

Współczesne strategie przeszukiwania baz struktur i związków chemicznych

Grzegorz FIC, Mariusz SKOMRA, Barbara DĘBSKA - Wydział Chemiczny, Politechnika Rzeszowska

Prosimy cytować jako: CHEMIK 2016, 70, 8, 410–418

Wprowadzenie

Współczesne bazy struktur i związków chemicznych cechują się bardzo dużym przyrostem informacji. W Tabelicy I pokazano przyrost zasobów bazy Pubchem w ostatnich latach. Baza ta powstała w 2004r. i jest zarządzana przez National Center for Biotechnology Information (NCBI), instytucję podległą US National Institutes of Health (NIH). Jest to aktualnie najliczniejszy i dostępny bezpłatnie zbiór struktur chemicznych na świecie. Składa się z trzech baz, w których gromadzone są informacje o związkach małowcząsteczkowych (nie więcej niż 1000 atomów oraz wiązań): PubChem Substance – zawiera pochodzące z wielu innych baz informacje o substancjach (mieszaniny, ekstrakty, związki kompleksowe i in.), PubChem Compound – zawiera informacje o strukturach chemicznych zapisanych w PubChem Substance, oraz PubChem BioAssay – zawiera informacje o różnych projektach badawczych w zakresie testowania ukierunkowanej biologicznej aktywności (w poszczególnych projektach liczebność grupy testowanych związków od 1 do kilkuset tysięcy).

Większość dużych baz chemicznych jest kompilacją zasobów pochodzących z wielu innych baz. Np. baza PubChem Substance gromadzi informacje z prawie 400 baz źródłowych – jedne z nich udostępniają miliony rekordów informacji (np. baza Aurora Fine Chemicals LLC – ponad 33 mln substancji, ZINC – 25,7 mln rekordów), inne zaś (np. laboratoria czy też niewielkie zespoły badawcze) niewielką liczbę, a nawet jeden rekord informacji. Warta odnotowania jest także kompatybilność struktur danych, co umożliwia powstawanie dużych systemów wyszukiwawczych. Przykładem może tutaj być system Entrez [1], udostępniający zasoby ok. 30 baz chemicznych, biologicznych i pokrewnych. Wszystkie bazy mogą być przeszukiwane poprzez sformułowanie i przesłanie jednej kwerendy wyszukiwawczej, z wykorzystaniem interfejsu globalnej wyszukiwarki Entrez.

Tablica I

Przyrost zasobów bazy Pubchem w latach 2007-2016 (mln).
Wartości uzyskano przez zastosowanie kwerendy
wyszukiwawczej: all[fil]

	29.10.07	16.09.08	11.11.09	19.02.10	18.03.13	6.11.14	9.02.16
PubChem Substance	21.3	44.6	61.2	62.5	116.8	178.4	216.9
PubChem Compound	18.1	19.3	25.7	26.1	46.7	62.0	87.2
PubChem BioAssay	0.000667	0.001188	0.001917	0.002153	0.649147	1.112105	1.154427

Przytoczone cechy baz związków chemicznych, a przede wszystkim duża ilość zgromadzonych w nich informacji, implikują konieczność stosowania wyrafinowanych narzędzi do wyszukiwania i odzyskiwania informacji. W dalszej części artykułu przedyskutowano wybrane

zagadnienia związane z formułowaniem kwerend wyszukiwawczych (tj. zestawem warunków, jakie musi spełniać wyszukana informacja) oraz strategiami przeszukiwania internetowych baz struktur i związków chemicznych. Szczególną uwagę zwrócono na najnowsze rozwiązania w tym zakresie, np. na strategię bazującą na koncepcjach rozmytych podstruktur, 2D i 3D podobieństwa chemicznego, ontologiczne oraz podobieństwa typu struktura-właściwości.

Strategie przeszukiwania baz związków chemicznych

Najogólniej, stosowane aktualnie strategie przeszukiwania baz związków chemicznych można podzielić na:

