

# Podstawy chemoinformatyki leków

Podręczniki i Skrypty



Uniwersytetu Śląskiego  
w Katowicach  
nr 191

50<sup>lat</sup>  
**Uniwersytetu**  
**Śląskiego**  
w Katowicach

Jarosław Polański, Andrzej Bąk

# Podstawy chemoinformatyki leków

Wydanie drugie rozszerzone

Wydawnictwo Uniwersytetu Śląskiego • Katowice 2018

## Recenzenci

*prof. dr hab. Barbara Malawska, Collegium Medicum UJ*  
*dr hab. inż. Barbara Dębska, prof. Politechniki Rzeszowskiej*

Reprodukcje rysunków za zgodą odpowiednio: American Chemical Society (ramki 79, 160, 180–182), Wiley-VCH Verlag GmbH&Co. (ramki 80, 90, 168, 169, 176).

# Spis treści

Przedmowa do wydania drugiego .....	9
1. Wstęp .....	11
2. Przedmiot chemii i chemoinformatyki .....	13
3. Komputer jako narzędzie pracy chemika – problemy wprowadzania danych .....	21
3.1. Kodowanie cząsteczek w chemoinformatyce .....	26
3.2. Grafy cząsteczek chemicznych .....	26
3.3. Notacja liniowa .....	28
3.3.1. Kody SMILES .....	30
3.3.1.1. SMILES – język kodowania grafów molekularnych .....	30
3.3.1.2. Gramatyka języka SMILES .....	31
3.3.2. Inne systemy notacji liniowej .....	41
3.4. Macierzowe systemy kodowania konstytucji cząsteczki .....	42
3.5. Edytory molekularne .....	50
3.6. Standardy wymiany informacji strukturalnej .....	60
3.7. Nomenklatura chemiczna .....	68
3.8. Związek chemiczny – <i>in vitro</i> oraz <i>in silico</i> .....	72
3.9. Ontologia chemiczna – formalna definicja związku chemicznego .....	73
3.10. Właściwości związków chemicznych i ich pomiary .....	76
3.11. Chemiczne bazy danych .....	78
3.11.1. Organizacja bazy danych .....	78
3.11.2. Typy baz danych .....	80
3.11.3. Strukturalne bazy danych ligandów .....	83
3.11.4. Strukturalne bazy danych makromolekuł .....	86
3.11.5. Bazy literaturowe .....	90
4. Chemia <i>in silico</i> – problemy przetwarzania danych .....	93
4.1. Deskryptory molekularne .....	93
4.1.1. Kodujące deskryptory konstytucyjne .....	95
4.1.2. Deskryptory daktyloskopowe cząsteczki .....	95
4.1.3. Deskryptory obliczane na podstawie atomowej reprezentacji cząsteczki ...	95
4.1.4. Deskryptory obliczane na podstawie fragmentów molekularnych .....	96
4.1.5. Deskryptory topologiczne .....	96
4.1.6. Proste deskryptory geometryczne .....	97
4.1.7. Złożone deskryptory geometryczne .....	97
4.1.7.1. Deskryptory pola oddziaływań cząsteczkowych .....	97
4.1.7.2. Deskryptory profilu konformacyjnego .....	98
4.1.7.3. Deskryptory wirtualnego miejsca receptorowego .....	98
4.1.7.4. Deskryptory receptorowe .....	100
4.1.8. Deskryptory złożonych systemów cząsteczkowych ligand – receptor .....	100
4.1.9. Deskryptory skorelowane z właściwościami .....	100
4.1.9.1. Korelaty właściwości globalnych .....	100
4.1.9.2. Korelaty fragmentów molekularnych – stałe Hammetta .....	101

