

## LABORATORNÍ INFORMAČNÍ SYSTÉMY PRO TESTOVÁNÍ S VYSOKOU PROPUSTNOSTÍ

**TOMÁŠ MÜLLER, DAVID SEDLÁK a PETR BARTŮŇEK**

*CZ-OPENSREEN: Národní infrastruktura pro chemickou biologii, Ústav molekulární genetiky AV ČR v.v.i. Vídeňská 1083, 142 20 Praha 4  
bartunek@img.cas.cz, tomas.muller@img.cas.cz*

Došlo 8.8.17, přijato 9.10.17.

**Klíčová slova:** testování s vysokou propustností; laboratorní informační systém

### Obsah

1. Úvod
2. Testování s vysokou propustností
3. Požadavky na laboratorní informační systém
4. Implementace laboratorního informačního systému
5. Pracovní postup
6. Závěr

### 1. Úvod

V dnešní době je používání laboratorního informačního systému (LIMS) v podstatě nutnost pro každou laboratoř zpracovávající větší množství chemických látek. Pro samotnou správu chemických látek existuje mnoho volně či komerčně dostupných softwarových řešení<sup>1</sup>. Výběr softwarových systémů pro správu testování s vysokou propustností (HTS) je však již podstatně omezenější. Existující volně použitelná řešení buď nejsou aktivně vyvíjena, nebo jsou dokonce zcela nedostupná<sup>2,3</sup>. Z tohoto důvodu vytváříme, v rámci české Národní infrastruktury chemické biologie, systém ChemGenDB, který primárně odráží aktuální potřeby naší laboratoře, ale je zároveň vyvíjen tak, aby mohl být použit i v jiných laboratořích s podobným zaměřením.

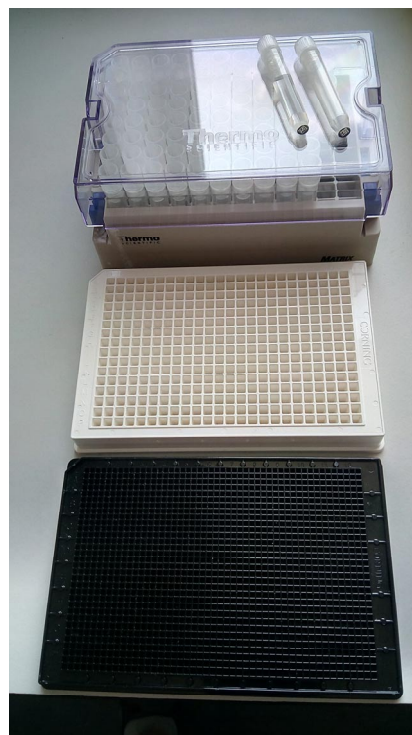
### 2. Testování s vysokou propustností

Jako testování s vysokou propustností neboli HTS (High Throughput Screening), nazýváme metodu vědeckých experimentů, kdy je velké množství chemických látek (obvykle řádově tisíce až miliony) testováno na biologickou aktivitu pro zvolený biologický cíl. Průběh experimentu je do velké míry automatizován pomocí soustavy

přístrojů a robotického ramene, které tyto přístroje obsluhuje. Chemikálie i biologický materiál jsou během testování uchovávány a přenášeny v plastových destičkách (mikrotitrační destičky). Tyto destičky (obr. 1) obsahují vícero jamek (nejčastěji 96, 384 nebo 1536), které dokáží pojmout kapalinu o objemu od desítek nanolitrů až po několik mililitrů. Biologickou složkou experimentů může být např. buňka, enzym nebo nukleová kyselina, chemickou složkou je pak nejčastěji chemická sloučenina rozpuštěná většinou v dimethylsulfoxidu (DMSO). Cílem HTS experimentů může být např. hledání potenciálních kandidátů pro vývoj nových léčiv nebo pochopení či validace některých biologických procesů.

