

Введение в химическую информатику

Лекция 1



Вводная информация о курсе

Лектор: Киреева Наталья Вадимовна

Материалы лекций будут доступны здесь: <http://cheminformatics.mozello.com>

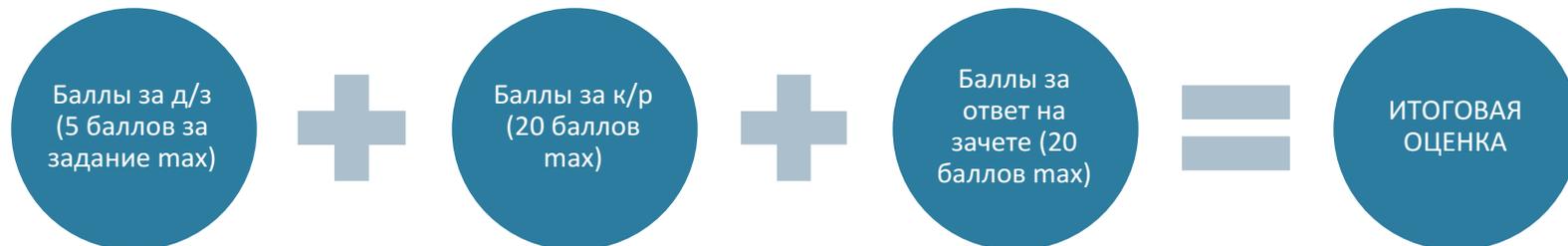
Пожаловаться/ полюбопытствовать/ пообщаться можно по e-mail:

nkireeva@gmail.com

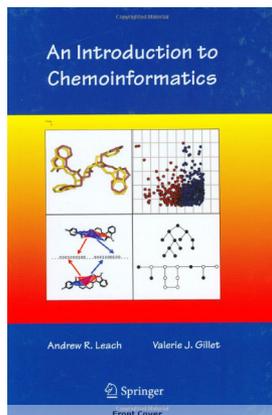
Форма итогового контроля: зачет

Допуск к зачету:

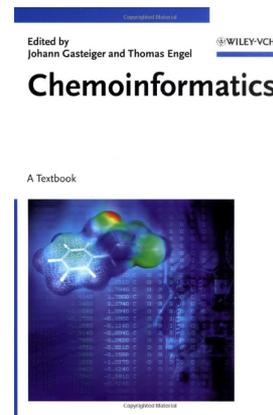
- ❖ 2 выполненных домашних задания
- ❖ написанная контрольная работа



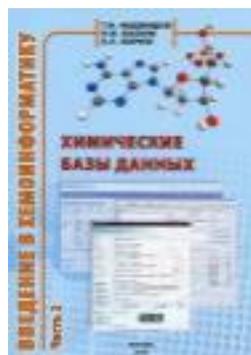
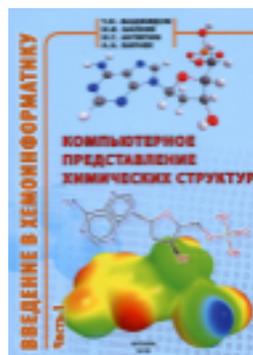
Основная рекомендованная литература



An Introduction to Chemoinformatics,
Andrew R. Leach, Valerie J. Gillet,
Springer 2007.



Chemoinformatics -A Textbook, Johann
Gasteiger and Thomas Engel, Wiley-VCH
2003



Введение в хемоинформатику, Т.И.Маджидов, И.И.Баскин, И.С. Антипин, А.А.Варнек

Varnek, A. and Baskin, I. I. (2011), Chemoinformatics as a Theoretical Chemistry
Discipline. Mol. Inf., 30: 20-32. doi: 10.1002/minf.201000100

Предмет химической (хемо-) информатики

F. K. Brown (1998) Chemoinformatics: what it is and how does it impact drug discovery?
Annual Reports in Medicinal Chemistry, 33, 375-384

“Химическая информатика означает совместное использование информационных ресурсов для преобразования данных в информацию и информации в знания для быстрого принятия наилучших решений при поиске соединений-лидеров в разработке лекарств и их оптимизации.”

G. Paris (August 1999 ACS meeting)

“Химическая информатика это научная дисциплина, охватывающая дизайн, создание, организацию, управление, поиск, анализ, распространение, визуализацию и использование химической информации.”

J. Gasteiger and T. Engels (editors) (2003). *Chemoinformatics: a textbook*. Wiley-VCH.

“Химическая информатика это применение методов информатики для решения химических задач.”