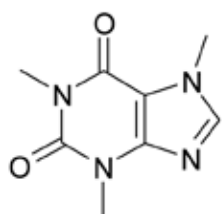
- 1. tekstowo-liczbowe**, wykorzystują pola typu: nazwa, autor, producent, właściwości, i inne
- 2. strukturalne**, przeszukują pola zawierające informację o strukturze, np. kody SMILES oraz MOL. Można tutaj wskazać różne metody, w tym (podano także angielskie nazwy strategii):
 - wyszukiwanie wszystkich struktur tożsamyh ze zdefiniowaną w kwerendzie (*Exact*), w tym także tautomerów (*All Tautomers* w bazie ChemSpider)
 - wyszukiwanie wszystkich struktur zawierających podstrukturę zdefiniowaną w kwerendzie (*Substructure*)
 - wyszukiwanie wszystkich struktur będących podstrukturami struktury zdefiniowanej w kwerendzie (*Superstructure*) – strategia przydatna podczas komputerowo wspomaganego projektowania syntez organicznych (poszukiwanie *building blocks*)
 - wyszukiwanie struktur podobnych, tj. spełniających zdefiniowane przez użytkownika warunki podobieństwa strukturalnego. W połączeniu ze strategią *Substructure* pozwala na tworzenie zbiorów struktur do badań przesiewowych (np. w procesie projektowania nowych leków)
 - wyszukiwanie prekursorów, tj. struktur, z których można otrzymać zdefiniowaną strukturę, np. strategia *Precursors* w bazie przemian metabolicznych BioPath
 - wyszukiwanie wszystkich struktur zawierających zdefiniowany szkielet strukturalny, np. strategię: *Flex*, *Flexplus* w bazie ChemIDplus, *Same Skeleton* w ChemSpider
 - wyszukiwanie wszystkich izomerów strukturalnych zdefiniowanej struktury, np. *All Isomers* w bazie ChemSpider.
- 3. hybrydowe**, kwerenda złożona jest z wielu warunków różnego typu (strukturalnych, tekstowych, liczbowych i logicznych), określających wymagane lub/ oraz zabronione właściwości wyszukiwanych związków chemicznych
- 4. ontologiczne**, stosowane przede wszystkim w bazach chemiczno-biologicznych, umożliwiają przeszukiwanie i przeglądanie treści baz wg terminów charakteryzujących poszczególne substancje i wymagają zaimplementowania jakiejś ontologii
- 5. typu struktura-właściwości** – wyszukiwanie związków posiadających podobne właściwości jak wskazana struktura/podstruktura.

Metody wprowadzania informacji strukturalnej

Informacje o strukturze chemicznej, będące istotnym elementem kwerend wyszukiwawczych, mogą być wprowadzane na kilka sposobów, jako:

Autor do korespondencji:
Dr inż. Grzegorz FIC, e-mail: gfic@prz.edu.pl

- **tekstowe identyfikatory struktury** – łańcuchy tekstowe, które mogą reprezentować struktury chemiczne w kwerendach wyszukiwawczych, np. nazwa systematyczna (IUPAC), nazwy zwyczajowe, numery identyfikacyjne (np. *CAS Registry Number*, *EINECS Number*), liniowy kod struktury SMILES [2], wektory InChI (*IUPAC International Chemical Identifier*) [3] oraz InChIKey (shaszowana postać wektora InChI) [4]. Wektor InChI składa się z kilku warstw informacji, oddzielonych znakiem „/” i prefiksem (Rys. 1): IS oznacza wersję InChI, następnie wzór sumaryczny, /c to początek listy połączeń atomów, /h – subwarstwa opisująca atomy wodoru połączone z innymi atomami (mogą jeszcze wystąpić inne warstwy, np. stereochemiczna, ładunków, izotopowa). InChIKey (Rys. 1) to wektor o stałej długości 27 znaków, zawierający informacje nieczytelne dla człowieka. Jest jednoznaczna reprezentacją struktury, chociaż znane są odstępstwa od tej reguły [5]. Powstaje w wyniku shaszowania wg algorytmu SHA-256 informacji zawartej w InChI. Składa się z 3 bloków, rozdzielanych znakiem pauzy. Pierwszy blok jest charakterystyczny dla związków zawierających określony szkielet strukturalny, drugi blok różnicuje izomery. Jest to wykorzystywane podczas przeszukiwania baz struktur. Kod SMILES jest następstwem linearyzacji struktury chemicznej, tj. „przecięcia” jednego wiązania w każdym pierścieniu, zaś początek oraz zakończenie tak utworzonego „łańcucha” reprezentowane są przez parę tożsamyh cyfr po symbolach atomów (na Rys. 1 są to atomy: N1 i C1 oraz C2 i N2).
- **graficzna informacja strukturalna** – narysowanie wzoru struktury/podstruktury i wysłanie jego kodowej postaci (np. jako kod SMILES) do serwera bazy (wyszukiwarki) przy wykorzystaniu specjalnego programu, który może mieć postać wtyczki do przeglądarki internetowej lub apletu przesyłanego przez serwer do komputera użytkownika
- **plikowa informacja strukturalna** – wczytanie pliku zawierającego informację o strukturze w jednym z formatów zapisu informacji o strukturach (np. mol, cml, smi, pdb). Plik może być zapisany na lokalnym komputerze lub w Internecie
- **plikowa informacja graficzna** – wczytanie pliku zawierającego obraz struktury (np. odręczny rysunek, skan ilustracji z publikacji) zapisany w jednym z akceptowanych przez wyszukiwarkę formatów graficznych (np. gif, jpg, tif, pdf); technologia aktualnie dostępna w wyszukiwarkach ChemSpider.com oraz Chemical Structure Lookup Service 2008.



