

Strokovni članek ■

## DrugBank – namenska spletna podatkovna zbirka o zdravilih

## DrugBank – Specialised Web- Enabled Drug Database

**Polonca Ferk, Brane Leskošek**

**Izvleček.** DrugBank je namenska prosto dostopna spletna podatkovna zbirka o zdravilih, ki vključuje podrobne, kakovostne in najsodobnejše farmacevtske, kemijske, fiziološke, farmakološke in klinične podatke o zdravilih in njihovih tarčnih molekulah. Zbirka je edinstvena z bioinformacijskega in kemijsko informacijskega vidika, saj ima vgrajena napredna orodja za brskanje, iskanje, rudarjenje po besedilu, za prikaz, razvrščanje, povzemanje podatkov ter za preverjanje ustreznosti podatkov ob vnosu. Mnoga podatkovna polja iz zbirke so povezana z drugimi spletnimi podatkovnimi viri. Široka uporabnost zbirke DrugBank v farmaciji in farmakologiji je daleč preseгла začetna pričakovanja.

**Abstract.** DrugBank is a specialised freely-available web-enabled drug database. It contains detailed, high-quality and up-to-date pharmaceutical, chemical as well as physiological, pharmacological and clinical data about drugs and drug targets. The database is a unique bioinformatics and cheminformatics resource with built-in advanced tools for data browsing, searching, mining, viewing, sorting, extracting and improved data handling. Many data fields from the database are hyperlinked to other web data resources. DrugBank's wide range of applications in pharmacy and pharmacology has highly exceeded initial expectations.

■ **Infor Med Slov:** 2010; 15(1): 30-38

---

Institucije avtorjev: Medicinska fakulteta, Univerza v Mariboru (PF, BL); Inštitut za biostatistiko in medicinsko informatiko, Medicinska fakulteta, Univerza v Ljubljani (BL).

Kontaktna oseba: Polonca Ferk, Medicinska fakulteta, Univerza v Mariboru, Slomškov trg 15, 2000 Maribor.  
e-pošta: polonca.ferk@guest.arnes.si

Prejeto / Received: 17.05.2010  
Sprejeto / Accepted: 30.06.2010

## Uvod

Farmacija in farmakologija sta veda, ki se vsaka s svojega zornega kota ukvarjata z zdravili; prva celovito, tj. z načrtovanjem zdravil, zagotavljanjem njihove kakovosti, učinkovitosti in varnosti izdelave in uporabe ter z optimizacijo farmakoterapije, druga pa z medsebojnimi interakcijami med zdravilom in telesom. V okviru farmakologije farmakokinetika preučuje gibanje zdravila po telesu (absorpcija, distribucija, metabolizem, eliminacija zdravila - ADME) in farmakokinetične interakcije med zdravili, farmakodinamika pa obravnava učinke in mehanizme delovanja zdravil ter farmakodinamične interakcije med zdravili<sup>1</sup>. S številnimi novimi dognanji je znanje o zdravilih in njihovih tarčnih molekulah vse bolj obsežno. V zvezi s tem se je pokazala nujna potreba po ustreznih spletnih podatkovnih zbirkah s smiselno in logično urejenimi podatki.

Prva tovrstna zbirka o zdravilih je bila TTD (angl. *Therapeutic Target Database*), ki danes vsebuje seznam imen več kot 1100 nizkomolekularnih zdravil in tarčnih molekul za zdravila. V nadaljevanju so razvili obsežnejše podatkovne zbirke o zdravilih (npr. KEGG,<sup>2</sup> ChEBI,<sup>3</sup> PubChem<sup>4</sup> in UniProt/Swiss-Prot<sup>5</sup>), ki poleg imen spojin vsebujejo še sinonime, slike, podatke o strukturi ter povezave do drugih (spletnih) podatkovnih zbirk, ne vsebujejo pa podatkov o tarčnih molekulah za zdravila. Iz njih lahko dobimo le osnovne pregledne informacije o zdravilih. Navedene zbirke so namenjene zlasti farmacevtskim in sinteznim kemikom. Na drugi strani imamo na voljo specifične podatkovne zbirke o zdravilih (npr. PharmGKB<sup>6</sup> in RxList<sup>7</sup>), ki vsebujejo podrobnejše farmacevtske in farmakološke informacije o zdravilih. Njihova omejitev je, da ne vsebujejo strukturnih, kemijskih in fizikalno-kemijskih podatkov. Takšne zbirke so namenjene pretežno farmacevtom in zdravnikom, delno tudi bolnikom.