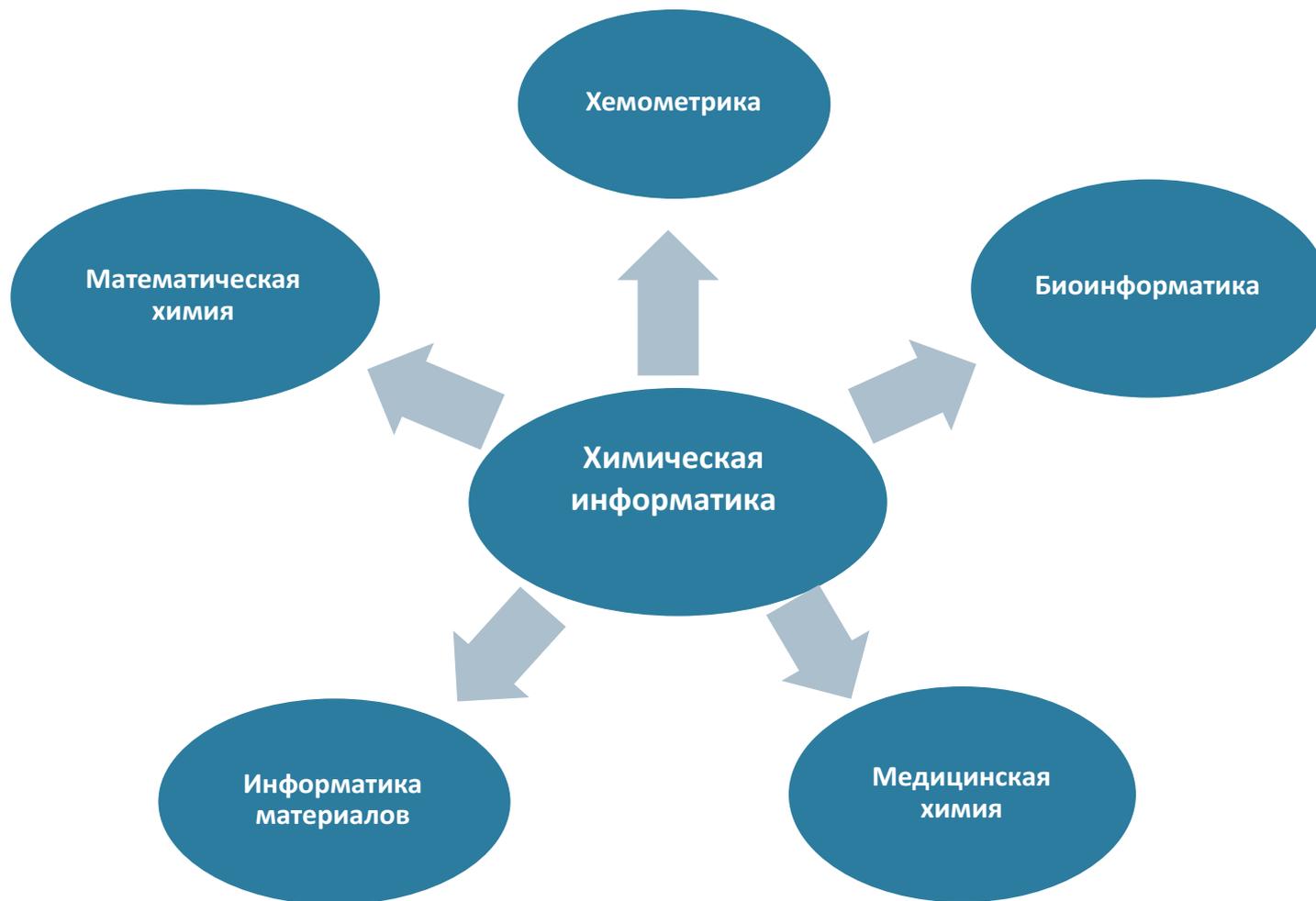
Химическая информатика (хемоинформатика)

Область с одной из наиболее долгих историй среди прикладных компьютерных дисциплин, вместе с тем выделившаяся в отдельное направление относительно недавно

Направление развития определяется междисциплинарным характером этой области



Химическая информатика и смежные дисциплины



Химическая информатика: молекулярная модель

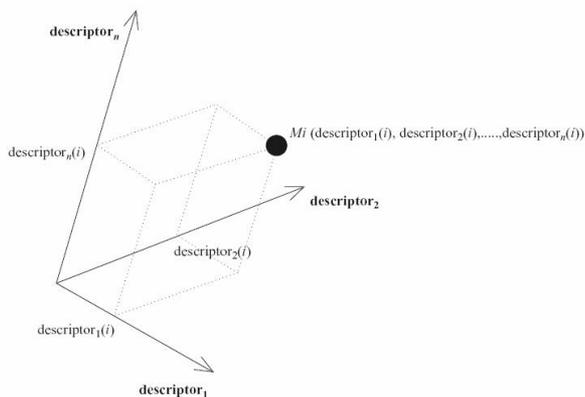
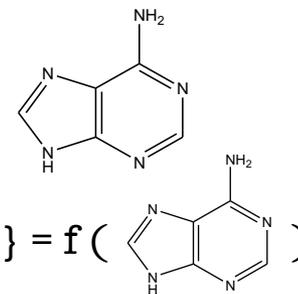


