková R.², Křenek A.³, Pazúriková J.^{1,3}, Geidl S.², Koča J.²¹ *Fakulta informatiky, Masarykova univerzita, Botanická 68a, 602 00 Brno*² *Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta a CEITEC – Středoevropský technologický institut, Masarykova Univerzita Brno, Kamenice 5, 625 00 Brno-Bohunice*³ *Ústav výpočetní techniky, Masarykova univerzita, Botanická 68a, 602 00 Brno*

Parciální atomové náboje popisují rozložení elektronové hustoty v molekule a poskytují tak informace o jejím chemickém chování. Nacházejí uplatnění v mnoha odvětvích počítačové chemie – v rámci molekulového modelování, při výpočtech molekulové dynamiky, dokování a dalších; v chemoinformatice se pak také často užívají jako deskriptory. Parciální atomové náboje jsou však pouze teoretický koncept, nejsou tedy přímo měřitelné. Nicméně lze je určit výpočtem.

Standardní metody výpočtu atomových nábojů jsou založeny na kvantové mechanice, avšak vysoká časová náročnost omezuje jejich použitelnost na relativně malé systémy. Druhým hlavním přístupem k výpočtu nábojů jsou pak metody empirické, které si zachovávají velmi dobrou přesnost kvantové mechanických metod, nicméně jejich výpočetní náročnost je výrazně nižší. Hlavní nevýhodou je ale nutnost použití empirických parametrů. Electronegativity Equalization Method (EEM) [1, 2] patří mezi nejčastěji používané metody s množstvím aplikací v literatuře. Parametrizace EEM je však netriviální proces.

Proto jsme vytvořili software NEEMP, který je schopen nalézt pro EEM velmi kvalitní a univerzální parametry ($R^2 > 0,95$ pro tréninkové i testovací sady, aplikovatelnost na více než 95 % molekul ze známých databází sloučenin) [3]. Software jsme navrhli s důrazem na numerickou stabilitu a efektivitu výpočtů, při nichž se využívá vysoce výkonných a robustních optimalizačních přístupů včetně automatického hledání ideální podmnožiny tréninkové sady. NEEMP lze dále použít také k výpočtu EEM nábojů a ověření kvality a aplikovatelnosti EEM parametrů.

Reference

- [1] W. J. Mortier, S. K. Ghosh, and S. Shankar, “Electronegativity-equalization method for the calculation of atomic charges in molecules,” *Journal of the American Chemical Society*, vol. 108, no. 15, pp. 4315–4320, 1986.
- [2] R. S. Vařeková and J. Koča, “Optimized and parallelized implementation of the electronegativity equalization method and the atom-bond electronegativity equalization method,” *Journal of computational chemistry*, vol. 27, no. 3, pp. 396–405, 2006.

- [3] S. Geidl, T. Bouchal, T. Raček, R. Svobodová Vařeková, V. Hejret, A. Křenek, R. Abagyan and J. Koča. High-quality and universal empirical atomic charges for chemoinformatics applications. *Journal of Cheminformatics*, London: BioMed Central Ltd, 2015. ISSN 1758-2946. doi:10.1186/s13321-015-0107-1.

P-33

Vliv kořenů rostlin na složení a funkční potenciál mikrobiálních komunit v dlouhodobě kontaminované zemině

Rídl J.¹, Kolář M.¹, Strejček M.², Strnad H.¹, Štursa P.², Pačes J.¹, Macek T.², Uhlík O.²

¹ *Oddělení genomiky a bioinformatiky, Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4*

² *Ústav biochemie a mikrobiologie, Vysoká škola chemicko-technologická v Praze, Technická 3, 166 28 Praha 6*

Vztah kořenů rostlin a půdních mikroorganismů je významný pro procesy bioremediace probíhající v kontaminované zemině. Pro studium vlivu vybrané vegetace a aplikace hnojiva na mikrobiální komunity jsme využili shotgunové a 16S rRNA ampliconové pyrosekvenování. Aplikace bioinformatických metod původně navržených pro hodnocení rozdílné míry exprese genů (na základě RNA-seq dat) umožnila zpracování různých sekvenčních dat unifikovaným postupem a statistické hodnocení míry abundance genů v metagenomických vzorcích. Výsledky dokládají vliv kořenů křenu, lilku a tabáku na taxonomické složení a funkční potenciál konsorcií v zemině dlouhodobě kontaminované polychlorovanými bifenoly.

P-34

Algoritmus pro rekonstrukci ncRNA BLAST hitů s nevýznamným skóre využívající suboptimálních struktur RNA

Schwarz M.¹, Pánek J.¹

¹ *Akademie věd České republiky, Mikrobiologický ústav, Vídeňská 1083, 142 20 Praha*

Pro vyhledávání homologních sekvencí jsou běžně využívány programy pro sekvenční alignment, ty jsou pro ncRNA použitelné pouze částečně, protože je známo, že u ncRNA je evolučně konzervována struktura RNA namísto sekvence a sekvenční podobnost je většinou omezena na příbuzné organismy. Proto jsme při sekvenčním vyhledávání homologů např. BLASTem omezení na hity s dostatečně významným skóre. Ve výstupu se však mnohdy nachází sekvence částečné či s mnoho nepárujícími se bázemi, které mají nevýznamné skóre, a nelze říci, zda jsou či nejsou homologní. Cílem prezentovaného algoritmu je rekonstrukce nalezených sekvencí do plné délky a predikce homologie dané sekvence podle podobnosti struktur vyhledávané ncRNA a rekonstruované sekvence. Z porovnání funkčních struktur známých ncRNA vyplývá, že mnohdy je daná struktura lépe predikována jednou z energeticky suboptimálních struktur vzhledem ke struktuře s energií minimální. Proto v našem algoritmu používáme suboptimální struktury pro určení nejlépe skórující sekvence a predikci homologie.