## Osnovne značilnosti podatkovne zbirke DrugBank

DrugBank<sup>8</sup> je posebna namenska spletna podatkovna zbirka o zdravilih, ki poskuša združiti lastnosti prej omenjenih zbirk. Vključuje tako podrobne in kakovostne fizikalno-farmaceutsko-kemijske (strukture, zaporedja) kot tudi farmakološke (farmakodinamične, farmakokinetične, farmakogenomske), fiziološke, molekularne in klinične podatke o zdravilih in njihovih tarčnih molekulah. Na ta način DrugBank podaja (skoraj) celovito informacijo o zdravilih. DrugBank je tudi edinstvena bioinformacijska in kemijsko informacijska podatkovna zbirka zaradi uporabe sodobnih orodij za iskanje informacij in zaradi možnosti urejenega povezovanja podrobnih podatkov o zdravilih in njihovih tarčnih molekulah, pri čemer uporabnika hkrati usmerja tudi na ustrezne zunanje spletne podatkovne vire.<sup>2-7,9-12</sup>

V DrugBank so vključeni podatki za skoraj 4800 spojin (učinkovin), razvrščenih v šest kategorij (Drug Types), ter za več kot 14000 tarčnih molekul za zdravila (proteinov oz. deoksiribonukleinskih kislin - DNA). Med (potencialnimi) učinkovinami jih je trenutno:

1. več kot 1350 nizkomolekularnih, odobrenih s strani Ameriške agencije za zdravila, FDA (angl. *Food and Drug Administration*),
2. 123 biotehnoških (proteinov, peptidov oz. nukleinskih kislin), odobrenih s strani FDA,
3. 71 nutricevtikov oz. mikronutrientov, npr. aminokislin, vitaminov, drugih metabolitov,
4. več kot 3243 eksperimentalnih spojin - potencialnih zdravilnih učinkovin v fazi predkliničnih oz. kliničnih študij (v to kategorijo so uvrščene tudi s strani FDA neodobrene učinkovine, inhibitorji encimov in potencialni toksini),
5. 57 učinkovin, ki so izgubile dovoljenje za promet,
6. 188 prepovedanih učinkovin.<sup>8</sup>

Kriteriji za vnos podatkov o spojinah v DrugBank so, da molekula vsebuje več kot en tip atoma, da ima znano kemijsko strukturo ter znane učinke oz. je opredeljena kot potencialna zdravilna učinkovina v vsaj enem od zanesljivih podatkovnih virov.<sup>8,9</sup>

Zbirka DrugBank je razdeljena na preglednice s kratkimi povzetki (Summary Tables) za vsako učinkovino oz. spojino posebej. Po preglednicah s povzetki je možno hitro brskanje, razvrščanje in preoblikovanje podatkov. Povzetki so prikazani na podoben način kot povzetki v iskalnem sistemu PubMed zbirke Medline.<sup>8,13</sup> Preglednice so povezane s karticami o zdravilih (DrugCards), narejenimi po analogiji že preverjenega in odlično sprejetega koncepta GeneCards.<sup>14,15</sup> S pritiskom na gumb DrugCard v povzetku se odpre spletna stran, ki podaja podrobnejše informacije o željeni spojin. Kartica o zdravilu vsebuje v trenutni različici 107 podatkovnih polj,<sup>8</sup> pri čemer se približno polovica polj nanaša na zdravilo (tabela 1), druga polovica pa na tarčne molekule za to zdravilo (tabela 2). Mnoga podatkovna polja so preko spletnih povezav povezana še z drugimi spletnimi podatkovnimi zbirkami, npr. KEGG,<sup>1</sup> ChEBI,<sup>3</sup> PubChem,<sup>4</sup> Swiss-Prot,<sup>5</sup> PDB,<sup>16</sup> GeneCards,<sup>15</sup> GenBank,<sup>17</sup> Wikipedia,<sup>18</sup> PDRHealth,<sup>19</sup> DPD,<sup>20</sup> HGNC<sup>21</sup> in GenAtlas,<sup>22</sup> ter s povzetki, digitalnimi slikami, različnimi interaktivnimi javanskimi programi za prikaz struktur.<sup>8,12</sup>

DrugBank je bila zasnovana predvsem z namenom lažjega računalniškega odkrivanja tarčnih molekul za zdravila, bolj optimalnega načrtovanja kemijskih struktur potencialnih zdravilnih učinkovin, sidranja (angl. *docking*) ali presejanja (angl. *screening*) učinkovin, napovedovanja metabolizma zdravil, (3D) interakcij med zdravili, primerjave strukturnih podobnosti med učinkovinami.<sup>8,10</sup> Kasneje se je z vključitvijo dodatnih kategorij podatkov o zdravilih (npr. podatkov o klinično pomembnih genetskih polimorfizmih posameznih nukleotidov (SNP, angl. *Single Nucleotide Polymorphism*), relevantnih za napoved farmakološkega odgovora apliciranega zdravila) uporabnost zbirke DrugBank razširila še na druga, tudi interdisciplinarna področja (npr.