HTS experiment obvykle probíhá v následujících krocích:

1. příprava biologického materiálu a jeho přenos na destičky,
2. přenos testovaných chemických látek,
3. inkubace,



Obr. 1. Příklad typů destiček používaných při HTS. Nahoře stojánek na 96 uzavíratelných zkumavek opatřených 2D čárovým kódem na spodní straně, schopných pojmout až 1,4 ml kapaliny. Používá se pro uchovávání vzorků. Uprostřed a dole jsou destičky pro testování ve formátu 384 a 1536

4. měření,
5. analýza.

HTS experimenty můžeme rozdělit na dvě velké skupiny: 1) primární testování, kdy se zpravidla testuje velké množství látek, každá se v experimentu vyskytuje pouze jednou a v jediné koncentraci a 2) sekundární testování – do sekundárního (též tzv. validačního) experimentu jsou vybrány látky, které prokázaly aktivitu v primárním experimentu a jejichž aktivitu je potřeba lépe charakterizovat, a proto jsou testovány v multiplikátech a ve více koncentracích.

Měření výsledků pokusu může mít mnoho podob v závislosti na konkrétním experimentu. Například je možno detegovat fluorescenci, luminiscenci nebo počítačem zpracovávat pořízené mikroskopické záběry. Pro každou jamku každé destičky tedy získáme jednu nebo více číselných hodnot. K samotným naměřeným hodnotám je často třeba ještě dodat kontext pomocí kontrol, což může být chemická látka, u které známe mechanismus působení v daném experimentu, nebo naopak jamka bez přidané chemické látky<sup>4</sup>. Jako pozitivní kontrolu používáme chemickou látku, u které v daném experimentu očekáváme biologickou aktivitu. U negativní kontroly naopak očekáváme, že biologickou aktivitu vykazovat nebude. Tyto dvě základní kontroly nám určí očekávané aktivní minimum a maximum, ve kterém by se měly pohybovat ostatní látky.

HTS experiment je nejen časově, ale i finančně náročný a je důležité se ujistit, že podmínky a postup experimentu jsou správné. Proto se často před samotným HTS provádí pilotní experiment pouze s kontrolními látkami. Z naměřených hodnot pak lze spočítat statistickou míru Z-faktor, která nám určuje vhodnost podmínek experimentu pro HTS na základě rozlišitelnosti biologické odezvy pozitivních a negativních kontrol<sup>5</sup>.

### 3. Požadavky na laboratorní informační systém

Z velice zjednodušeně nastíněné problematiky HTS je vidět, že i v základní podobě se jedná o velmi komplexní proces a jeho řízení se neobejde bez specializovaného softwarového systému, tzv. LIMS (Laboratory Information Management System). Ten by měl obsahovat následující části:

- databáze chemických struktur,
- správa vzorků,
- zadání experimentů,
- analýza experimentů,
- tvorba reportů.

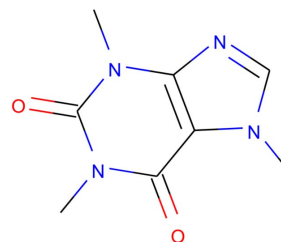
Databáze chemických struktur je základní vrstvou LIMS pro HTS, která přispívá k celkové přehlednosti celého systému a velmi významným způsobem zjednodušuje a zpřehledňuje analýzu výsledků.

Existuje mnoho způsobů počítačové reprezentace molekul<sup>6</sup>. Tím asi nejjednodušším a nejrozšířenějším je formát SMILES<sup>7</sup>, který je poměrně dobře čitelný i pro člověka. Jeho kanonická verze teoreticky umožňuje i iden-

tifikaci molekuly. V praxi se však využívá formát InChI<sup>8</sup>, který dokáže jednoznačně identifikovat každou chemickou strukturu. Jeho hašovaná forma InChIKey má podobu textového řetězce o konstantní délce 27 znaků. Ztrácí informaci o struktuře, avšak zachovává si schopnost jednoznačné identifikace, a proto je často využívána v databázích. Třetím často používaným formátem je Molfile<sup>9</sup>, který je oproti dříve zmíněným formátům mnohem větší a není lineární, ale jako jediný z této trojice si uchovává informaci o souřadnicích jednotlivých atomů a je tedy nejvhodnější k uložení grafické podoby molekuly a tvorbu obrázkových výstupů (obr. 2).