P-35

Molpher-lib: softwarová knihovna pro systematickou exploraci chemického prostoru

Molpher-lib: software library for systematic chemical space exploration

Šicho M.¹, Svozil D.^{1,2}

¹ CZ-OPENSOURCE & Laboratoř informatiky a chemie, Fakulta chemické technologie, VŠCHT Praha, Technická 5, 166 28 Praha 6

² CZ-OPENSOURCE, Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

Molpher je uživatelský software, který implementuje novou metodu počítačově řízené explorační chemického prostoru zvanou molekulární morphing [1]. Ačkoliv je Molpher velice sofistikovaný a dobře zpracovaný software, trpí několika nedostatky – především nízkou flexibilitou zdrojového kódu, která znemožňuje větší zásahy do samotného algoritmu a snadné rozšíření o nové funkce. V této práci představujeme softwarovou knihovnu, Molpher-lib, která je založená na původní implementaci, ale klade si za cíl umožnit uživatelům snadnou modifikaci stávajícího exploračního algoritmu, větší možnosti konfigurace a lepší přístup k datům generovaných za běhu programu. Původní implementace v C++ byla zrefaktorována a přepsána tak, aby vznikla modulární struktura s bohatým aplikačním programovacím rozhraním (API). Zároveň je toto API přístupné skrz programovací jazyk Python, který je ve vědecké komunitě velice populární. Knihovna si tak zachovává rychlost původní implementace, ale zároveň umožňuje jak velice snadnou integraci molekulárního morphingu do stávajících aplikací, tak i větší možnosti konfigurace a analýzy výsledků explorační.

Reference

- [1] Hoksza D., Škoda P., Voršilák M., Svozil D. Molpher: a software framework for systematic chemical space exploration. *J Cheminform.* 2014 Mar 21;6(1):7

P-36

Reprodukovatelná analýza efektivity fingerprintů

Škoda P.¹, Hoksza D.¹

¹ *Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Malostranské nám. 25, 118 00 Praha*

Využití virtuálního screeningu pro prioritizaci sloučenin ve velkých chemických knihovnách je běžnou praxí v raných fázích vývoje léčiv. Jednou z existujících variant virtuálního screeningu je ligand-based virtuální screening (LBVS). Úspěšnost použití LBVS se výrazně liší dle použité molekulární reprezentace. Často používanou reprezentací jsou molekulární otisky, tzv. fingerprinty. Jedním z významných parametrů ovlivňující výsledek virtuálního screeningu je množství informace o screenovaných molekulách, například počet již známých aktivních molekul. Většina publikovaných testů LBVS a souvisejících benchmarkovacích dat není dostatečně podrobně popsána, aby umožnila spolehlivou reprodukovatelnost provedených experimentů. Důsledkem jsou zhoršené možnosti vzájemného porovnání výkonnosti různých metod. Z těchto důvodů jsme provedli standardizaci a obohacení vybraných datasetů o informace, jež umožní jejich reprodukovatelné použití nejen při testování metod LBVS. V rámci prezentace představíme výsledky metod na upravených datasetech a také analýzu úspěšnosti metod v závislosti na dostupné míře znalostí o screenovaných datech. Přínosem takovéto analýzy je pak možnost využít fingerprintovou reprezentaci v závislosti na dostupných datech.

P-37

ChemSpace.js: Vizualizace chemického prostoru pro webové aplikace

Škuta C.^{1,2}, Svozil D.^{1,2}, Bartůněk P.¹

¹ CZ-OPENSOURCE, Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

² Laboratoř informatiky a chemie, CZ-OPENSOURCE, Fakulta chemické technologie, VŠCHT Praha, Technická 5, 166 28 Praha 6

Chemický prostor je multidimenzionální prostor všech teoreticky možných, energeticky stabilních, chemických sloučenin. Jedná se o užitečný nástroj pro systematické podobnostní prozkoumávání množin chemických sloučenin, vizuální inspekci vztahů mezi strukturou a aktivitou látek (SAR analýz), a prezentace chemických knihoven. Přestože existují desktopová řešení pro vizualizaci chemického prostoru, většina z nich je jen velmi úzce zaměřena bez možnosti adaptace ve specifitějších nebo zcela odlišných případech. Z těchto důvodů jsme vyvinuli ChemSpace.js, jednoduše použitelnou, univerzální knihovnu pro vizualizaci 2D chemického prostoru v rámci webové stránky. Pro vykreslování je využívána rastrová technologie HTML5 canvas prostřednictvím open-source javascriptové knihovny Konva.js. ChemSpace.js je tedy multiplatformní a funguje ve všech běžně používaných internetových prohlížečích. ChemSpace.js umožňuje jednoduchou navigaci v rámci chemického prostoru prostřednictvím přibližování/oddalování jeho vybraných částí, přiřazováním barev a rozměrů na základě kategorie/atributu nebo vykreslováním cest definovaných body/souřadnicemi ve 2D prostoru. Za použití definovaných událostí a metod může knihovna komunikovat s dalšími elementy na stránce a pracovat s externími datovými zdroji. Za účelem podpory používání a možností vylepšování je ChemSpace.js k dispozici pod MIT licencí.

P-38

Hľadanie substrátovej špecificity DUSP fosfatáz

Sladeček S.¹, Novotný M.¹

¹ *Univerzita Karlova v Praze, Přírodovědecká fakulta, Katedra buněčné biologie, Albertov 6, 128 43 Praha 2*

Duálne špecifické fosfatázy (DUSP) sú enzýmy, ktoré dokážu na jednom substráte defosforylovať fosfoserín/threonínové a fosfotyrozínové zvyšky. Všeobecne DUSP-y sú fosfatázy, ktoré defosforylujú mitogénami aktivované proteín kinázy (MAPK). Napriek tomu sú medzi nimi rozdiely v špecificite, keďže niektoré DUSPy dokážu defosforylovať iba malý počet MAPK, kým iné defosforylujú širšiu škálu proteínov z tejto rodiny. Počet DUSPov narastá s komplexnosťou genómu a u človeka ich bolo doposiaľ opísané 25. Medzi nimi sú rozdiely v exprésii, bunkovej lokalizácii ako aj rozdiely v štruktúre. Doteraz nie je jasne definovaný faktor, ktorý by zapríčiňoval rozdiely v ich defosforylačnej špecificite. V predloženej práci sme sa použitím *in silico* metód ako sú vytvorenie fylogenetických stromov, ConSurf, InParanoid, DSSP atď. snažili zmapovať niektoré spoločné a odlišné charakteristiky DUSPov, ktoré by mohli mať vplyv na ich špecificitu voči substrátu.

P-39

Genomový projekt *Pseudomonas alcaliphila* JAB1

Strnad H.¹, Ridl J.¹, Hroudová M.¹, Uhlík O.², Macek T.², Pačes V.¹

¹ Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

² Vysoká škola chemicko-technologická v Praze, Ústav biochemie a mikrobiologie, Technická 5, 166 28 Praha 6

Bakterie *Pseudomonas alcaliphila* JAM1 je studována pro své bioremediační schopnosti jako je například degradace chlorbenzoátů a polychlorovaných bifenylů.