farmakogenomske raziskave).<sup>11</sup> Zbirka predstavlja tudi izjemno obsežen in podroben vir informacij za splošno farmacevtsko izobrazbo in je namenjena za uporabo raziskovalcem, strokovnjakom in učiteljem iz različnih področij, ki se ukvarjajo z zdravili.<sup>8</sup>

Zbirko DrugBank je razvila raziskovalna skupina prof. Wisharta z Univerze v Alberti iz Kanade, pri čemer so sodelovali strokovnjaki s področij računalniških znanosti, bioinformatike, farmacije, kemije, medicine in biologije. Njen razvoj je bil dolgotrajen in vsebinsko obsežen projekt, v okviru katerega so sestavili, preverili in potrdili raznolike podatke iz različnih virov v enotno zbirko. Relevantne informacije so črpali iz več deset učbenikov, več sto člankov iz revij s faktorjem vpliva in iz okoli 30 elektronskih zbirk podatkov. Pri tem so uporabili številna spletna programska orodja. Razvili so tudi lastna programska orodja, zlasti orodja za napredno iskanje podatkov po zbirki. V vodstveno skupino projekta sta bila vključena dva akreditirana farmacevta, zdravnik in trije bioinformatiki, dodatno izobraženi s področja računalniških znanosti ter molekularne biologije in (bio)kemije.

Zbirka je prosto dostopna na spletni strani <http://www.drugbank.ca/>. Prva različica je bila 1.0, sledila je različica 2.0, najnovejša pa je različica 2.5. Trenutno zajema zbirka okoli 16 GB podatkov (besedilo, kemijska zaporedja, strukture in slike). Večina teh podatkov je javno dostopnih in jih uporabnik lahko prenese k sebi.<sup>8</sup>

**Tabela 1** Tipi podatkov o zdravilu na kartici zdravila iz zbirke DrugBank.

datum prvega vnosa v kartico o zdravilu
datum zadnje posodobitve kartice o zdravilu
disociacijska konstanta (pKa) za nizkomolekularne spojine oz. izoelektrična točka (pI) (če je učinkovina proteinska molekula)
eksperimentalna in predvidena vodotopnost (logS)
eksperimentalna in predvidena vrednost porazdelitvenega koeficienta za nizkomolekularne spojine oz. hidrofobnosti za proteine
eksperimentalni permeabilnostni koeficienti na celicah Caco-2
farmaceutske oblike
farmakokinetični podatki
fizikalno stanje
generično ime, sinonimi in zaščitena imena zdravila
identifikacijske številke (ID): PharmGKB ID, CAS ID, KEGG ID, PubChem ID, ChEBI ID, GenBank ID, SwissProt ID, PDB ID, InChI ID, HPRD ID, HGNC ID, GeneCards ID, Gene Atlas ID
ime encima faze 1, ki presnavlja učinkovino
ime po IUPAC oz. standardno kemijsko ime učinkovine
ime po SwissProt, če je zdravilo peptid oz. protein
imena organizmov, pri katerih je zdravilo najbolj učinkovito
informacije za pacienta / informacije za zdravnika
interakcije med zdravili, interakcije s hrano
izomerni in standardni nizi SMILES, ki se nanašajo na strukturo učinkovine
kategorija
kemijska formula učinkovine
mehanizem delovanja, indikacije, kontraindikacije, farmakoterapevtska skupina
monozotopna molekulska masa
opis učinkovine
oznaka FDA, oznaka InChIKey
povprečna molekulska masa učinkovine
primarna in sekundarna številka vnosa
referenca ali številka patenta, ki se nanaša na sintezo učinkovine
reference o drugih informacijah o učinkovini
slika in besedilna različica 3D strukture v formatu PDB
slika in besedilna različica strukture spojine v datoteki formata MOL ter besedilna različica strukture spojine v datoteki formata SDF
slika kemijske strukture za nizkomolekularne učinkovine oz. zaporedje za biotehnoške učinkovine
slike spektrov: NMR, masni
spletne povezave do RxList, Wikipedia, PDRhealth, MSDA, FDA, GenBank, Swiss-Prot, PubChem, KEGG, ChEBI
SwissProt ID encima faze 1, ki presnavlja učinkovino
temperatura tališča oz. vrelišča
varnostni list (MSDS; angl. <i>Material Safety Data Sheet</i> )
zaporedje encima faze 1, ki presnavlja učinkovino

Legenda: CAS - oddelek Ameriškega kemijskega društva, pooblaščen za dodelitev enoznačnega številčnega identifikatorja vsaki kemikaliji (angl. *Chemical Abstract Service*); DIN - ID po kanadskem ID sistemu zdravil; NMR - jedrska magnetna resonanca (angl. *Nuclear Magnetic Resonance*); InChI - mednarodni kemijski identifikator (angl. *International Chemical Identifier*); HPRD - podatkovna zbirka o človeških proteinih (angl. *Human Protein Reference Database*); IUPAC - Mednarodno združenje za čisto in uporabno kemijo (angl. *International Union of Pure and Applied Chemistry*); SMILES - posebna množica simbolov (angl. *Simplified Molecular Input Line Entry Specification*).