4.1.9.3. Prognozowanie właściwości związków chemicznych .....	104
4.2. Modelowanie struktur 3D .....	107
4.2.1. Generatory struktur 2D .....	110
4.2.2. Generatory struktur 3D .....	110
4.2.2.1. Semiempiryczne metody chemii kwantowej .....	116
4.2.2.2. Mechanika molekularna .....	117
4.2.2.3. Dynamika molekularna .....	125
4.2.2.4. Techniki generowania konformerów .....	133
4.2.3. Pakiety do modelowania molekularnego .....	134
4.3. Podobieństwo strukturalne cząsteczek .....	139
4.4. Synteza i retrosynteza .....	152
4.4.1. Synteza organiczna – odwzorowanie reagentów w produkty .....	154
4.4.2. Projektowanie syntezy organicznej – odwzorowanie produktów w reagenty .....	155
4.4.3. Retroreakcje – nomenklatura syntonów .....	160
4.4.4. Operacje na syntonach .....	160
4.4.4.1. Przekształcenie syntonu w reagent .....	160
4.4.4.2. Modyfikacje syntonów .....	162
4.4.4.3. Odwrócenie reaktywności .....	163
4.4.5. Komputerowo wspomagane projektowanie syntezy (CASD) .....	164
4.5. Eksploracja baz danych .....	166
4.6. Chemometria – analizy złożonych danych chemicznych .....	169
4.6.1. Wybrane metody analizy struktury danych .....	170
4.6.2. Podstawowe pojęcia chemometrii .....	171
4.6.3. Zmienne i ich korelacje .....	171
4.6.4. Wstępna transformacja danych .....	173
4.6.5. Eksploracja danych .....	175
4.6.6. Jednoczynnikowa analiza wariancji .....	175
4.6.7. Wieloczynnikowa analiza wariancji .....	178
4.6.8. Algorytmy genetyczne .....	180
4.6.9. Wybrane techniki projekcji i wizualizacji wielu zmiennych .....	183
4.6.10. Analiza czynników głównych .....	183
4.6.11. Sieci neuronowe .....	185
4.6.12. Analiza regresji .....	198
4.6.13. Metoda najmniejszych kwadratów .....	201
4.6.14. Analiza regresji czynników głównych .....	205
4.6.15. Metoda częściowych najmniejszych kwadratów .....	205
4.6.16. Walidacja krzyżowa .....	207
4.7. Komputerowe wspomaganie projektowania molekularnego .....	212
4.8. Synteza ukierunkowana na właściwości .....	214
4.8.1. Intuicja i szczęśliwy traf w poszukiwaniach i odkryciach nowych leków .....	215
4.8.2. Schematy racjonalne w poszukiwaniu nowych leków – badanie zależności między budową a działaniem związków .....	215
4.8.3. Badania przesiewowe metodami chemii kombinatorycznej .....	223
4.8.4. Projektowanie molekularne <i>in silico</i> .....	224
4.8.5. Projektowanie leków, kiedy dostępne są dane strukturalne receptora .....	224
4.8.6. Dokowanie molekularne .....	227
4.8.7. Projektowanie oparte na szeregu aktywnych ligandów .....	233
4.8.8. Ilościowe modele zależności struktura – aktywność QSAR .....	235
4.8.8.1. Modele 0D QSAR .....	236
4.8.8.2. Modele 1D QSAR .....	237
4.8.8.3. Modele 2D QSAR .....	237
4.8.8.4. Modele 3D QSAR .....	239
4.8.8.5. Modele 4D QSAR .....	243
4.8.8.6. Modele 5D QSAR .....	244

4.8.8.7. Modele 6D QSAR .....	245
4.8.8.8. Modele RI-QSAR z wirtualnym receptorem modelowanym danymi innymi niż szereg jego aktywnych ligandów .....	245
4.8.8.9. Modele RD QSAR z rzeczywistymi danymi receptorowymi .....	246
4.8.9. Lekotyipy .....	247
<b>5. Internetowe zasoby i nowe technologie chemoinformatyczne <i>on-line</i> .....</b>	<b>249</b>
<b>Literatura dodatkowa .....</b>	<b>255</b>
<b>Dodatek .....</b>	<b>257</b>

## Przedmowa do wydania drugiego

Zainteresowanie pierwszym wydaniem podręcznika przerosło nasze oczekiwania. Był on dostępny w ograniczonym nakładzie, który od dawna jest wyczerpany, gdyż od czasu publikacji minęło już kilka lat. Okazało się też, że w pierwszym wydaniu podręcznika wiele drobnych błędów wymagało korekt. Chemoinformatyka to prężnie rozwijająca się nowa gałąź chemii, wciąż powstają nowe koncepcje i pomysły. Z nadzieją na wypełnienie luki na rynku podręczników z tej dziedziny postanowiliśmy przygotować nowe, poprawione i uzupełnione wydanie naszej publikacji.