Na chemické sloučeniny těsně navazuje databáze chemických vzorků. Chemickou sloučeninu chápeme jako abstraktní entitu definovanou především svou strukturou, chemický vzorek je oproti tomu konkrétní fyzický objekt s definovaným dodavatelem, šarží, čistotou a dalšími parametry. Dále je v systému třeba zaznamenat veškeré instance daného vzorku, tzn. jeho umístění, koncentrace, použité rozpouštědlo aj. Konkrétní implementace se bude odvíjet od systému správy (Compound Management, CM<sup>1</sup>) vzorků daného pracoviště. V případě použití automatizovaného CM systému je dobré ho s LIMS přímo propojit.

V dalším kroku je potřeba ukládat informace o provedení konkrétního experimentu. K popisu experimentu lze přistupovat mnoha způsoby, od volného textu až po strukturovaná data. Nejvhodnější implementace se opět odvíjí od potřeb daného pracoviště. Ke strukturovanému popisu



A) CN1C=NC2=C1C(=O)N(C(=O)N2)C

B)

InChI=1S/C8H10N4O2/c1-10-4-9-6-5/1017(13)12(3)8(14)11(6)2/h4H,1-3H3

C)

RYYYLVZUVIJVGH-UHFFFAOYSA-N

D)

```

3519
50Chem-0171114420
24 25 9 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 7320 2 0000 0 0000 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 0000 1 0000 0 0000 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5 5440 0 0047 0 0000N 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 0000 0 0000 0 0000 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5 5440 0 0047 0 0000N 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4 0000 0 0000 0 0000C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 000 0 0000 0 0000C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 7320 1 0000 0 0000C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 0000 0 0000 0 0000C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4 1774 0 0000 0 0000C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 0000 1 0000 0 0000C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 0000 1 0000 0 0000C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1774 0 0000 0 0000C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 1120 0 0000 0 0000H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 7320 0 0000 0 0000H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6 4442 1 5426 0 0000H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6 4476 1 5446 0 0000H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6 4482 1 5426 0 0000H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6 4476 1 5446 0 0000H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 0000 1 5446 0 0000H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 4602 1 1000 0 0000H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 8990 0 4931 0 0000H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 12 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4 11 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4 11 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5 10 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5 9 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6 11 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
7 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
11 15 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
12 14 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
12 21 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
13 20 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
13 20 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
14 22 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
14 21 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
H END
9151

```

Obr. 2. Různé počítačové reprezentace struktury kofeinu. A) SMILES, B) InChI, C) InChIKey D) Molfile

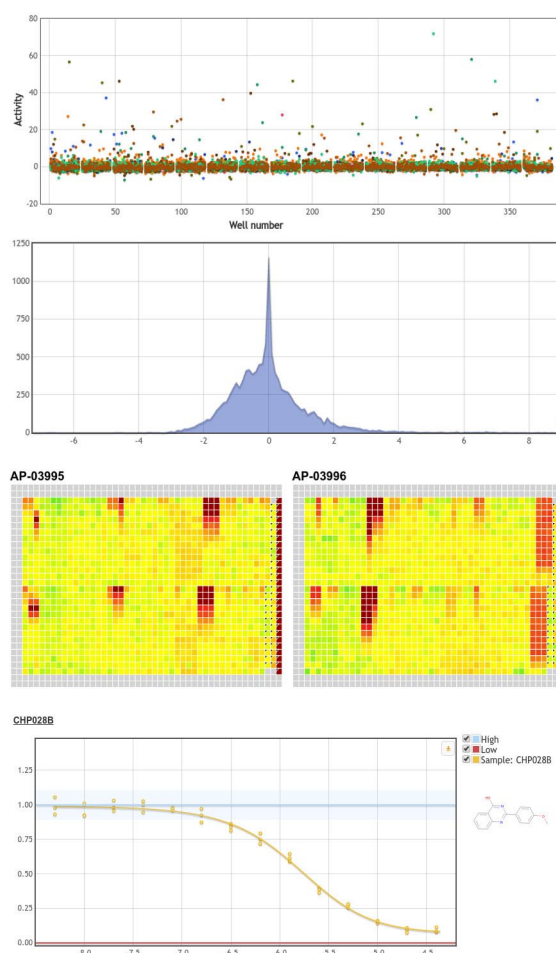
lze s výhodou využít BioAssay Ontology<sup>10</sup>, což je ontologie vytvořená přímo pro testování s vysokou propustností. Hlavní přednost použití BioAssay Ontology a ontologií obecně je správná anotace dat v souhrnných databázích, které přejímají data z mnoha různých zdrojů. Z hlediska systému je nejdůležitější správné přiřazení vzorků v destičkách a následné nahrání naměřených dat<sup>11</sup>.