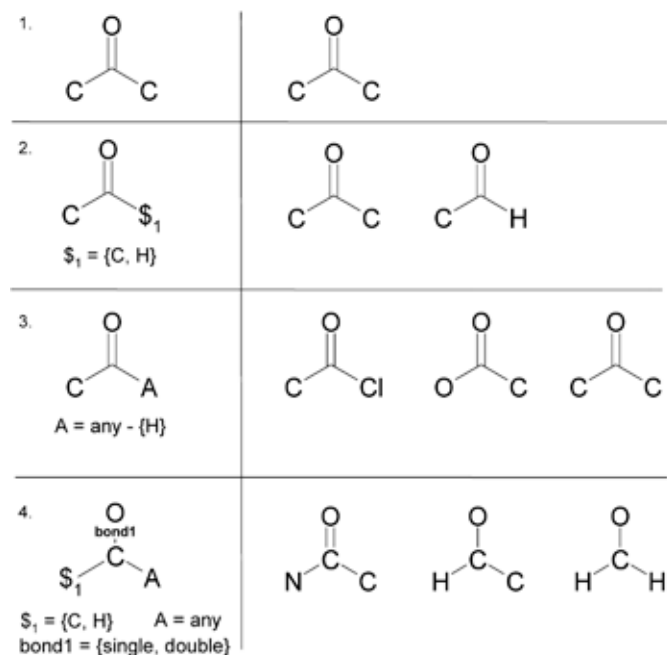
1,3,7-trimethylpurine-2,6-dione
 InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
 RYYVLZVUVIJVGH-UHFFFAOYSA-N
 CN1C=NC2=C1C(=O)N(C(=O)N2C)C
 58-08-2
 200-362-1

Rys. 1. Wybrane tekstowe identyfikatory kofeiny. Kolejno, od góry: nazwa IUPAC, InChI, InChIKey, kod SMILES, CAS Registry Number, EINECS Number

Rozmyte podstruktury

Ważnym elementem wielu kwerend wyszukiwawczych jest fragment struktury chemicznej – podstruktura. Podstruktury mogą być definiowane w sposób ostry (tzn. każdy atrybut opisujący podstrukturę ma dokładnie jedną wartość) lub rozmyte (Rys. 2). Prekursorem rozmytych podstruktur są struktury Markusha – struktury chemiczne zawierające zmienne fragmenty. Koncepcja tych struktur została wprowadzona w celu zwartej reprezentacji klas lub zbiorów związków o podobnych strukturach. W rozmytych

podstrukturach [6] atrybuty określające właściwości węzłów-atomów (np. rodzaj atomu, liczba sąsiadów, liczba wolnych elektronów, położenie atomu: w pierścieniu, układzie aromatycznym,



Rys. 2. Przykłady rozmytych podstruktur. Po lewej definicja podstruktury rozmytej, po prawej przykłady ze zbioru podstruktur spełniających tę definicję (poz. 1: podstruktura zdefiniowana ostro, bez rozmytości)

łańcuchu alifatycznym, i inne) lub/oraz krawędzi-wiązań (np. typ wiązania, położenie przestrzenne) mogą przyjmować więcej niż jedną wartość – dla każdego atrybutu określa się zbiory dozwolonych lub zakazanych wartości. Każda rozmyta podstruktura reprezentuje zatem nie jedną, ale zbiór konkretnych podstruktur. Informacje o rozmytości podstruktur wprowadza się w interfejsach wyszukiwarek z wykorzystaniem specjalnych formularzy lub za pomocą dedykowanej notacji SMARTS (Rys. 3) [7], należącej do rodziny notacji SMILES.

<chem>[n;R]&*C(=O)O</chem>	wyszukiwanie wszystkich związków zawierających grupę karboksylową oraz atom azotu leżący w jakimś pierścieniu aromatycznym
<chem>[OH]c1ccccc1</chem>	wyszukiwanie wszystkich związków zawierających pierścień benzenowy z przyłączoną grupą OH
<chem>(C(=O)O).(OCC)>>C(=O)OCC.O</chem>	wyszukiwanie reakcji międzycząsteczkowej estryfikacji
<chem>(C(=O)O.OCC)>>C(=O)OCC.O</chem>	wyszukiwanie reakcji wewnątrzcząsteczkowej estryfikacji

Rys. 3. Przykłady kwerend w notacji SMARTS (rozszerzenie notacji SMILES umożliwiające formułowanie strukturalnych kwerend wyszukiwawczych z uwzględnieniem koncepcji rozmytych podstruktur)

Podobieństwo struktur chemicznych

Opracowano różne metody komputerowego określania podobieństwa dwóch struktur chemicznych. Każda z nich wymaga zastosowania odpowiedniej metody reprezentacji struktury oraz zdefiniowania miary podobieństwa chemicznego (strukturalnego).