Genom *P. alcaliphila* JAB1 byl sekvenován za použití 454 technologie. Párová i nepárová sekvenční čtení (500k) byla složena v programu Newbler (2.5) do jednoho scaffoldu. Kompletní genomová sekvence byla připravena uzavřením sekvenčních mezer v programu Gap4 (Staden) bez nutnosti další cílené sekvenace. Genom sestává z jednoho chromosomu (5,9 Mb). Chromozóm má relativně vysoký obsah GC (62,5%). Byla provedena anotace genomu (CDS, rRNA, tRNA, tmRNA, ncRNA).

Financováno: GAČR: 13-28283S, MŠMT: 2B08062, RVO 68378050.

P-40

SEED v2.0 – uživatelsky příjemné prostředí pro zpracování amplikonových sekvencí z NGS

Větrovský T.¹, Štursová M.¹, Baldrian P.¹

¹ *Mikrobiologický ústav AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4*

Je mnoho zavedených programů na zpracování amplikonových dat z NGS z nichž naprostá většina je založena na práci v příkazovém řádku, a proto je hůře stravitelná pro biology bez základních znalostí bioinformatiky. Program SEED (<http://www.biomed.cas.cz/mbu/lbwrf/seed/>) je volně dostupnou “okenní” aplikací pro zpracování amplikonových dat vyvinutý pro 64-bitovou platformu Windows. Program byl vytvořen s cílem poskytnout intuitivní prostředí pro rychlou bioinformatickou analýzu amplikonových sekvencí odvozených od různých taxonomických markerů (houbových ITS oblastí, bakteriálních 16S rDNA, aj.) na stolním počítači. SEED verze 2.0 je plně optimalizován pro zpracování amplikonových dat ze sekvenátoru Illumina a umožňuje pohodlnou práci s daty o velikosti okolo 8 milionů sekvencí (~4 GB) na běžném počítači s 8 GB RAM. Nová verze 2.0 má široké spektrum funkcí zahrnujících editaci sekvencí a jejich názvů, spojování “pair-end” sekvencí, filtrování podle kvality, třídění a seskupování sekvencí, “alignování” a klastrování sekvencí, tvorbu OTU tabulek, vyhledávání pomocí algoritmu BLAST, taxonomickou identifikaci, konstrukci a editaci fylogenetických stromů a mnoho dalších.

P-41

Databáze látek zpomalujících stárnutíVoller J.^{1,2}, Kadlecová A.²¹ *Ústav molekulární a translační medicíny, Univerzita Palackého, Hněvotínská 5, 775 15 Olomouc*² *Laboratoř růstových regulátorů, Centrum regionu Haná, Univerzita Palackého & Ústav experimentální botaniky AV ČR, Šlechtitelů 11, 783 71 Olomouc*

Stárnutí a s ním spojená kumulace poškození a narušení homeostázy jsou u většiny druhů včetně člověka a klasických modelových organismů nevratné procesy. Přestože naše pochopení příčin stárnutí je pouze omezené, jsme schopni v některých případech modulovat rychlost vzniku poškození případně jeho reparaci. Zatím nejefektivnější zásahy proti stárnutí zpočívají v genové manipulaci nebo režimových opatřeních typu kalorické restrikce. Farmakologická modulace stárnutí i délky života je zatím v počátcích. Tomu odpovídá i stav databází katalogizujících známé intervence proti stárnutí. Naším cílem je vytvořit specializovanou anotovanou databázi nízkomolekulárních látek schopných zpomalovat stárnutí. V první fázi jsme vytvořili databázi věnovanou farmakologické modulaci délky života *C. elegans*. Rozšíření o další organismy a protektivní efekty na molekulární, buněčné, tkáňové a orgánové úrovni je předmětem probíhajících studentských prací.

P-42

Počítačová predikce syntetické dostupnosti

Voršilák M.¹, Svozil D.^{1,2}

¹ CZ-OPENSSCREEN & Laboratoř informatiky a chemie, Fakulta chemické technologie, VŠCHT Praha, Technická 5, 166 28 Praha 6

² CZ-OPENSSCREEN, Ústav molekulární genetiky AV ČR, v. v. i., Vídeňská 1083, 142 20 Praha 4

Počet generovaných struktur při počítačovém *de novo* návrhu běžně dosahuje desítek či stovek tisíc, a proto je potřeba pro další zpracování odfiltrovat takové, které nemají požadované vlastnosti. Mezi tyto vlastnosti patří např. biologická aktivita, kdy v první fázi pouze rozhodujeme o tom, zda je či není daná látka aktivní, toxicita či syntetická dostupnost. Pro takové filtrování se běžně používají klasifikační metody, pro jejichž naučení je třeba mít k dispozici trénovací množinu obsahující příklady z obou dvou tříd. Takovou množinu je relativně snadné zkonstruovat v případě biologické aktivity či toxicity, kteréžto vlastnosti je možné experimentálně určit. Syntetická dostupnost je ovšem vlastnost, kterou nelze experimentálně charakterizovat a pro konstrukci klasifikátoru tak nejsou k dispozici příklady struktur obtížně syntetizovatelných. Proto jsme navrhli metodu Nonpher pro návrh synteticky nedostupných sloučenin založenou na algoritmu Molpher. Za použití Molpheru byly synteticky dostupné struktury transformovány pomocí jednoduchých strukturních změn (např. přidej/změň/uber atom/vazbu) na sloučeniny mnohem komplexnější a tudíž obtížně syntetizovatelné. Nonpher metoda byla porovnána s dalšími možnými přístupy pro návrh obtížně syntetizovatelných látek. Pro porovnání byla použita klasifikace pomocí random forest klasifikátoru. Kvalita jeho predikce byla vyhodnocena za použití externí testovací množiny sestávající z publikovaných látek, obtížnost jejichž syntézy byla konsensuálně určena několika medicínálními či organickými chemiky. Námi navržený postup Nonpher dosáhl ze všech testovaných postupů nejvyšší přesnosti (85%) a může proto významným způsobem přispět k dalšímu rozvoji počítačové predikce syntetické dostupnosti.