**Tabela 2** Tipi podatkov o tarčni molekuli na kartici zdravila iz zbirke DrugBank.

GenBank ID za gen tarčne molekule, GenBank ID za proteinsko tarčno molekulo, GeneCard ID, GenAtlas ID, HGNC ID, SwissProt ID, PDB ID, DIN
ime gena tarčne molekule
ime po SwissProt
ime tarčne molekule in sinonimi
imena in ID domen Pfam
kromosomska lokacija, lokus, seznam SNP-jev in mutacij
lokacija signalnega peptida ali drugih lokalizacijskih signalov v zaporedju tarčne molekule
lokacija v celici (citoplazma, jedro, membrana) oz. v okolici celice
molekulska masa
obdelava podatkov za tarčno molekulo: ročno ali s posebnim programom za rudarjenje po besedilu (PolySearch)
pomen tarčne molekule za sposobnost preživetja organizma
razvrstitev tarčne molekule glede na ontologijo genov, vključujoč funkcijo, celične procese in lokacijo
reakcije in signalne poti, v katerih je tarčna molekula udeležena
reference
splošne in specifične funkcije tarčne molekule
število aminokislin (če je tarčna molekula protein) oz. nukleotidov (če je tarčna molekula DNA)
število in lokacija transmembranskih regij tarčne molekule
teoretična izoelektrična točka
zaporedje aminokislin (če je tarčna molekula protein) oz. nukleotidov (če je tarčna molekula DNA)
zaporedje DNA (gena) za tarčno molekulo

Legenda: Pfam - podatkovna zbirka o družinah proteinov. Posamezna polja se lahko večkrat ponovijo, kajti učinkovina ima lahko več tarčnih molekul.

V najnovejšo različico so dodatno vključeni številni novi podatki. Med njimi so z aplikativnega vidika še zlasti zanimivi in uporabni podatki o interakcijah med zdravili in interakcijah zdravil s hrano ter eksperimentalni farmakokinetični podatki. S tem se je uporabnost zbirke DrugBank še dodatno razširila zlasti v strokovnih kliničnih krogih. Zbirka je trenutno najpopolnejša in najsodobnejša javno dostopna spletna podatkovna zbirka o interakcijah med zdravili in interakcijah zdravil s hrano.<sup>8</sup>

## Orodja v zbirki DrugBank

DrugBank omogoča preko spletnega vmesnika iskanje po vseh podatkovnih poljih (angl. *fully searchable*). Vgrajena ima mnoga orodja za brskanje (angl. *browsing*), iskanje (angl. *searching*), prikaz (angl. *viewing*), razvrščanje (angl. *sorting*) in povzemanje (angl. *extracting*) podatkov o učinkovinah in njihovih tarčnih molekulah.<sup>8</sup>

Ključna prednost zbirke DrugBank v primerjavi z drugimi spletnimi viri podatkov o zdravilih so zlasti naprednejša iskalna orodja z več izbirnimi funkcijami, npr. lastno iskalno orodje za primerjavo zaporedij BLAST (angl. *Basic Local Alignment Search Tool*), običajni iskalnik z Boolovimi operatorji, orodje za iskanje (podobnosti) kemijskih struktur (ChemQuery), orodje za oblikovanje (povzemanje) informacij iz manj strukturiranih relacijskih podatkov (Data Extractor).<sup>8</sup>

### Brskala orodja

Zbirka podpira tako splošno brskanje po podatkih (z uporabo izbire DrugBank Browse) kot tudi usmerjeno brskanje po podatkih. Usmerjeni brskalni orodji sta PharmaBrowse, namenjeno zlasti farmacevtom, zdravnikom in farmacevtskim kemikom, ki iščejo informacije o zdravilu v smislu indikacij in farmakoterapevtske skupine, ter orodje GenoBrowse, namenjeno brskanju po genetskih podatkih o zdravilu ali njegovih tarčnih molekulah.<sup>8</sup>

### Iskalna orodja

DrugBank omogoča standardne besedilne poizvedbe, preko vnosnega polja za besedilo. V okviru splošnega iskanja po besedilu je v spustnem meniju možno omejiti iskanje z izbiro ene od šestih kategorij zdravila. Rezultati poizvedb so podani v obliki standardnega tabelaričnega prikaza, pri čemer je iskana beseda na vseh mestih označena.<sup>8</sup>