Metody reprezentacji struktur. Uwzględniają tylko budowę konstytucyjną (podobieństwo chemiczne 2D) lub pełną przestrzenną strukturę (podobieństwo chemiczne 3D). Można wskazać trzy główne typy 2D reprezentacji struktur stosowane przy określaniu ich podobieństwa chemicznego:

- klucze struktury (*fingerprints*) – zero-jedynkowe wektory o różnej długości; każdy element potwierdza (wartość 1) bądź też wyklucza (wartość 0) obecność jakiejś cechy strukturalnej w cząsteczce. Opracowano wiele różnych kluczy struktury, które różnią się: (i) liczbą bitów, (ii) znaczeniem poszczególnych bitów oraz (iii) uwzględnianiem wybranych elementów opisujących 3D strukturę. Najczęściej stosowane to: PubChem fingerprint (881 cech struktur) [8], FP3 oraz FP4 (pakiet OpenBabel, uwzględniają 55/307 typów podstruktur) [9], MACCS [10] – występuje w kilku postaciach: 166/322-bitowy fingerprint oraz 166/322-elementowy liczbowy deskryptor (elementy wektora są tutaj liczbami określającymi ile konkretnych cech strukturalnych znajduje się w strukturze cząsteczki). Poważną wadą tego typu fingerprint jest niemożność uwzględnienia wszystkich cech strukturalnych cząsteczek – wynik porównania nie jest w pełni obiektywny. Wyznaczony stopień podobieństwa zależy przede wszystkim od zastosowanego klucza struktury – może się okazać, że dwie struktury, bardzo podobne wg metody bazującej na jakimś kluczu, okażą się dużo mniej podobne po zastosowaniu innego fingerprint. Próba eliminacji tej wady są haszowane klucze struktury. Tutaj zbiór cech strukturalnych nie jest uprzednio ustalony. W wyniku totalnej fragmentacji struktury generowany jest zbiór wszystkich podstruktur: od jedno-, dwu-, ..., aż do n-atomowych (n jest liczbą atomów w cząsteczce). Następnie, po zastosowaniu algorytmów haszujących, generowany jest zero-jedynkowy fingerprint o ustalonej długości (najczęściej 1024 lub 512 bitów). Niezależnie od tego, ile podstruktur powstało w wyniku fragmentacji wygenerowany fingerprint ma zawsze taką samą liczbę elementów. W wyniku haszowania jakiś bit może reprezentować więcej niż jeden element struktury (tzw. kolizja bitów); taka sytuacja jest dozwolona, aczkolwiek niewskazana. Najczęściej stosowane haszowane klucze struktury: FP2 (pakiet OpenBabel) – fragmentacja struktury jest ograniczona do fragmentów 1–7 atomowych, ECFP [11] – grupa szaszowanych fingerprint nowej generacji, początkowy zestaw cech jest następnie rozszerzany i uwzględnia, po kolejnych iteracjach, wpływ sąsiadów kolejnych sfer (mierzonych liczbą wiązań od stanu początkowego).
- deskryptory molekularne [12] – wartości liczbowe, z których każda reprezentuje ilościowo jedną lub więcej konkretnych cech strukturalnych cząsteczki
- molekularne wartości własne – jedna wartość liczbowa reprezentuje strukturę całej cząsteczki, np. deskryptory BCUT [13].

Miary podobieństwa. Określają w sposób ilościowy podobieństwo chemiczne dwóch struktur; niektóre z nich, stosowane są dla określenia podobieństwa z użyciem *fingerprint* (Rys. 4). Podobieństwo Tanimoto (TS): a jest liczbą par bitów [1,0] (tj. sytuacji, gdy jakaś cecha strukturalna reprezentowana w *fingerprint* występuje w jednej, a nie występuje w drugiej strukturze), b – liczbą par bitów [0,1], c – liczbą par bitów [1,1]. Należy tutaj wyraźnie podkreślić, że wartość $TS=1$ nie świadczy o identyczności dwóch porównywanych struktur, a jedynie o tożsamości ich wektorów *fingerprint*. Przyjmuje się, że wartości TS powyżej 0.85 świadczą o dużym podobieństwie dwóch struktur chemicznych i jest wielce prawdopodobne, że te dwie struktury będą np. wykazywać podobną aktywność biologiczną. Modyfikacją TS jest asymetryczny indeks Tversky'ego (TI): α oraz β są tutaj współczynnikami wagowymi, zezwalającymi na uwzględnienie czynnika asymetrii. Inna miara podobieństwa, to odległość euklidesowa (ED): a_i oraz b_i są wartościami i -tych elementów *fingerprint* dla dwóch porównywanych struktur A i B. Im większa wartość ED tym mniejsze podobieństwo struktur. ED jest stosowany przede wszystkim jako miara w metodach porównywania podobieństwa chemicznego bazujących na deskryptorach. Znane są również inne miary podobieństwa chemicznego [14].

Podobieństwo chemiczne 3D. Warunkiem wyznaczenia podobieństwa chemicznego 3D jest znajomość położenia poszczególnych atomów w przestrzeni 3D dla różnych konformacji porównywanych cząsteczek. Spośród coraz większej liczby metod wyznaczania podobieństwa 3D warto wymienić:

- zastosowanie deskryptorów związanych z odległościami i kątami w przestrzeni 3D (np. kąty wiązań, odległości międzyatomowe)
- określanie podobieństwa pól molekularnych (pola potencjału elektrostatycznego, kształtu geometrycznego, gęstości elektronowej i inne), np. metody CoMFA (*Comparative Molecular Field Analysis*), CoMSIA (*Comparative Molecular Similarity Index Analysis*)
- porównanie momentów molekularnych, np. CoMMA (*Comparative Molecular Moments Analysis*) – uwzględnienie deskryptorów będących momentami molekularnymi względem centrum masy, centrum ładunku, centrum dipola
- zastosowanie deskryptorów związanych z kształtami cząsteczek, np. objętość van der Waals, parametr steryczny Tafta, parametry STERIMOL umożliwiające ilościowy opis kształtów podstawników, deskryptory WHIM – *Weighted Holistic Invariant Molecular*.