Po provedení experimentu jsou naměřená data uložena do databáze a experiment lze analyzovat. HTS experimenty jsou obecně velice náchylné k systematickým chybám, které mohou být způsobeny např. problémem při pipetování, postupnou hydratací rozpouštědla, nerovnoměrnou expozicí vůči vnějším vlivům atp. Známé jsou také okrajové efekty, ke kterým dochází proto, že jamky na krajích a především v rozích destiček nemají stejné podmínky jako zbytek destičky – nejsou ze všech stran obklopeny dalšími jamkami. Proto je často žádoucí samot-

ná naměřená data před analýzou standardizovat, normalizovat či aplikovat určité korekce. Takto lze data předzpracovat mnoha způsoby, mezi často používané algoritmy patří například:

- Z-skóre – transformace jednorozměrných dat na základě jejich směrodatné odchylky,
- median polish – transformace dvourozměrných dat, počítá zvlášť mediány sloupců a řádků a snaží se odstraňovat sloupcové a řádkové efekty,
- B-score<sup>12</sup> – normalizace vyvinutá přímo pro HTS z metody median polish.

Pro porovnávání dat z různých experimentů není možné použít surová naměřená data, neboť citlivost HTS vede typicky k posunu hodnot aktivit v různých experimentech stejného typu. Z tohoto důvodu je nejprve třeba data normalizovat na hodnoty relativní vůči kontrolním látkám.



Obr. 3. Ukázka typů vizualizací dat v systému ChemGenDB. Odshora: naměřená data z primárního experimentu, normalizovaná metodou B-score vynesena v bodovém diagramu. Pod ním jsou stejná data vynesena v histogramu. Následují data ze sekundárního experimentu zobrazená ve formě destiček, kde teplotní mapa znázorňuje změřenou aktivitu vzorků. Vespod jsou data koncentrační řady jednoho vzorku v multiplikátech proložena logistickou křivkou včetně kontrol

Při analýze primárního experimentu si zvolíme hranici aktivity, od které považujeme látku v tomto experimentu za aktivní. Předpokládáme, že převážná většina testovaných vzorků nebude vykazovat biologickou aktivitu a pouze malé procento látek se bude svou naměřenou aktivitou výrazně lišit. S těmito látkami pak můžeme postoupit do dalších fází testování<sup>13</sup>.

Sekundární testování má mnohem více podob, například potvrzení aktivity, kdy provedeme celý test znovu za stejných podmínek pouze s vybranými látkami, abychom ověřili reprodukovatelnost předchozích výsledků. Častým cílem sekundárního testování je též proměření závislosti biologické aktivity na koncentraci, kde se snažíme charakterizovat biologickou odpověď proložením naměřených bodů logistickou křivkou<sup>14</sup>.

V poslední fázi by měl systém být schopen vytvořit report s kompletními výsledky experimentu.