ELIXIR CZ



L6-01

Software Development for Life Sciences (SD4LS)

Pergl R.¹, Svobodová Vařeková R.², Černý J.³, Štourač J.²

¹ *České vysoké učení technické v Praze, Fakulta informačních technologií, Thákurova 9, 160 00 Praha 6*

² *Masarykova univerzita, Přírodovědecká fakulta, Kamenice 753/5, Bohunice, Brno*

³ *Biotechnologický ústav AV ČR, v. v. i., Průmyslová 595, 252 50 Vestec*

Vědečtí pracovníci v oblasti bioinformatiky, chemoinformatiky a dalších oborů z oblasti moderních life sciences tráví stále více času návrhem, vytvářením a údržbou databází a softwarových nástrojů. Bohužel, velká část těchto vědeckých pracovníků neměla v průběhu své kariéry možnost se hlouběji seznámit s doporučenými metodikami a praktikami pro vývoj softwarových nástrojů [1].

Software Development for Life Sciences je pracovní skupina v rámci infrastruktury ELIXIR, která byla ustanovena s cílem vytvoření procedur pro zlepšení kvality a udržitelnosti vývoje software pro oblast Life Sciences. Tyto procedury by měly posléze být zařazeny mezi doporučované přístupy infrastruktury ELIXIR a potenciálně i ostatních biomedicínských výzkumných infrastruktur. Konkrétní řešené oblasti jsou - Vytvoření vodítek pro „good practices“ vývoje SW a hodnocení kvality; - Poskytnutí merik a nástrojů pro hodnocení kvality vývoje SW; - Vytvoření vodítek pro Life Sciences open-source software.

V rámci tohoto příspěvku vám představíme hlavní myšlenky a principy a zdroje informací. V diskusi se poté můžeme podrobněji věnovat vybraným otázkám.

Reference

- [1] Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., et. al (2014). Best practices for scientific computing. PLoS Biol, 12(1), e1001745.

	<i>strana</i>
L1-01 Metódy pre spracovanie sekvenačných dát z moderných sekvenátorov Vinař Tomáš , <i>Univerzita Komenského v Bratislave</i>	7
L1-02 Charakterizace satelitních sekvencí z NGS dat Novák Petr , <i>Biologické Centrum AV ČR, v. v. i. České Budějovice</i>	8
L1-03 Role retrotransponů při expanzích genových rodin v genomu člověka a myši Janoušek Václav , <i>Katedra zoologie, Přírodovědecká fakulta UK v Praze</i>	9
L1-04 PredictSNP2: Konsenzuální prediktor vlivu nukleotidových mutací na rozvoj monogenních chorob Bendl Jaroslav , <i>Fakultní nemocnice u sv. Anny v Brně</i>	10
L1-05 Rychlá deterministická vizualizace shotgun metagenomických dat Sedlář Karel , <i>Vysoké učení technické v Brně</i>	11
L2-01 Sémantická dvojhluková analýza dat genové exprese Malinka František , <i>České vysoké učení technické v Praze</i>	15
L2-02 Hammock: nástroj pro shlukování velkého množství krátkých peptidových sekvencí Krejčí Adam , <i>Regionální centrum aplikované molekulární onkologie, Masarykův onkologický ústav Brno</i>	16
L2-03 CeCe - simulátor mezibuněčné komunikace v dynamickém prostředí Fatka Jiří , <i>Západočeská univerzita v Plzni</i>	17
L2-04 Virtuální paralelní infrastruktury pro velká data v metabolomice Fesl Jan , <i>Jihočeská univerzita v Českých Budějovicích, Přírodovědecká fakulta, Ústav aplikované informatiky</i>	18
L2-05 Chameleon 2: algoritmus pro hierarchické shlukování dat Bartoň Tomáš , <i>Ústav molekulární genetiky AV ČR, v. v. i. Praha 4</i>	19

	strana
L3-01 Predikce toxicity nanočástic <i>Anděl Michael, Ústav experimentální medicíny AV ČR Praha</i>	23
L3-02 Léčiva na biologických membránách pohledem výpočetní chemie <i>Berka Karel, Univerzita Palackého v Olomouci</i>	25
L3-03 QSAR modelování jako nástroj pro zkoumání chemogenomického prostoru <i>Škuta Ctibor, CZ-OPENSOURCE, Ústav molekulární genetiky AV ČR, v. v. i. Praha 4</i>	26
L3-04 Jak najít jehlu v kupce sena aneb orientace v dnešních databázích <i>Svoboda Ondřej, ChemAxon s.r.o. Praha</i>	27
L3-05 Pokročilé dotazování jazykem SPARQL v databázích malých molekul <i>Galgonek Jakub, Ústav organické chemie a biochemie AV ČR, v. v. i. Praha 6</i>	28
L3-06 Vývoj nového farmakoforového modelu <i>Mareška Václav, VŠCHT Praha</i>	29
L3-07 Využití nábojových deskriptorů v chemoinformaticce <i>Geidl Stanislav, Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta, Masarykova univerzita Brno</i>	30
L4-01 Suboptimální struktury hledají optimální RNA <i>Pánek Josef, MBÚ AV ČR Praha</i>	33
L4-02 Homologní predikce terciární struktury RNA <i>Hoksza David, Univerzita Karlova Praha</i>	34
L4-03 Lokální konformery DNA a jejich automatické přiřazování <i>Božíková Paulína, Biotechnologický ústav, AV ČR Vestec</i>	35

	<i>strana</i>
L4-04	36
Detekce biomolekulárních substruktur pomocí PatternQuery <i>Pravda Lukáš, Masarykova univerzita - Přírodovědecká fakulta Brno</i>	
L4-05	37
Výskyt sekundárních struktur v sekvenčním prostoru proteinů <i>Trejšačenko Vjačeslav, Přírodovědecká fakulta Univerzity Karlovy v Praze</i>	
L4-06	39
HotSpot Wizard 2.0: Automatický návrh aminokyselinových mutací a chytrých knihoven pro proteinové inženýrství <i>Štourač Jan, Fakultní nemocnice u sv. Anny v Brně</i>	
L4-07	41
Extrakce biologicky relevantních informací ze 3D struktur biomakromolekul <i>Svobodová Vařeková Radka, Masarykova Univerzita CEITEC - Středoevropský technologický institut Brno</i>	
L4-08	43
Altruistická metadynamika <i>Hošek Petr, VŠCHT Praha</i>	
L5-01	47
Efektivní stabilizace proteinů a identifikace nových substrátů pro biotechnologie: hybridní experimentální a výpočetní přístupy <i>Brezovský Jan, Fakultní nemocnice u sv. Anny v Brně</i>	
L5-02	48
Aktivace genu NKD1 je známkou aberantní signalizace Wnt dráhy v nádorech epitelu: bioinformatická analýza velkých veřejně dostupných datasetů <i>Kolář Michal, Ústav molekulární genetiky AV ČR Praha 4</i>	
L5-03	50
Identifikace a lokalizace mutací v repetitivní oblasti genu MUC1 <i>Přistoupilová Anna, Univerzita Karlova v Praze, 1. lékařská fakulta</i>	
L5-04	51
Evoluční a speciální genomika na modelu dvou druhů slaviček <i>Reifová Radka, Přírodovědecká fakulta, Univerzita Karlova v Praze</i>	
L6-01	105
Software Development for Life Sciences (SD4LS) <i>Pergl Robert, Fakulta informačních technologií, ČVUT v Praze</i>	