$$TS = \frac{c}{a+b+c} \quad \Pi = \frac{c}{\alpha \cdot a + \beta \cdot b + c}$$

$$ST = \frac{V_{AB}}{V_{AA} + V_{BB} - V_{AB}} \quad CT = \frac{\sum V_{AB}}{\sum V_{AA} + \sum V_{BB} - \sum V_{AB}}$$

$$ED = \sqrt{\sum (a_i - b_i)^2}$$

$$T_3S = \frac{\sum_{i=1}^N X_{iA} X_{iB}}{\sum_{i=1}^N X_{iA}^2 + \sum_{i=1}^N X_{iB}^2 - \sum_{i=1}^N X_{iA} X_{iB}}$$

Rys. 4. Definicje wybranych miar podobieństwa chemicznego (objaśnienia w tekście)

Miary podobieństwa 3D konformerów. Do ilościowej oceny podobieństwa chemicznego 3D opracowano specjalne miary [15]. Jedną z nich jest podobieństwo Tanimoto, w postaci zmodyfikowanej dla pól molekularnych (T_3S na Rys. 4): X_{iA} oraz X_{iB} to wartości liczbowe jakiejś właściwości porównywanych cząsteczek A oraz B w i -tym punkcie siatki pola. W bazie PubChem stosowane są dwie miary (Rys. 4) [16]. Parametr ST (*shape Tanimoto*) jest miarą podobieństwa 3D kształtu: V_{AA} , V_{BB} to objętości tych fragmentów konformerów odpowiednio A i B, które nie są wspólne w ich maksymalnej superpozycji, V_{AB} – wspólna objętość cząsteczek A i B w tej superpozycji. Natomiast CT (*color Tanimoto*) jest miarą zgodności sześciu cech strukturalnych: donory oraz akceptory wiązań wodorowych, kationy, aniony, hydrofobowość, pierścienie. Sumowanie jest po cechach: V_{AB} – objętość fragmentów zgodnych dla jakiejś cechy, V_{AA} oraz V_{BB} – objętości fragmentów w których konkretne cechy się nie pokrywają. Procedura wyznaczenia ST i CT jest dwuetapowa: (i) utworzenie takiej superpozycji (nałożenia) 3D struktur dwóch konformerów, w której wspólna ich część (V_{AB}) jest maksymalna i obliczenie ST, oraz (ii) obliczenie CT w poszczególnych punktach tej superpozycji. W algorytmie PubChem przyjmuje się, że jakieś dwa konformery są podobne, gdy $ST \geq 0.8$ oraz $CT \geq 0.5$.

Strategie struktura-właściwości

Oprócz metod, które wyznaczają podobieństwo chemiczne w wyniku analizy i porównania struktur chemicznych, zaproponowano metody, w których kryteriami porównawczymi są właściwości cząsteczek chemicznych. Wg tej koncepcji podobnymi mogą okazać się cząsteczki, których struktury są zupełnie różne i których

podobieństwo wyznaczane wg klasycznych metod jest niewielkie. Ma to szczególne znaczenie dla tworzenia zbiorów struktur cechujących się z jednej strony dużą różnorodnością strukturalną (*chemical diversity*), a z drugiej strony zbliżonymi właściwościami, np. podobną aktywnością biologiczną. Przykładem takiej metody jest LASSO (*Ligand Activity in Surface Similarity Order*) [17]: dwa ligandy wykazują zbliżoną aktywność, jeżeli ich powierzchnie mają podobne właściwości – związki o podobnych powierzchniach wiążą się z tymi samymi białkami. Dla każdej cząsteczki w bazie danych wyznacza się tzw. deskryptor LASSO – wektor, którego każdy element jest liczbą punktów na powierzchni cząsteczki (reprezentowanej w postaci siatki) cechujących się określonymi właściwościami. Wyróżnia się 23 różne typy punktów powierzchni, np. miejsca hydrofobowe, miejsca oddziaływań typu π , miejsca obsadzone przez donory wiązań wodorowych i inne. Przyjęto założenie, że dwie cząsteczki wykazują podobną aktywność, jeżeli ich deskryptory LASSO są identyczne lub bardzo podobne. Wykorzystując metodę sieci neuronowych można dokonać szybkiej selekcji związków, które mogą wykazywać określoną aktywność biologiczną (autorzy podają, że przeskanowanie 1 mln związków trwa poniżej 1 minuty). Metoda LASSO została zaimplementowana m.in. w wyszukiwarce chemicznej ChemSpider.com.