Z výčtu výše uvedených požadavků vyplývá, že LIMS pro HTS by měl být dostatečně flexibilní, měl by být schopen ukládat různorodé množství typů výsledků a současně by měl nabízet základní vizualizační a analytické nástroje (obr. 3).

#### 4. Implementace laboratorního informačního systému

V Národní infrastruktuře chemické biologie jsme z důvodu větší flexibility přistoupili k implementaci vlastního LIMS systému ChemGenDB. ChemGenDB je přístupný jako webová aplikace a nikoliv jako samostatný klient. Toto řešení skýtá mnoho výhod, mezi které patří například velký výběr existujících nástrojů a knihoven, simultánní práce více uživatelů nebo skutečnost, že jediný potřebný software, který je třeba instalovat na klientské zařízení, je webový prohlížeč. Samozřejmě lze najít i negativa tohoto přístupu, jako jsou výkonnostní či paměťové limity klientského zařízení nebo samotného prohlížeče při práci s větším množstvím dat, rozdílné interpretace webových stránek a skriptů na straně klienta v závislosti na použitém prohlížeči nebo nutnost oddělení časově náročnějších výpočtů od serverových odpovědí<sup>15</sup>.

Veškeré použité nástroje pro ChemGenDB byly vybírány z množiny volně dostupného neplaceného softwaru. Tento výběr je poměrně velký, a proto není nutné sahat po komerčních řešeních. Také není problém výsledný software dále šířit. Pro serverovou část aplikace byl zvolen programovací jazyk Python, který představuje dobrý kompromis mezi výkonem a jednoduchostí. Zároveň pro něj existuje skutečně velká paleta užitečných knihoven<sup>16</sup>. Jednou z těchto knihoven, kterou jsme použili, je i Django<sup>17</sup>, která poskytuje mnoho nástrojů pro vývoj webových aplikací a velmi tak usnadňuje jejich vývoj. Data jsou uchovávána pomocí relačního databázového systému PostgreSQL, který byl vybrán pro svou robustnost a velmi dobrý výkon<sup>18,19</sup>. Pro manipulaci s chemickými daty byla zvolena knihovna RDKit<sup>20,21</sup>, která je zároveň schopna rozšířit funkcionalitu PostgreSQL (RDKit database cartridge) inte-

grací funkcí pro práci s chemickými strukturami na úrovni databáze (např. hledání dle podobnosti nebo podstruktury), které jsou tak řádově rychlejší.

Náročnější výpočetní úlohy jsou realizovány paralelně za pomoci úkolové fronty Celery<sup>22</sup>. Například pouhé získání a zpracování dat pro analýzu většího experimentu (statisíce či miliony látek) nebo nahrání takového počtu naměřených dat může trvat řádově i desítky vteřin, což už je pro webovou aplikaci nevhodné. Stejně tak by mohlo dojít k zahlcení aplikace současným odesláním několika takto náročných úloh najednou. Úkolová fronta Celery funguje nezávisle na zbytku aplikace a lze snadno omezit počet úloh, které se zpracovávají najednou. Ostatní úlohy se řadí do fronty a spustí se až ve chvíli, kdy na ně dojde řada.

Systém je řešen modulárně, protože v mnoha případech nelze z důvodu velké complexity řešit dané situace dostatečně obecně a zároveň uživatelsky přívětivě. Proto je možné pro některé části systému naprogramovat zásuvný modul, který řeší konkrétní problém. Toho se dá využít zejména při integraci LIMS s laboratorními přístroji, aby bylo možné například nahrávat data přímo z přístrojů měřících aktivitu testovaných látek nebo vytvářet instrukce pro automatizovaný CM a minimalizovat tak vznik chyb při správě vzorků.