P-01	55
Identifikácia a štúdium domén ľudského ryanodínového receptora 2 zodpovedných za ochorenia srdca	
Bauerová Vladena , Ústav molekulárnej biológie SAV, Bratislava	
P-02	57
Molekulární dokování růstových regulátorů.	
Bazgier Václav , Katedra fyzikální chemie, Přírodovědecká fakulta, Univerzita Palackého v Olomouci	
P-03	58
Studium molekulárních základů evoluce feromonové komunikace hmyzu prostřednictvím RNA-seq	
Buček Aleš , Ústav organické chemie a biochemie, AV ČR, Praha 6	
P-04	59
Dolbico: Databáze lokálních konformací biomolekul	
Čech Petr , VŠCHT Praha	
P-05	60
Návrh chemické knihovny potenciálních agonistů glukokortikoidního receptoru pomocí algoritmu Molpher	
Čmelo Ivan , VŠCHT Praha	
P-06	61
Uživatelsky orientované webové nástroje Loschmidtových laboratoří	
Damborský Jiří , Fakultní nemocnice u sv. Anny v Brně	
P-07	62
Computer aided drug design towards new inhibitors of Kynurenine Aminotransferase	
Dehaen Wim , VŠCHT Praha	
P-08	63
Nástroj pro generování elementárního složení z dat hmotnostních spektrometrů s vysokým rozlišením	
Doležalová Marie , Jihočeská univerzita, Přírodovědecká fakulta, Ústav aplikované informatiky, České Budějovice	
P-09	64
Analýza zmesných vzoriek vo forenznej genetike pomocou metód sekenovania novej generácie	
Đuriš František , Geneton s.r.o., Bratislava	

P-10	65
Shluková analýza lidské starodávné mtDNA: 30 000 let pod maternální dědičností Ehler Edvard , <i>Katedra biologie a environmentálních studií, Univerzita Karlova v Praze</i> <i>Pedagogická fakulta</i>	
P-11	67
Štatistický model pre anotáciu genomických mikrosatelitov Gazdarica Juraj , <i>Geneton s.r.o., Bratislava</i>	
P-12	68
ChrgDescCalc.py – nástroj na výpočet nábojových deskriptorů Hejret Václav , <i>Masarykova univerzita CEITEC MU, Brno</i>	
P-13	69
Analýza genové exprese u novorozenců narozených v prostředí s různou úrovní znečištění ovzduší Hoňková Kateřina , <i>Oddělení genetické ekotoxikologie, Ústav experimentální medicíny</i> <i>AV ČR, v.v.i., Praha 4</i>	
P-14	70
Evoluční historie nejstaršího lentiviru: charakterizace integrací ELVgv a pozitivní selektce TRIM5 v řádu Dermoptera Hron Tomáš , <i>Ústav molekulární genetiky AV ČR, v. v. i., Praha 4</i>	
P-15	71
Analýza výskytu k-merov aminokyselin na rozhraní komplexov protein – DNA: hľadanie kovariáci s rozpoznávanými nukleotidovými sekvenciami Jakubec Dávid , <i>Ústav organickej chemie a biochemie AV ČR, Praha 6</i>	
P-16	72
Na cestě k plně automatické identifikaci molekul Jarkovská Karla , <i>ChemAxon s.r.o., Praha</i>	
P-17	73
Využití báze znalostí pro predikci protein-protein interakčních míst Jelínek Jan , <i>Univerzita Karlova v Praze; Matematicko-fyzikální fakulta, Praha 2</i>	
P-18	74
Komplexní bioinformatická analýza expresních a mutačních profilů dlaždicobuněčných karcinomů hlavy a krku Kolář Michal , <i>Ústav molekulární genetiky AV ČR, Praha 4</i>	

P-19	75
Parametrizace ekvalizace elektronegativity (EEM) metodou diferenciální evoluce Křenek Aleš , Ústav výpočetní techniky, MU, Brno	
P-20	77
Dynamika skládání RNA pomocí „Basin Hopping“ grafů Kucharík Marcel , Geneton s. r. o., Bratislava	
P-21	78
Nové ligandy G-kvadruplexu c-myc Kužmová Erika , Ústav organické chemie a biochemie AV ČR, v. v. i., Praha 6	
P-22	79
Jak citlivá je identifikace inzerčních míst transpozonů přímo ze sekvenačních dat? Lexa Matej , Masarykova univerzita, Fakulta informatiky, Brno	
P-23	80
Automatická anotácia sekundárnych štruktúr v cytochrómoch P450 a hodnotenie ich štruktúrnej podobnosti Midlik Adam , NCBR, Masarykova univerzita, Brno	
P-24	82
ChemGenDB – integrovaná platforma pro sběr a analýzu chemických sloučenin a HTS dat Müller Tomáš , Ústav molekulární genetiky AV ČR, v. v. i., Praha 4	
P-25	83
FireProt: výpočetní platforma pro návrh termostabilních vícebodových mutantů Musil Miloš , Vysoké učení technické v Brně, Fakulta informačních technologií, Brno	
P-26	84
Korelace mezi Bayesovským skóre a biologickou aktivitou ligandů aktivních na estrogenních receptorech α a β Novotný Jiří , VŠCHT Praha	
P-27	85
Evoluce genomu Burkholderia contaminans během chronické infekce pacienta s cystickou fibrózou Nunvář Jaroslav , Ústav lékařské mikrobiologie 2. LF UK a FN Motol, Praha	