Strategie ontologiczne

W wyniku dynamicznego rozwoju różnych dziedzin wiedzy doszło do sytuacji, że naukowcy z całego świata, odkrywając nowe fakty, tworzyli często swoje własne nazewnictwo do opisu poznawanej wiedzy. W efekcie, naukowcy z różnych ośrodków na świecie mieli problemy z porównaniem i analizą wyników badań. Próba rozwiązania tego problemu jest tworzenie ontologii dla konkretnych dziedzin wiedzy. Ontologia to zbiór terminów (termów) opisujących jakąś dziedzinę wiedzy. Jej głównym zadaniem jest zapewnienie jednoznaczności przekazu wiedzy. W tym celu ontologia wykorzystuje kategoryzację oraz hierarchizację. Jako przykład takiego podejścia można wskazać ontologię ChEBI (*Chemical Entities of Biological Interest*) [18], będącą częścią bazy związków bioaktywnych ChEMBL. Ontologia ChEBI (Rys. 5) wykorzystuje trzy subontologie: (i) jednostki chemiczne (*chemical entity*), w której związki są klasyfikowane ze względu na ich skład i strukturę,

CHEBI:24431 chemical entity – CHEBI:23367 molecular entity – CHEBI:33579 main group molecular entity – CHEBI:33675 p-block molecular entity – CHEBI:33582 carbon group molecular entity – CHEBI:50860 organic molecular entity – CHEBI:33285 heteroorganic entity – CHEBI:36962 organochalcogen compound – CHEBI:36963 organooxygen compound – CHEBI:37622 carboxamide – CHEBI:29347 monocarboxylic acid amide – CHEBI:22645 arenecarboxamide – CHEBI:22702 benzamides – CHEBI:7496 nelfinawir

Rys. 5. Jedna z gałęzi drzewa ontologii ChEBI dla nelfinawiru – inhibitora proteazy HIV. Pełne drzewo zawiera ok. 50 gałęzi: <http://www.ebi.ac.uk/chebi/chebiOntology.do?chebild=CHEBI:7496&treeView=true#vizualisation> (17.02.2016)

np. węglowodory, kwasy karboksylowe, aminy, trzeciorzędowe aminy, (ii) funkcję (*role*), w której jednostki są klasyfikowane ze względu na ich biologiczne funkcje (np. antybiotyki, koenzymy, hormony) albo ze względu na sposób ich wykorzystania przez ludzi (np. pestycydy, leki antyreumatyczne, paliwa), albo ze względu na funkcje chemiczne (np. akceptor, donor, rozpuszczalnik, ligand), oraz (iii) jednostki subatomowe (*subatomic particle*), która klasyfikuje jednostki mniejsze od atomów (np. elektrony, fotony, nukleony).

Podsumowanie

Niektóre z opisanych tutaj strategii wyszukiwania informacji strukturalnych są coraz powszechniej stosowane przez chemików (np. strategie bazujące na podobieństwie chemicznym czy koncep-

cji rozmytych podstruktur), inne mają tylko specyficzne zastosowania (np. strategie *Superstructure* oraz *Precursors*, stosowane podczas komputerowo wspomaganego projektowania reakcji i syntez organicznych [19]).

Wydaje się, że aktualnie jednym z najważniejszych zastosowań różnych narzędzi do przeszukiwania baz związków chemicznych jest tworzenie wirtualnych bibliotek struktur i związków, niezbędnych dla wirtualnego screeningu, który stał się nieodłącznym elementem procesów projektowania związków chemicznych o pożądanym właściwościach (np. nowych leków). Jest to następstwem rozwoju metod chemoinformacyjnych, a przede wszystkim algorytmów określania podobieństwa chemicznego i strategii przeszukiwania dużych baz struktur chemicznych. Bardzo pomocnym narzędziem wspomagającym tego typu prace jest klasteryzacja, tj. grupowanie struktur we względnie jednorodne klasy (klastry), zawierające struktury podobne. Klasteryzacja zbiorów struktur może być dokonywana z wykorzystaniem podobieństwa 2D lub 3D (PubChem udostępnia obie metody).

Warto zwrócić uwagę na strategię ontologiczną, umożliwiającą szybkie wyszukanie związków o podobnych właściwościach chemicznych czy biologicznych. Jest to jeszcze narzędzie mało rozpowszechnione i zaimplementowane w niewielu wyszukiwarkach. Wyróżnia się tutaj baza Pubchem, udostępniająca kilka różnych ontologii, w tym wspomnianą uprzednio ChEBI.

Duże nadzieje należy wiązać z dalszym rozwojem strategii typu: struktura-właściwości. Już obecnie strategia tego typu umożliwia znalezienie w bazie ChemSpider.com związków cechujących się dużą bioaktywnością do konkretnego białka i niewielką do jakiegokolwiek innego białka.