#### 5. Pracovní postup

První krok práce v systému ChemGenDB je vytvoření obsahu databáze chemických vzorků. Toho lze dosáhnout jednoduše pomocí nahrání textového souboru s informacemi o chemických vzorcích, jejich strukturách a umístění na destičkách. Samozřejmostí je i ruční zadávání vzorků a destiček včetně kreslení nových chemických struktur. Při nahrání nové struktury jsou automaticky spočítány dostupné fyzikálně-chemické parametry a systém se pokusí propojit danou strukturu s veřejně dostupnými chemickými databázemi, jako jsou například PubChem, ChemSpider, ChEMBL nebo ZINC (obr. 4). Z těchto databází jsou též získány triviální a systematické názvy odpovídající dané struktuře. Dále je možno pracovat s celými destičkami, reformátovat je a připravovat tak pro experiment. Propojení mezi systémem a fyzickými vzorky je realizováno přes čárové kódy, které systém umožňuje přidělovat, tisknout a spravovat.

Experiment se v systému vytváří pomocí formuláře, kam je možné zadat jeho atributy, především počet a typ měření a množinu destiček, které do experimentu vstupují. Naměřená data jsou po provedení experimentu automaticky nahrána do systému. Následně lze experiment analyzovat. Ještě před samotnou analýzou je možné data normalizovat spuštěním jedné nebo více přednastavených normalizačních metod, případně je možné zadat i vlastní vzorec normalizace. Samotná analýza poskytuje různé pohledy na naměřená data pomocí několika vizualizací. Teplotní mapa (heatmap) vytváří pohled na destičky postupně tak, jak byly měřené, barevná škála znázorňuje změřenou aktivitu

Obr. 4. Ukázka rozhraní systému ChemGenDB: databáze chemických struktur. Typické třísloupcové rozložení – v levém sloupci seznam objektů, uprostřed hlavní pracovní oblast a v pravém sloupci vybraný seznam s možností exportu

(obr. 3 a obr. 5). Pokud zde nalezneme nějaká problematická místa, můžeme je zkusit ošetřit normalizací nebo přímo na destičce místo označit jako chybu a vyřadit tyto body z další analýzy. Bodový diagram a histogram nám pak poskytuje celkový pohled na aktivitu všech látek v experimentu, což nám především při analýze primárního

experimentu umožní snadné stanovení hranice aktivity pro přijetí látek do dalšího testování (obr. 3).

Při analýze sekundárního experimentu je k dispozici modul, který prokládá naměřenými daty logistické křivky s možností nastavení či omezení všech parametrů, filtrování na základě dat i spočítaných parametrů a posouzení

Obr. 5. Rozhraní analýzy sekundárního experimentu. V prostředním sloupci je nástroj na výpočet křivky závislosti aktivity na koncentraci. Odshora – vstupní parametry výpočtu, možnosti stránkování a filtrování křivek, volby zobrazení, uložení křivek a export, samotné zobrazení křivek a naměřených dat. Vespod je část teplotních map destiček



kvality výsledné křivky. Parametry těchto křivek pak poslouží jako indikátor aktivity (obr. 5). Systém obsahuje také nástroje pro porovnávání výsledků z více experimentů, souhrnné zobrazení dat, nástroje pro shlukovou analýzu a analýzu vztahu struktury a aktivity.

## 6. Závěr

V současné době probíhá vývoj systému ChemGenDB pro potřeby infrastruktury CZ-OPENSREEN na Ústavu molekulární genetiky Akademie věd ČR. Design systému se primárně odvíjí od aktuálních potřeb a postupů laboratoře, ale zároveň je kladen důraz na možnost širšího použití a nasazení. Nyní systém pokrývá potřeby pro kompletní správu chemických vzorků, flexibilní design experimentu a několik nástrojů analýzy výsledků. Pro komplexnější analýzy nad rámec systému je možné provést export veškerých relevantních dat.

V blízké budoucnosti se počítá s vývojem podsystému pro správu inventáře, rozvojem analytických nástrojů a především s rozšiřováním možností navrhování složitějších experimentů.