P-28	86
Porovnání výsledků genové exprese při použití různých anotací genomu <i>Oppelt Jan, NCBR & CEITEC MU, Brno</i>	
P-29	88
Potlačení aktivity imunitního systému v primárním nádoru u pacientů s kolorektálním karcinomem consensus molecular subtype 2 je důležité pro vznik metastázy. <i>Ostašov Pavel, Lékařská fakulta v Plzni, Univerzita Karlova v Praze</i>	
P-30	89
Vytěžování znalostí ze souborů chemických dat na základě interpretace QSAR modelů <i>Polishchuk Pavlo, Lékařská fakulta Univerzity Palackého v Olomouci - IMTM</i>	
P-31	90
Stanovení efektivity sestřihu pre-mRNA z RNA-seq dat <i>Převorovský Martin, Univerzita Karlova v Praze, Přírodovědecká fakulta, Praha 2</i>	
P-32	91
NEEMP – nástroj pro parametrizaci EEM <i>Raček Tomáš, Masarykova Univerzita; CEITEC - Středoevropský technologický institut, Brno</i>	
P-33	93
Vliv kořenů rostlin na složení a funkční potenciál mikrobiálních komunit v dlouhodobě kontaminované zemině <i>Rídl Jakub, Ústav molekulární genetiky AV ČR, v. v. i., Praha</i>	
P-34	94
Algoritmus pro rekonstrukci ncRNA BLAST hitů s nevýznamným skóre využívající suboptimálních struktur RNA <i>Schwarz Marek, Mikrobiologický ústav AV ČR, Praha</i>	
P-35	95
Molpher-lib: softwarová knihovna pro systematickou exploraci chemického prostoru <i>Šicho Martin, Vysoká škola chemicko-technologická v Praze</i>	
P-36	96
Reprodukovatelná analýza efektivity fingerprintů <i>Škoda Petr, Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Praha 2</i>	

P-37	97
ChemSpace.js: Vizualizace chemického prostoru pro webové aplikace	
Škuta Ctibor , CZ-OPENSREEN, Ústav molekulární genetiky AV ČR, v. v. i., Praha 4	
P-38	98
Hľadanie substrátovej špecificity DUSP fosfatáz	
Sladěček Stanislava , Univerzita Karlova v Praze	
P-39	99
Genomový projekt <i>Pseudomonas alcaliphila</i> JAB1	
Strnad Hynek , Ústav molekulární genetiky AV ČR, v. v. i., Praha 4	
P-40	100
SEED v2.0 – uživatelsky příjemné prostředí pro zpracování ampliconových sekvencí z NGS	
Větrovský Tomáš , Mikrobiologický ústav AV ČR, v. v. i., Praha	
P-41	101
Databáze látek zpomalujících stárnutí	
Voller Jiří , IMTM, Olomouc	
P-42	102
Počítačová predikce syntetické dostupnosti	
Voršilák Milan , VŠCHT Praha	

Anděl Michael	23
Bartoň Tomáš	19
Bartůněk Petr	60, 82, 97
Bauerová Vladena	55
Bazgier Václav	57
Bendl Jaroslav	10, 39, 47, 61, 83
Berka Karel	25, 57, 80
Božíková Paulína	35
Brezovský Jan	10, 39, 47, 61, 83
Buček Aleš	58
Budiš Jaroslav	64, 67
Cehák Jiří	18, 63
Čech Petr	35, 59
Čmelo Ivan	60
Damborský Jiří	10, 39, 47, 61, 83
Dehaen Wim	62
Doležalová Marie	18, 63
Đuriš František	64, 67
Ehler Edvard	65
Fatka Jiří	17
Fesl Jan	18, 63
Galgonek Jakub	28
Gazdarica Juraj	67
Geidl Stanislav	30, 41, 68, 75, 91
Hejret Václav	30, 68
Hoksza David	34, 73, 96
Hoňková Kateřina	69
Horský Vladimír	41
Hošek Petr	43
Hron Tomáš	70
Jakubec Dávid	71
Janoušek Václav	9, 51
Jarkovská Karla	27, 72
Jelínek Jan	73
Jindřich Jindřich	82
Kolář Michal	48, 74, 78, 93
Krejčí Adam	16

Křenek Aleš	75, 91
Kucharík Marcel	77
Kužmová Erika	78
Lexa Matej	16, 79
Macas Jiří	8
Malinka František	15
Mareška Václav	29
Midlik Adam	80
Mořkovský Libor	51
Müller Tomáš	82
Musil Miloš	10, 39, 61, 83
Novák Petr	8
Novotný Jiří	84
Novotný Marian	98
Nunvář Jaroslav	85
Oppelt Jan	86
Ostašov Pavel	88
Pačes Jan	51, 70, 93, 99
Pánek Josef	33, 94
Pazúriková Jana	75, 91
Pech David	27, 72
Pitule Pavel	88
Polishchuk Pavlo	89
Pravda Lukáš	36
Převorovský Martin	90
Přistoupilová Anna	50
Raček Tomáš	41, 75, 91
Reifová Radka	51
Rídl Jakub	93, 99
Sarker Martin	71
Schwarz Marek	33, 94
Sedlák David	60, 82
Sedlář Karel	11
Sehnal David	36, 41
Sladeček Stanislava	98
Strnad Hynek	74, 93, 99
Svoboda Ondřej	27, 72

Svobodová Vařeková Radka	30, 36, 41, 68, 75, 80, 91
Svozil Daniel	26, 35, 59, 60, 62, 82, 84, 95, 97, 102
Šauer Max	27, 72
Šícho Martin	95
Škoda Petr	96
Škuta Ctibor	26, 82, 97
Štourač Jan	10, 39, 61
Treťjačenko Vjačeslav	37
Větrovský Tomáš	100
Vinař Tomáš	7
Vohradský Jiří	33
Voller Jiří	101
Vondrášek Jiří	37, 71
Voršilák Milan	102

Anděl Michael (*andelmi2@fel.cvut.cz*)

Ústav experimentální medicíny AV ČR, Praha

Bartoň Tomáš (*tomas.barton@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Bartůněk Petr (*bartunek@img.cas.cz*)

ÚMG AV ČR, Praha 4

Bauerová Vladena (*vladena.bauerova@savba.sk*)

Ústav molekulárnej biológie SAV, Bratislava

Bazgier Václav (*vaclav.bazgier@upol.cz*)

Katedra fyzikální chemie, Přírodovědecká fakulta, Univerzita Palackého v Olomouci, Olomouc

Bendl Jaroslav (*jaroslav.bendl@gmail.com*)

Fakultní nemocnice u sv. Anny v Brně, Brno

Beňová Zuzana (*zu.benova@gmail.com*)

VŠCHT - příprava ke studiu, Praha

Berka Karel (*karel.berka@upol.cz*)

Univerzita Palackého v Olomouci, Olomouc

Božíková Paulína (*paulina.bozikova@gmail.com*)