Lastno iskalno orodje BLAST je algoritem za primerjavo preiskovanega zaporedja aminokislin oz. nukleotidov s podatki o zaporedjih, zbranih v zbirki DrugBank. Rezultat je kvalitativno in kvantitativno ujemanje testiranega in referenčnega zaporedja. V okvirček za poizvedbo SeqSearch je potrebno vnesti zaporedje v formatu FASTA, tj. format za prikaz zaporedij nukleotidov oz. aminokislin, pri čemer sta vsak nukleotid oz. aminokislina označena z enočrkovno kodo. Iskalno orodje SeqSearch je praktično uporabno zlasti za hitro in enostavno ugotavljanje spojin vodnic.<sup>8</sup>

Iskalnik z Boolovimi operatorji uporablja indekse, ki jih izdelava program GLIMPSE (angl. *GLobal IMPLICIT Search*). Gre za program, namenjen izdelavi in obnavljanju visoko strukturiranega besedilnega kazala, ki omogoča učinkovito indeksiranje in poizvedovanje (indeks zajema le 2-4% izvornega besedila).<sup>8,9</sup>

Orodje ChemQuery je namenjeno zlasti iskanju podobnosti kemijskih struktur, uporablja pa se tudi za enostavno iskanje spojin na osnovi njihove kemijske formule oz. molekulske mase. V smislu ugotavljanja podobnosti kemijskih struktur se uporablja na podoben način kot BLAST. Uporabniki lahko v ustreznem oknu skicirajo kemijsko strukturo z uporabo ustreznih javno dostopnih javanskih programov za risanje kot sta npr. kompleksen program ACD (ChemSketch), ki ponuja veliko možnosti za risanje, ali enostavnejši program MarvinSketch (na voljo v novi različici zbirke). Kemijsko strukturo lahko v iskalni okvirček vnesemo tudi v obliki niza SMILES (angl. *Simplified Molecular Input Line Entry Specification*), podanega kot zaporedje znakov ASCII (angl. *American Standard Code for Information*

*Interchange*). Nize SMILES je možno uvoziti iz večine računalniških programov za urejanje kemijskih struktur ter jih pretvoriti v dvodimenzionalne risbe ali tridimenzionalne modele molekul. V primerjavi z InChI, ki ga je razvil IUPAC in je standardno orodje za ugotavljanje istovetnosti kemijskih formul v besedilu, za njihov prikaz ter za kodiranje molekularnih informacij, je prednost nizov SMILES v tem, da predstavljajo še bolj človeku razumljivo in berljivo obliko. Rezultati poizvedbe, tj. »zadetki« z visoko stopnjo podobnosti v kemijski strukturi, se izpišejo v tabelarni obliki, z označenimi spletnimi povezavami na ustrezno kartico o zdravilu. V novi različici je dodan tudi gumb za prikaz podobnih struktur (*Show Similar Structure(s)*). S praktičnega vidika omogoča orodje ChemQuery hitro ugotavljanje, ali se preiskovana spojina lahko (specifično) veže na predvideno tarčno molekulo.<sup>8</sup>

### Orodje za povzemanje podatkov

Orodje za oblikovanje informacij iz manj strukturiranih podatkov (*Data Extractor*) je zasnovan na enostavnem sistemu za upravljanje z relacijskimi podatkovnimi zbirkami, ki uporabniku omogoča izbiro enega ali več zelenih podatkovnih polj ter iskanje po vrstah, pojavnosti ali delni pojavnosti besed, nizov ali števil. Uporablja spletne obrazce z izbirnimi vnosnimi polji, s čimer uporabnik intuitivno sestavi poizvedbo, kot bi bila napisana v jeziku, podobnem SQL (angl. *Structured Query Language*). S samo nekaj pritiski na miško je možno relativno enostavno sestaviti kompleksne poizvedbe (npr. »najdi vse učinkovine, ki imajo molekulska masa manjša od 700 daltonov, porazdelitveni koeficient manjši od 3,5 in spadajo v farmakoterapevtsko skupino peroralnih antidiabetikov«). Rezultat je podan v formatu HTML, s spletnimi povezavami do ustreznih kartic o zdravilu.<sup>8</sup>