*Tento článek vznikl za podpory MŠMT v rámci Národního programu udržitelnosti I projekt LO1220 (CZ-OPENSREEN).*

## LITERATURA

1. Popr M., Sedlák D., Bartůňek P.: *Chem. Listy* 111, 772 (2017).
2. Tai D., Rathnam C., Jianwen F.: *Comb. Chem. High Throughput Screen.* 14, 757 (2011).
3. Tolopko A. N., Sullivan J. P., Erickson S. D., Wrobel D., Chiang S. L., Rudnicki K., Rudnicki S., Nale J., Selfors L. M., Greenhouse D., Muhlich J. L., Shamu C. E.: *BMC Bioinf.* 11, 260 (2010).
4. Pereira D. A., Williams J. A.: *Br. J. Pharmacol.* 152, 53 (2007).
5. Zhang J.-H., Chung T. D. Y., Oldenburg K. R.: *J. Biomol. Screening* 4, 67 (1999).
6. Jiráť J., Svozil D.: *Chem. Listy* 111, 710 (2017).
7. Weininger D.: *J. Chem. Inf. Comput. Sci.* 28, 31 (1988).
8. Heller S., McNaught A., Stein S., Tchekhovskoi D., Pletnev I.: *J. Cheminformatics* 5, 7 (2013).
9. Dalby A., Nourse J. G., Hounshell W. D., Gushurst A. K. I., Grier D. L., Leland B. A., Laufer J.: *J. Chem. Inf. Comput. Sci.* 32, 244 (1992).
10. Visser U., Abeyruwan S., Vempati U., Smith R. P., Lemmon V., Schürer S. C.: *BMC Bioinf.* 12, 257 (2011).
11. Kevorkov D., Makarenkov V.: *J. Biomol. Screening* 10, 557 (2005).
12. Brideau C., Gunter B., Pikounis B., Liaw A.: *J. Biomol. Screening* 8, 634 (2003).
13. Keseru G. M., Makara G. M.: *Drug Discov. Today* 11, 741 (2006).
14. Parham F., Austin C., Southall N., Huang R., Tice R., Portier C.: *J. Biomol. Screening* 14, 1216 (2009).
15. Leff A., Rayfield J. T.: *Fifth IEEE International Enterprise Distributed Object Computing Conference: Web-application development using the Model/View/Controller design pattern, 7th July 2001* (bez editora), str. 118.
16. Pilgrim M.: *Dive Into Python*. Apress, Berkeley, CA: New York 2004.
17. Forcier J., Bissex P., Chun W. J.: *Python Web Development with Django*. Addison-Wesley Professional, Upper Saddle River, NJ 2008.
18. Momjian B.: *PostgreSQL: Introduction and Concepts*. Pearson Education, Boston, Mass. 2000.
19. <https://www.postgresql.org>, staženo 24. 5. 2017.
20. Landrum G.: *RDKit UGM 2012: The RDKit: History and Status, London, 4th October 2012*.
21. <http://www.rdkit.org/>, staženo 24. 5. 2017.
22. <http://www.celeryproject.org>, staženo 24. 5. 2017.

**T. Müller, D. Sedlák, and P. Bartůňek** (CZ-OPENSREEN: National Infrastructure for Chemical Biology, Institute of Molecular Genetics of the Academy of Sciences of the Czech Republic, Prague): **Laboratory Information Systems to High-Throughput Screening**

High-throughput screening is a very complex process of large-scale testing of thousands, potentially millions of chemical compounds for their biological activity in a single experiment. Because of its large scale nature there is a great need for some LIMS (Laboratory Information Management System), that would keep track of your chemical samples and screening results. The cornerstone of the whole system should be a database of chemical structures, potentially enriched by their physical and chemical properties linked with actual chemical samples under study and their physical location in a storage system. Having this information, one should be able to create a screening experiment and assign measured values to one's samples in specified concentrations. With all these data combined, one can process, visualize and analyze the results and either to confirm that no mistakes have been made or to backtrack the error if some problems have arisen. ChemGenDB is such a LIMS. It is being developed at Institute of Molecular Genetics and it should cover all demands mentioned above and much more.