# 1. Wstęp

Chemia leków (ang. *medicinal chemistry*) jest interdyscyplinarną nauką obejmującą chemię organiczną, biologię molekularną, chemię obliczeniową, a także farmację i farmakologię. Angielski termin *medicinal chemistry*, odnoszący się do leków, często tłumaczony jest na polski jako „chemia medyczna” i czasami interpretowany w sensie poszerzonym o problemy chemii w medycynie. Istnieją jednak inne nauki, jak biochemia, których przedmiot odnosi się do takich zastosowań, znacznie precyzyjniej je opisując. Warto zauważyć, że także w literaturze angielskiej termin *medicinal chemistry* rozumie się czasami nieco szerzej niż tylko jako chemię leków. Rzadziej spotyka się tam pojęcie *medical chemistry*, które jest jeszcze bliższe polskiemu terminowi „chemia medyczna”.

Ze względu na cechujący *homo sapiens* instynkt przetrwania sztuka leczenia ludzi, czyli medycyna, zawsze była przedmiotem ogromnego zainteresowania człowieka. Leki poszukiwane są od czasów prehistorycznych. Pierwsze takie substancje odkrywano przypadkowo. Zadaniem chemii jest wytwarzanie molekuł wykazujących pożądane działania. Stosownie do tego lekiem określa się substancję o pożądanym profilu aktywności biologicznej. Ponieważ farmaceutyki (potocznie: leki) reprezentują najważniejszą część rynku, problemy związane z otrzymywaniem takich substancji określa się często mianem *projektowania i poszukiwania leków* (ang. *drug design and discovery*). Nawet dzisiaj poszukiwanie nowych farmaceutyków jest bardzo skomplikowanym procesem, który z trudem poddaje się racjonalizacji. Złożoność opracowania nowego leku podkreśla stosowany czasami termin „racjonalne projektowanie” (ang. *rational design*) leku. Pojęcie to przeczy logice i jest tautologią językową – bo czy możliwe jest projektowanie nieracjonalne? Dostrzegając te problemy nomenklaturowe, w ostatnim czasie termin „(racjonalne) projektowanie leków” zastępuje się często bardziej precyzyjnym określeniem „*projektowanie molekularne*”, które można zdefiniować jako konstruowanie nowych cząsteczek o określonym profilu aktywności chemicznej lub biologicznej.

## Ramka 1

Najwcześniejszy przekaz pisany dotyczący terapii datuje się na XVI wiek p.n.e. Na papierze zanotowano wówczas ponad 800 receptur farmaceutycznych, w tym recepturę preparatu opium. Wiedza na temat właściwości terapeutycznych ziół, roślin i praktyk leczniczych rozprzestrzeniła się w starożytnych cywilizacjach na terytoriach, przez które prowadziły szlaki handlowe i na których toczyły się wojny. Przepisy adaptowano powszechnie z innych kultur, wzbogacając własne umiejętności i sztukę leczenia. W starożytnej Grecji i Rzymie ważną częścią farmakologii była umiejętność wytwarzania odtrutek przeciw truciznom,