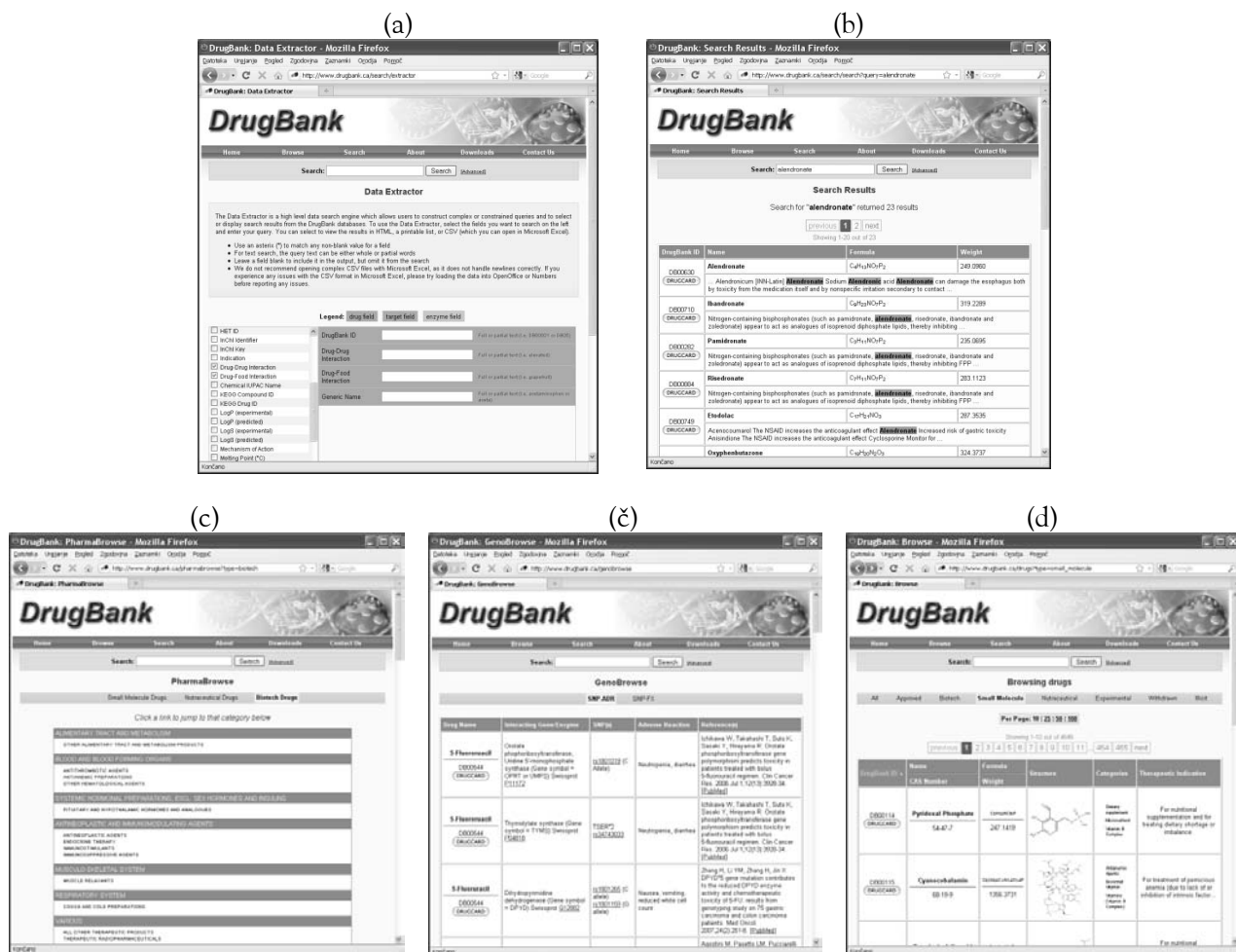
## Zagotavljanje kakovosti, posodabljanje podatkov in vnašanje popravkov

Posamezno kartico o zdravilu vnese v zbirko strokovnjak iz skupine, nato jo neodvisno preveri drug strokovnjak iz skupine, končni nadzor nad njeno vsebino pa imajo štirje po položaju najvišji strokovnjaki in raziskovalci iz skupine, in sicer zdravnik, akreditirani farmacevt ter dva biokemika z doktoratom znanosti.<sup>8</sup>

Za vnos in preverjanje podatkov v zbirki so prilagodili oz. na novo razvili več programskih paketov, vključujoč orodja za rudarjenje po besedilu, računalna za izračun kemijskih parametrov, orodja za določitev (angl. *annotation*) oz. pripis funkcije proteinom na osnovi njihovih strukturnih značilnosti. S temi orodji je možno pregledovati oz. prikazovati besedila ali slike iz različnih elektronskih podatkovnih virov ter podatke nato primerjati, oceniti, popraviti ter na koncu vnesti v DrugBank.<sup>8</sup>

Za programske pakete v okviru zbirke DrugBank se uporablja sistem za spremljanje različic programa (*CVS*, angl. *Concurrent Versioning System*), vse spremembe in popravki v centralni zbirki pa se beležijo.<sup>8</sup> Razvit je bil poseben sistem za nadzor in preverjanje pravilne izpolnjenosti vsakega polja v kartici o zdravilu.<sup>8</sup>

Sodobnost zbirke (npr. dodajanje novih spojin v bazo) zagotavljajo s pomočjo neprestano delujočih orodij za avtomatsko pridobivanje strukturiranih podatkov iz ostalih spletišč, povezanih s podatki o zdravilih, tarčnih molekulah ter strukturah spojin.<sup>5,8</sup> Ekipa strokovnjakov stalno posodablja zbirko DrugBank z novimi podatki, ki se v glavnem nanašajo na nova zdravila (dovoljenje za promet pridobi približno 50 novih zdravil na leto). Za starejša, manj uporabljena zdravila in zdravila sirote (angl. *orphan drugs*) izpolnjujejo podatkovna polja v kartici o zdravilu ročno, saj je ustrezne podatke zanje možno najti večinoma v neelektronskih podatkovnih virih.<sup>5,8</sup>



**Slika 1** Spletni vmesnik podatkovne zbirke DrugBank, ki vključuje primere rabe (a) splošnega iskanja, (b) orodja DataExtractor, (c) osnovnega prikazovalnika podatkov ter naprednejših prikazovalnikov (č) PharmaBrowse in (d) GenoBrowse.