Literatura

- Entrez, the Life Sciences Search Engine; <http://www.ncbi.nlm.nih.gov/sites/gquery>.
- SMILES – A Simplified Chemical Language, Daylight Chemical Information Systems, Inc. 2011, <http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>.
- The IUPAC International Chemical Identifier, IUPAC, 2012. <http://www.iupac.org/home/publications/e-resources/inchi.html>; Heller S., McNaught A., Stein S., Tchekhovskoi D., Pletnev I.: *InChI – the worldwide chemical structure identifier standard*. J. Cheminform. 2013, 5, 7. <http://jcheminf.springeropen.com/articles/10.1186/1758-2946-5-7>.
- Heller S.R., McNaught A., Pletnev I., Stein S., Tchekhovskoi D. *InChI, the IUPAC International Chemical Identifier*. J. Cheminform. 2015, 7, 23. <http://jcheminf.springeropen.com/articles/10.1186/s13321-015-0068-4>.
- Williams A. *An InChIkey Collision is Discovered and NOT Based on Stereochemistry*, 2011, <http://www.chemconnector.com/2011/09/01/an-inchikey-collision-is-discovered-and-not-based-on-stereochemistry/>.
- Hippe Z.S., Fic G., Nowak G.: *Representation of Common Sense in Chemical Syntheses by Means of Molecular Graphs*, Found. Comput. Decision Sci. 1994, 19, 21–30; Dębska B., Guzowska-Świder B.: *Fuzzy Definition of Molecular Fragments in Chemical Structures*. J. Chem. Inf. Comput. Sci., 2000, 40, 325–329.
- SMARTS – A Language for Describing Molecular Patterns, Daylight Chemical Information Systems, Inc. 2011, <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>.
- dokładna specyfikacja: ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.pdf.
- Open Babel v2.3.0 documentation, <http://openbabel.org/docs/dev/>.
- MayaChemTools: MACCS, 2015, <http://www.mayachemtools.org/docs/modules/html/MACCSKeys.html>.
- Rogers D., Hahn M.: *Extended-Connectivity Fingerprints*. J. Chem. Inf. Model. 2010, 50(5), 742–754.
- Todeschini R., Consonni V.: *Molecular Descriptors for Chemoinformatics*. Wiley-VCH, 2009.
- Pearlman R.S., Smith K.M.: *Metric Validation and the Receptor-Relevant Subspace Concept*. J. Chem. Inf. Comput. Sci. 1999, 39 (1), 28–35.
- www.daylight.com/dayhtml/doc/theory/theory.finger.html.
- Todeschini R., Consonni V.: *Handbook of Molecular Descriptors*. John Wiley & Sons, 2008.

16. Bolton E.E., Kim S., Bryant S.H.: *PubChem3D: Similar conformers*. J. Cheminform. 2011, 3, 13, <http://www.jcheminf.com/content/3/1/13>.
17. Reid D., Sadjad B.S., Zsoldos Z., Simon A.: *LASSO-ligand activity by surface similarity order: a new tool for ligand based virtual screening*. J. Comput. Aided Mol. Des. 2008, 22(6-7), 479-87.
18. Hastings J., de Matos P., Dekker A., Ennis M., Harsha B., Kale N., Muthukrishnan V., Owen G., Turner S., Williams M., Steinbeck C.: *The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013*, Nucleic Acids Res. 2013, 41(Database issue):D456-63, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531142/>.
19. Nowak G., Fic G.: *Komputerowe projektowanie wieloetapowych syntez organicznych – stan aktualny i kierunki rozwoju*. Przem. Chem. 2011, 90 (1), 78-83.

(Otrzymano 14.07.2016 r.)

Mgr inż. Mariusz SKOMRA – absolwent Politechniki Rzeszowskiej (biotechnologia). Doktorant Wydziału Chemicznego Politechniki Rzeszowskiej. Zainteresowania naukowe – komputerowe projektowanie reakcji i syntez organicznych, komputerowe przetwarzanie informacji strukturalnych.
e-mail: d180@stud.prz.edu.pl, tel.: 17 865 1838

Dr inż. Grzegorz FIC – absolwent Politechniki Rzeszowskiej (technologia chemiczna); dr nauk chemicznych (UJ). Starszy wykładowca w Politechnice Rzeszowskiej (Wydział Chemiczny, Zakład Biotechnologii i Bioinformatyki). Autor 77 publikacji naukowych, systemów informacyjnych oraz programów komputerowych dla chemii i nauk pokrewnych. Zainteresowania naukowe – komputerowe projektowanie reakcji i syntez organicznych, komputerowe przetwarzanie informacji strukturalnych, bioinformatyka.
e-mail: gfic@prz.edu.pl, tel.: 17 865 1838

Dr hab. inż. Barbara DĘBSKA – absolwentka WSI w Rzeszowie; dr hab. nauk technicznych w zakresie technologii chemicznej i informatyki chemicznej. Zatrudniona w Politechnice Rzeszowskiej jako profesor ndzw., Zakład Biotechnologii i Bioinformatyki. Odznaczona Krzyżem Kawalerskim Orderu Odrodzenia Polski. Autorka 187 publikacji naukowych. Zainteresowania naukowe – chemometria, inteligentne systemy informatyczne dla nauk chemicznych i pokrewnych, e-learning.
e-mail: bjdebska@prz.edu.pl, tel.: 17 865 1358