Biotechnologický ústav, AV ČR, Vestec

Brezovský Jan (*brezovsky@mail.muni.cz*)

Fakultní nemocnice u sv. Anny v Brně, Brno

Brož Petr (*bioinformatician.cz@gmail.com*)

Institute of Applied Biotechnologies a.s., Praha

Buček Aleš (*bucek.ales@gmail.com*)

Ústav organické chemie a biochemie, AV ČR, Praha 6

Budiš Jaroslav (*budis@dcs.fmph.uniba.sk*)

Geneton s. r. o., Bratislava

Burgetová Ivana (*burgetova@fit.vutbr.cz*)

Vysoké učení technické v Brně, Fakulta informačních technologií, Brno

Cehák Jiří (*jiri.cephak@outlook.cz*)

Jihočeská univerzita, České Budějovice, Přírodovědecká fakulta, Ústav aplikované informatiky

Čech Petr (*cechp@vscht.cz*)

VŠCHT Praha, Praha 6

Černý Martin (*cerny.m@gmail.com*)

Mikrobiologický ústav AV ČR, v. v. i., Praha

Čmelo Ivan (*cmelo.ivan@gmail.com*)

VŠCHT Praha, Praha 6

Damborský Jiří (*jiri@chemi.muni.cz*)

Fakultní nemocnice u sv. Anny v Brně, Brno

Dehaen Wim (*wimdehaen@gmail.com*)

VŠCHT Praha, Praha 6

Dohnalová Hana (*hannicka12@seznam.cz*)

Hana Dohnalová, Praha 412

Doležalová Marie (*d.marienka@seznam.cz*)

Jihočeská univerzita, Přírodovědecká fakulta, Ústav aplikované informatiky, České Budějovice

Dvořák Pavel (*pavel.dvorak@uochb.cas.cz*)

Ústav organické chemie a biochemie AV ČR, v. v. i., Praha 6

Đuriš František (*mejlnafsetko@gmail.com*)

Geneton s.r.o., Bratislava

Ehler Edvard (*eda.ehler@seznam.cz*)

Katedra biologie a environmentálních studií, Univerzita Karlova v Praze Pedagogická fakulta

Faltýnková Květoslava (*kveta.faltynkova@lag.agel.cz*)

Laboratoře AGEL a.s., Nový Jičín

Fatka Jiří (*fatkaj@ntis.zcu.cz*)

Západočeská univerzita v Plzni, Plzeň

Fesl Jan (*fesl@post.cz*)

Jihočeská univerzita, Přírodovědecká fakulta, Ústav aplikované informatiky, České Budějovice

Fišer Karel (*karel.fiser@lfmotol.cuni.cz*)

CLIP - Laboratorní centrum Klinika dětské hematologie a onkologie 2. lékařská fakulta, UK, Praha

Galgonek Jakub (*jakub.galgonek@uochb.cas.cz*)

Ústav organické chemie a biochemie AV ČR, v. v. i., Praha 6

Gazdarica Juraj (*juraj.gazdarica@gmail.com*)

Geneton s.r.o., Bratislava

Geidl Stanislav (*standag@chemi.muni.cz*)

Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta, Masarykova univerzita, Brno

Hejret Václav (*408320@mail.muni.cz*)

Masarykova univerzita CEITEC MU, Brno

Hoksza David (*david.hoksza@mff.cuni.cz*)

Univerzita Karlova, Praha

Hon Jiří (*ihon@fit.vutbr.cz*)

Fakulta informačních technologií, VUT v Brně, Opava-Milostovice

Hoňková Kateřina (*katerina.honkova@astronomie.cz*)

Oddělení genetické ekotoxikologie, Ústav experimentální medicíny AV ČR, v. v. i., Praha 4

Horský Vladimír (*vl.horsky@gmail.com*)

Masarykova Univerzita, CEITEC - Středoevropský technologický institut, Brno

Hošek Petr (*hosekp@vscht.cz*)

VŠCHT Praha, Praha 6

Hošťáková Nina (*nhostakova@gmail.com*)

Biologické centrum AV ČR, v. v. i., České Budějovice

Hron Tomáš (*tomas.hron@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Jakubec Dávid (*jakubecd@natur.cuni.cz*)

Ústav organické chemie a biochemie AV ČR, Praha 6

Janoušek Václav (*vjanousk@gmail.com*)

Katedra zoologie, Přírodovědecká fakulta UK v Praze, Praha 2

Jansík Branislav (*branislav.jansik@vsb.cz*)

IT4Innovations, VSB-TUO, Ostrava Poruba

Jarkovská Karla (*kjarkovska@chemaxon.com*)

ChemAxon s.r.o., Praha

Jelínek Jan (*jelinek@ksi.mff.cuni.cz*)

Univerzita Karlova v Praze; Matematicko-fyzikální fakulta, Praha 2

Jindřich Jindřich (*jindrich@img.cas.cz*)

Ústav molekulární genetiky, AV ČR, v. v. i, Praha 4

Jirát Jiří (*Jiri.Jirat@vscht.cz*)

VŠCHT Praha, Praha 6

Kahle Michal (*kahle@img.cas.cz*)

ÚMG AV ČR, Praha 4

Kolář Michal (*kolarmi@img.cas.cz*)

Ústav molekulární genetiky AV ČR, Praha 4

Krejčí Adam (*adam.krejci@mou.cz*)

Regionální centrum aplikované molekulární onkologie, Masarykův onkologický ústav, Brno

Křenek Aleš (*ljocha@ics.muni.cz*)

Ústav výpočetní techniky, MU, Brno

Kucharík Marcel (*mkmarcelgg@gmail.com*)

Geneton s.r.o., Bratislava

Kužmová Erika (*kuzmovae@gmail.com*)

Ústav organické chemie a biochemie AV ČR, v. v. i., Praha 6

Lexa Matej (*lexa@fi.muni.cz*)

Masarykova univerzita, Fakulta informatiky, Brno

Macas Jiří (*macas@umbr.cas.cz*)

Biologické centrum AV ČR, České Budějovice

Malinka František (*f.malinka@centrum.cz*)

České vysoké učení technické v Praze, Praha

Mareška Václav (*mareskav@vscht.cz*)

VŠCHT Praha, Praha 6

Martínek Tomáš (*martinto@fit.vutbr.cz*)

Vysoké učení technické v Brně, Fakulta informačních technologií, Brno

Midlik Adam (*midlik@mail.muni.cz*)