## Izboljšave v novejših različicah

Glede na številne povratne informacije in konkretne predloge uporabnikov po zdaj že večletni uporabi zbirke so uvedli naslednje izboljšave:<sup>8</sup>

(i) razširitev prvotne različice glede števila vključenih spojin in vsebine podatkov

Velik napredek je bil narejen zlasti v smislu identifikacije velikega števila novih tarčnih molekul za zdravila, pri čemer je bilo v veliko pomoč posebno orodje za rudarjenje po besedilu

PolySearch. Tako je DrugBank danes vodilna zbirka s podatki o tarčnih molekul za zdravila.

(ii) povezava z dodatnimi spletnimi podatkovnimi zbirkami

(iii) vključitev dodatnih podatkovnih polj v kartico o zdravilu

(iv) posodobljena in naprednejša, a hkrati za uporabo enostavnejša programska orodja, zlasti za besedilne in strukturne poizvedbe ter prikaz struktur, kar omogoča hitrejšo in učinkovitejšo iskanje

Splošno iskanje po besedilu je izboljšano tako, da uporabnik za omejitev iskanja po posameznih podatkovnih poljih v kartici o zdravilu s pritiskom miške označuje ustrezna potrditvena polja. Obvezni polji pri vsakem besedilnem iskanju sta polji z imeni in sinonimi zdravila. Uporabnik ima tudi možnost, da izbere vsa besedilna podatkovna polja hkrati. Prav tako je omogočeno »inteligentno« iskanje po besedilu, ki v primeru (morebitnega) napačnega črkovanja ali nepopolnega vnosa imen zdravil in tarčnih molekul samodejno ponudi podobne alternativne besede.

- (v) izboljšana orodja za izbor ustreznih podatkov, vnos, izvažanje in označevanje podatkov ter za nadzor kakovosti podatkov in njihovo preverjanje

V ta namen so razvili posebno programsko opremo za upravljanje z znanstvenimi podatki (angl. *Scientific Data Management Software*), poimenovano DrugBank-SDMS. Različica zbirke DrugBank, ki je dostopna javnosti, je neposredno povezana z DrugBank-SDMS in vsako noč se podatki SDMS avtomatsko izvozijo na strežnik zbirke. Strokovnjaki, ki zbirko razvijajo, obnavljajo in dopolnjujejo, lahko preko tega sistema do zbirke ves čas oddaljeno dostopajo ter dodajajo podatke, preverjajo vnose, vnašajo popravke, ne da bi jim bilo potrebno pisati programe za prenos podatkov na strežnik. Oprema SDMS omogoča tudi učinkovitejšo ugotavljanje napak, že ob vnosu (avtomatsko preverjanje formata in črkovanja) ter nato enkrat tedensko (s splošnimi programi za napovedovanje in oblikovanje datotek). S SDMS je lažji tudi izvoz datotek, ki so javno na voljo za prenos, možno je hitro ustvarjanje besedilnih prenosov in lažje ustvarjanje datotek v obliki XML.<sup>8</sup>

Programerji zbirke so razvili tudi dodatna avtomatska orodja za rudarjenje po besedilu ali spletu, npr. BioSpider<sup>23</sup> ali PolySearch. BioSpider avtomatsko zbira biološke, kemijske in farmakološke podatke iz okoli 30 vsebinsko obsežnih in zanesljivih spletnih virov zgolj na osnovi imena pojine, niza STRING ali številke

CAS, medtem ko orodje PolySearch omogoča rudarjenje po izvlečkih iz iskalnega sistema PubMed. Vse tako pridobljene besedilne podatke pred vnosom v javno dostopno različico neodvisno pregledata še dva strokovnjaka s področja zdravil, od katerih ima vsaj eden doktorat iz biomedicinskih znanosti oz. je zdravnik.<sup>5,8</sup>