Aktualności z firm

News from the Companies

ZMIANY PERSONALNE

Zmiany w składzie Zarządu BIOMED-LUBLIN

29 lipca br. Pan Waldemar Sierocki złożył rezygnację z funkcji Prezesa Zarządu BIOMED-LUBLIN WSiS SA. Stanowisko Prezesa Zarządu zostało powierzone Panu Andrzejowi Stachnikowi, menedżerowi o wieloletnim doświadczeniu w branży farmaceutycznej. Rada Nadzorcza Spółki oraz Zarząd w komunikacie giełdowym podkreśliły wkład Pana Sierockiego w wieloletni rozwój Spółki oraz w proces zatwierdzenia układu przez zgromadzenie wierzycieli w postępowaniu restrukturyzacyjnym. (kk)

(<http://www.biomed.lublin.pl/>, 31.07.2016)

Zmiany w składzie Rady Dyrektorów CEEP

Robert Pietryszyn, prezes Zarządu Grupy LOTOS SA został wybrany do składu Rady Dyrektorów Central Europe Energy Partners (CEEP). CEEP jest jedyną organizacją branżową sektora energii w Europie Centralnej w Brukseli i odgrywa kluczową rolę w inicjowaniu oraz opiniowaniu dyrektyw unijnych. Działa na rzecz wspierania integracji środkowoeuropejskiego sektora energii w ramach wspólnej polityki energii oraz bezpieczeństwa w sektorze energii UE. Prowadzi także szczegółowe prace analityczne, stając się wyrazicielem poglądów swoich członków.

Aktualnie w skład organizacji wchodzi podmioty sektora energii z Czech, Rumunii, Chorwacji, Litwy i Polski. Są to między innymi: KGHM, Litgrid, Achema, Grupa Azoty, ArcelorMittal Poland, HEP, NWR, PSE czy PERN. Walne Zgromadzenie CEEP wybrało nową Radę Dyrektorów, w skład której, obok Roberta Pietryszyna, weszli: Henryk Baranowski, Prezes Zarządu PGE (do akceptacji przez Radę Nadzorczą PGE), Eryk Kłossowski, Prezes Zarządu PSE, Rafał Miland, Wiceprezes Zarządu PERN SA

Zgodnie ze statutem CEEP, Rada Dyrektorów wybiera ze swego grona Przewodniczącego. Oczekuje się, że taki wybór nastąpi w sierpniu br. Dotychczasowa Rada Dyrektorów CEEP podała się do dymisji dając szansę innym przedstawicielom firm sektora energii do prowadzenia działań odzwierciedlających potrzeby Polski i całej Europy

Centralnej na forum Unii Europejskiej. Poprzednia Rada Dyrektorów otrzymała skwitowanie za swoją dotychczasową działalność. (abc) (http://www.lotos.pl/174/grupa_kapitalowa/centrum_prasowe/aktualnosci, 29.07.2016; dostęp 11.08.2016)

SPOTKANIA

Minister Skarbu w LOTOS

Dawid Jackiewicz, minister Skarbu Państwa RP, spotkał się 22 lipca br. z Zarządem Grupy LOTOS SA. Tematem spotkania były kwestie związane z przyszłością przemysłu naftowego i rynku paliw płynnych w Polsce. W szczególności rozmawiano nt. prac nad strategią GL SA na lata 2016-2020 oraz kwestiami związanymi z dywersyfikacją dostaw ropy naftowej do rafinerii w Gdańsku. Do gdańskiego Naftoportu zmierza tankowiec Atlantis wiozący 2 mln baryłek ropy naftowej zakupionych przez Grupę LOTOS od irańskiego narodowego producenta tego surowca – firmy NIOC (National Iranian Oil Company).

– Minister Skarbu Państwa na spotkaniu z zarządem ponownie podkreślił strategiczne znaczenie dywersyfikacji dostaw surowców dla bezpieczeństwa energetycznego kraju – komentuje Robert Pietryszyn, prezes Zarządu Grupy LOTOS SA – Pan Minister Jackiewicz zaznaczył również istotność współpracy pomiędzy spółkami sektora naftowo-energetycznego dla wzmocnienia tego kierunku działań – dodał prezes Pietryszyn. Minister Jackiewicz zapoznał się również ze stanem zaawansowania kluczowej inwestycji GL SA – Projektu EFRA, wpisującego się w realizację Planu na Rzecz Odpowiedzialnego Rozwoju, przyjętego przez rząd RP. Dzięki nowym instalacjom, rafineria Grupy LOTOS, bez zwiększania dodatkowego przerobu ropy naftowej, wytworzy dodatkowych 900 tys. t wysokiej jakości paliw i 300 tys. t koksu petrochemicznego. Wpłyne to na zdecydowane zwiększenie konkurencyjności Grupy LOTOS na trudnym rynku rafineryjnym w Europie już w 2018 r. (abc)

(http://www.lotos.pl/174/grupa_kapitalowa/centrum_prasowe/aktualnosci, 22.07.2016; dostęp 11.08.2016)

Dokończenie na stronie 425