NCBR, Masarykova univerzita, Brno

Mořkovský Libor (*morkovsk@natur.cuni.cz*)

Univerzita Karlova v Praze, Praha 2

Müller Tomáš (*tomas.muller@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Musil Miloš (*imusilm@fit.vutbr.cz*)

Vysoké učení technické v Brně, Fakulta informačních technologií, Brno

Novák Petr (*petr@umbr.cas.cz*)

Biologické Centrum AV ČR, v. v. i., České Budějovice

Novotný Jiří (*fg-42@seznam.cz*)

VŠCHT Praha, Praha 6

Novotný Marian (*marian@natur.cuni.cz*)

Univerzita Karlova v Praze, Praha 2

Nunvář Jaroslav (*jaroslav.nunvar@lfmotol.cuni.cz*)

Ústav lékařské mikrobiologie 2. LF UK a FN Motol, Praha

Oppelt Jan (*jan.oppelt@mail.muni.cz*)

NCBR & CEITEC MU, Brno

Ostašov Pavel (*pavel.ostasov@lfp.cuni.cz*)

Lékařská fakulta v Plzni, Univerzita Karlova v Praze, Plzeň

Pačes Jan (*hpaces@img.cas.cz*)

Ústav molekulární genetiky AV ČR, Praha

Pánek Josef (*panek@biomed.cas.cz*)

MBÚ AV ČR, Praha

Pastva Ondřej (*Ondrej.Pastva@uhkt.cz*)

Ústav hematologie a krevní transfuze, Biochemie, Praha 2

Pazúriková Jana (*pazurikova@ics.muni.cz*)

Masarykova Univerzita, Brno

Pech David (*dpech@chemaxon.com*)

ChemAxon s.r.o., Praha

Pfeiferová Lucie (*pfeiferl@vscht.cz*)

Ústav organické chemie a biochemie Akademie věd České republiky, v. v. i., Praha

Pitule Pavel (*pavel.pitule@lfp.cuni.cz*)

Lékařská fakulta UK v Plzni, Plzeň

Polishchuk Pavlo (*pavlo.polishchuk@upol.cz*)

Lékařská fakulta Univerzity Palackého v Olomouci - IMTM, Olomouc

Popr Martin (*popr@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Pravda Lukáš (*xpravda@ncbr.muni.cz*)

Masarykova univerzita - Přírodovědecká fakulta, Brno

Převorovský Martin (*prevorov@natur.cuni.cz*)

Univerzita Karlova v Praze, Přírodovědecká fakulta, Praha 2

Přistoupilová Anna (*anna.pristoupilova@lf1.cuni.cz*)

Univerzita Karlova v Praze, 1. lékařská fakulta, Praha 2

Puterová Janka (*iputerova@fit.vutbr.cz*)

Biofyzikální ústav AV ČR, Brno

Raček Tomáš (*tom@krab1k.net*)

Masarykova Univerzita; CEITEC - Středoevropský technologický institut, Brno

Reifová Radka (*radka.reifova@natur.cuni.cz*)

Přírodovědecká fakulta, Univerzita Karlova v Praze, Praha 2

Rídl Jakub (*ridlj@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha

Sarker Martin (*sarkerm@natur.cuni.cz*)

Ústav organické chemie a biochemie, Praha

Schwarz Marek (*marek.schwarz@biomed.cas.cz*)

Mikrobiologický ústav AV ČR, Praha

Sedlák David (*sedlak@img.cas.cz*)

CZ-OPENSREEN, Institute of Molecular Genetics AS CR, Prague

Sedlář Karel (*sedlar@feec.vutbr.cz*)

Vysoké učení technické v Brně, FEKT, Brno

Sehnal David (*David.sehnal@gmail.com*)

Ceitec - MU, Brno

Sladeczek Stanislava (*stanislava.sladeczekova@gmail.com*)

Univerzita Karlova v Praze, Brno

Stoček Štěpán (*stoces@ueb.cas.cz*)

SEQme s.r.o., Dobříš

Strnad Hynek (*strnad@img.cas.cz*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Svoboda Ondřej (*osvoboda@chemaxon.com*)

ChemAxon s.r.o., Praha

Svobodová Vařeková Radka (*svobodova@chemi.muni.cz*)

Masarykova Univerzita CEITEC - Středoevropský technologický institut, Brno

Svozil Daniel (*svozild@vscht.cz*)

VŠCHT Praha, Praha 6

Šauer Max (*msauer@chemaxon.com*)

ChemAxon s.r.o., Praha

Šícho Martin (*sichom@vscht.cz*)

Vysoká škola chemicko-technologická v Praze, Praha

Škoda Petr (*skoda@ksi.mff.cuni.cz*)

Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Praha 2

Škuta Ctibor (*ctibor.skuta@img.cas.cz*)

CZ-OPENSREEN, Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Štourač Jan (*stourac.jan@gmail.com*)

Fakultní nemocnice u sv. Anny v Brně, Brno

Treťjačenko Vjačeslav (*tretyacv@natur.cuni.cz*)

Přírodovědecká fakulta Univerzity Karlovy v Praze, Praha

Větrovský Tomáš (*kostelecke.uzeniny@seznam.cz*)

Mikrobiologický ústav AV ČR, v. v. i., Praha

Vinař Tomáš (*tvinar@gmail.com*)

Univerzita Komenského v Bratislave, Bratislava

Vohradský Jiří (*vohr@biomed.cas.cz*)

Mikrobiologický ústav AV ČR, v. v. i., Praha 4

Voller Jiří (*jirivoller@gmail.com*)

IMTM, Olomouc

Vondrášek Jiří (*jiri.vondrasek@uochb.cas.cz*)

ÚOCHB AV ČR, Praha 6

Voršilák Milan (*vorsilam@vscht.cz*)

VŠCHT Praha, Praha 6

Zidkova Anastasiya (*azidkova@hpst.cz*)

HPST, s.r.o., Praha

Zigo Tomáš (*tomas.zigo@slovanet.sk*)

Ústav molekulární genetiky AV ČR, v. v. i., Praha 4

Konferenční sborník ENBIK2016

Vydala: Vysoká škola chemicko-technologická v Praze
Technická 5, 166 28 Praha 6

Editoři: **Petr Čech, Daniel Svozil**

Tisk: powerprint s. r. o., Brandejsovo nám. 1219/1
165 00 Praha 6 – Suchbát

Rok vydání: 2016

Počet stran: 130

Náklad: 120 ks