## Zaključek

Zbirka DrugBank ima danes izjemno široko uporabnost, ki je presegla vsa v začetku zastavljena pričakovanja. Gre za prosto dostopno namensko spletno podatkovno zbirko, ki vsebuje zelo podrobne, kakovostne in najsodobnejše podatke o zdravilih in njihovih tarčnih molekulah. Edinstvena je zlasti v smislu svoje povezovalne vloge med različnimi spletnimi (in drugimi) viri kemijskih, fizikalnih, molekularnih, farmacevtskih, bioloških, fizioloških, farmakoloških in kliničnih podatkov. Poleg tega je doslej edina javno dostopna podatkovna zbirka z informacijami o odobrenih biotehnoških zdravilih. Vgrajena ima mnoga specifična orodja za brskanje, iskanje, prikaz, razvrščanje in povzemanje podatkov o zdravilih in njihovih tarčnih molekulah; njena ključna prednost v primerjavi z drugimi spletnimi viri podatkov o zdravilih so zlasti naprednejša in za uporabo enostavna orodja za iskanje.<sup>8</sup> S tega stališča je zanimiva z bioinformacijskega vidika ter izjemno uporabna za strokovnjake na znanstvenih in strokovnih področjih, ki se ukvarjajo z zdravili.

Zbirko so razvili z namenom uporabe v raziskavah in razvoju novih zdravil, kasneje pa se je njena uporabnost široko razmahnila tudi na ostala področja farmacije ter na področja drugih naravoslovnih znanosti, povezanih z zdravili.<sup>8-12</sup> Njena uporaba se vse bolj širi tudi na področje biotehnologije, farmakogenomike in nutrigenomike.<sup>8,11</sup>

V članku smo želeli osvetliti uporabnost zbirke DrugBank. Glede na izkušnje v ameriškem prostoru je uporaba takšne zbirke v veliko pomoč strokovnjakom in znanstvenikom, ki se z različnih vidikov ukvarjajo z zdravili. Podobno zbirko o



zdravilih bi bilo smiselno razviti in uvesti (prenesti) tudi v slovenski prostor oz. na področje prometa z zdravili, ki ga nadzira Evropska agencija za zdravila, EMEA (angl. *European Medicines Agency*). Z ozaveščanjem in večanjem stopnje bioinformatijske izobraženosti strokovnjakov, ki se ukvarjajo z zdravili (predvsem s farmakološkega in kliničnega vidika), in tudi univerzitetnih učiteljev in študentov verjamemo v široko praktično uporabnost podobne zbirke o zdravilih tudi pri nas.

### Literatura

1. Rang HP, Dale MM, Ritter JM. *Pharmacology*. Churchill Livingstone, Edinburgh, London, Melbourne, New York, 2007.
2. KEGG (Kyoto Encyclopedia of Genes and Genomes): <http://www.genome.jp/kegg/> (dostop: 1.4.2010)
3. ChEBI (Chemical Entities of Biological Interest): <http://www.ebi.ac.uk/chebi/> (dostop: 1.4.2010)
4. PubChem: <http://pubchem.ncbi.nlm.nih.gov/> (dostop: 1.4.2010)
5. UniProt/Swiss-Prot: <http://www.ebi.ac.uk/uniprot/> (dostop: 1.4.2010)
6. PharmGKB (Pharmacogenomics Knowledge Base): <http://www.pharmgkb.org/> (dostop: 1.4.2010)
7. RxList: <http://www.rxlist.com/script/main/hp.asp> (dostop: 1.4.2010)
8. DrugBank: <http://www.drugbank.ca/> (dostop: 1.4.2010)
9. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006; 34 (Database issue): D668-672.
10. Wishart DS. In silico drug exploration and discovery using DrugBank. *Curr Protoc Bioinformatics* 2007; Chapter 14: Unit 14.4.
11. Wishart DS. DrugBank and its relevance to pharmacogenomics. *Pharmacogenomics* 2008; 9: 1155-1162.
12. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 2008; 36 (Database issue): D901-906.
13. PubMed/Medline: <http://www.ncbi.nlm.nih.gov/PubMed> (dostop: 1.4.2010)
14. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 1998; 14: 656-664.
15. GeneCards: <http://www.genecards.org/> (dostop: 1.4.2010)
16. PDB (Protein Data Bank): <http://www.pdb.org> (dostop: 1.4.2010)
17. GenBank: <http://www.ncbi.nlm.nih.gov/genbank/> (dostop: 1.4.2010)
18. Wikipedia: <http://en.wikipedia.org/> (dostop: 1.4.2010)
19. PDRHealth (Physicians' Desk Reference Health): <http://www.pdrhealth.com/> (dostop: 1.4.2010)
20. DPD (Health Canada Drug Product Database): <http://www.hc-sc.gc.ca/dhp-mps/prodpharma/databasdon/index-eng.php> (dostop: 1.4.2010)
21. HGNC (HUGO Gene Nomenclature Commission): <http://www.genenames.org/> (dostop: 1.4.2010)
22. GenAtlas: <http://genatlas.medecine.univ-paris5.fr/> (dostop: 1.4.2010)
23. Knox C, Shrivastava S, Stothard P, Eisner R, Wishart DS. BioSpider: a web server for automating metabolome annotations. *Pac Symp Biocomput* 2007:145-56.