które w tym czasie często stosowano. *Mithridaticum* stanowi przykład takiego preparatu zawierającego aż 54 składniki stanowiące remedium na różne trucizny. Środek ten przygotowany został w I wieku p.n.e. dla króla Pontu Mityrydatesa. Ważne odkrycia w farmakologii pochodzą z okresu średniowiecza. Paracelsus (1493–1541) usiłował znaleźć lek na każdą chorobę. Wraz z powstaniem nowoczesnej chemii przedefiniowano także przedmiot zainteresowania farmakologii. Już w 1809 roku w pierwszym numerze „Bulletin de Pharmacie” czytamy, że ulepszanie metod ekstrakcji jest istotnym czynnikiem, który warunkuje możliwość zastąpienia złożonych i słabo zdefiniowanych ekstraktów preparatami czystych substancji, na przykład alkaloidów. Dzisiaj farmakologia jest interdyscyplinarną nauką opierającą się z jednej strony na nowoczesnej chemii organicznej, chemii obliczeniowej i medycynie, a z drugiej strony na fizyce, która dostarcza technologii i urządzeń pozwalających na precyzyjne badania losów ksenobiotyków w żywych organizmach oraz mechanizmów działania tych substancji. Racjonalne metody poszukiwania leków są coraz ważniejsze w nowoczesnym przemyśle farmaceutycznym oraz jego sektorze badawczo-rozwojowym. Farmaceutyki coraz lepiej imitują naturalne efekторы, a współczesna cywilizacja nie może obyć się bez leków, które nie tylko leczą, lecz także poprawiają jakość naszego życia.

Mimo istotnych trudności projektowanie coraz częściej staje się integralną częścią poszukiwania nowych leków. Projektowanie molekularne jest nauką w dużym stopniu interdyscyplinarną. Współcześnie zaś w metodach projektowania molekularnego w coraz większym stopniu wykorzystuje się techniki obliczeniowe oraz informatykę. Konstruowaniem zaprojektowanych obiektów molekularnych zajmuje się synteza organiczna. Obecnie większość nowych leków to związki syntetyczne. Także współczesne metody syntezy organicznej korzystają ze wsparcia informatycznego. Interdyscyplinarne połączenie tradycyjnych kierunków chemii oraz informatyki doprowadziło ostatecznie do uformowania się chemoinformatyki (ang. *chemoinformatics*, *cheminformatics*, *chemiinformatics*), przedmiotu, w którego polu zainteresowania leży informatyka projektowania molekularnego oraz informatyka cząsteczkowej chemii jako podstawowego obiektu manipulacji w projektowaniu molekularnym.

Na polskim rynku księgarskim dostępnych jest wiele podręczników poświęconych chemii leków oraz projektowaniu leków, brakuje natomiast podręcznika poświęconego chemoinformatyce. W zamyśle autorów niniejszy skrypt ma za zadanie wypełnić tę lukę. Ponieważ w swych źródłach chemoinformatyka łączy się z projektowaniem molekularnym, wyeksponowaliśmy w szczególności problemy chemoinformatyki leków. Odpowiada to profilowi nowych specjalności: chemii leków oraz chemii informatycznej, uruchomionych w Instytucie Chemii Uniwersytetu Śląskiego w Katowicach.

W skrypcie umieszczono także dwa rozdziały poświęcone bazom danych chemicznych oraz wykorzystaniu zasobów internetowych w chemii i chemoinformatyce. Rozbudowana wersja tych rozdziałów dostępna jest w postaci ćwiczeń w projekcie iCSE ([icse.us.edu.pl](http://icse.us.edu.pl)).

Redakcja  
Magdalena Starzyk

Projekt okładki  
Hanna Olsza

Korekta  
Joanna Zwierzyńska

Łamanie  
Hanna Olsza

Copyright © 2018 by  
Wydawnictwo Uniwersytetu Śląskiego  
Wszelkie prawa zastrzeżone

**ISSN 1644-0552**

**ISBN 978-83-8012-896-5** (wersja drukowana)

**ISBN 978-83-8012-897-2** (wersja elektroniczna)

Wydawca  
**Wydawnictwo Uniwersytetu Śląskiego**  
**ul. Bankowa 12B, 40-007 Katowice**  
[www.wydawnictwo.us.edu.pl](http://www.wydawnictwo.us.edu.pl)  
e-mail: [wydawus@us.edu.pl](mailto:wydawus@us.edu.pl)

Wydanie II. Ark. druk. 18,0.  
Ark. wyd. 20,5. Papier offset. kl. III, 90 g  
Cena 30 zł (+ VAT)

Druk i oprawa:  
„TOTEM.COM.PL Sp. z o.o.” Sp.K.  
ul. Jacewska 89, 88-100 Inowrocław