Figure 1.2. N-dimensional chemical space with a molecule M at position i

Объекты:

- молекулярные графы;

- векторы дескрипторов $\{D_i\} = f(\text{structure})$



$$\text{Molecule } i: M(i) = \sum_1^j d_j(i)$$

При векторном описании химической структуре ставится в соответствие вектор молекулярных дескрипторов (числовых параметров, характеризующих структуру соединения).

Альтернативное описание возможно при помощи непосредственно молекулярных графов (графовые ядра (кernels))

Структура соединения

Способ
представления
(описания)
соединения

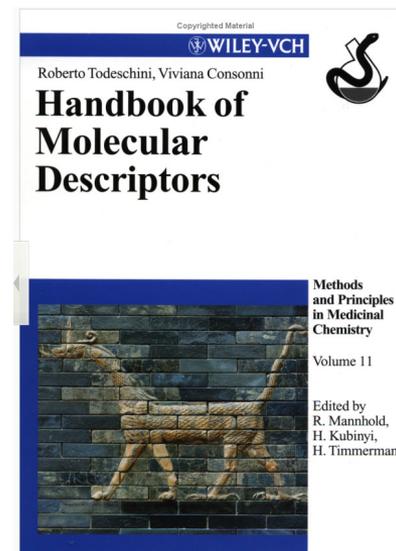
Методы
машинного
обучения

Прогнозируемые
свойства

Дескрипторное описание данных

> 7000 типов молекулярных дескрипторов

- ❖ Простейшие структурные и геометрические (молекулярный вес, число атомов определенного типа, размер молекулы, расстояние между функциональными группами)
- ❖ Топологические (индексы Рандича и Винера, ...)
- ❖ Физико-химические
- ❖ Молекулярные фрагменты (последовательности атомов и связей, атомы с окружением, ...)
- ❖ Фармакофорные дескрипторы
- ❖ Молекулярные отпечатки
- ❖



Информатика материалов

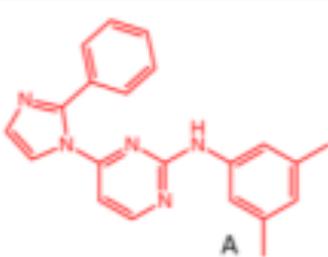
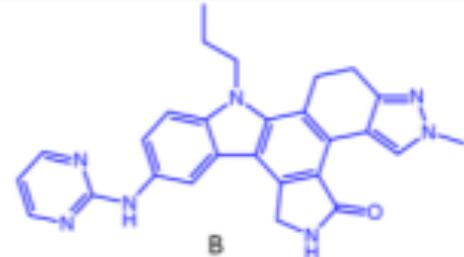
Более разрозненные источники информации по дескрипторам, отсутствие специализированных программ, большая по сравнению с молекулярными дескрипторами привязка к объекту исследования

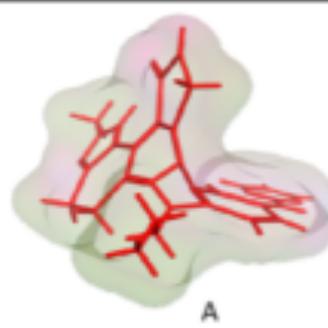
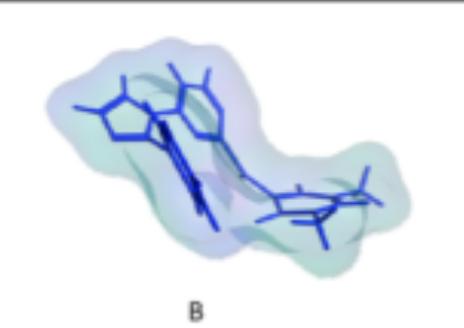
- ❖ Отпечатки различного типа
- ❖ Дескрипторы, описывающие свойства химических элементов
- ❖ Параметры, содержащие кристаллографическую информацию
- ❖ Дескрипторы описывающие состав соединений
- ❖ Различные физико-химические характеристики
- ❖ ...

Химическая информатика: молекулярная модель

Принцип молекулярного подобия (сходства): структурно сходные молекулы предположительно обладают близкими свойствами

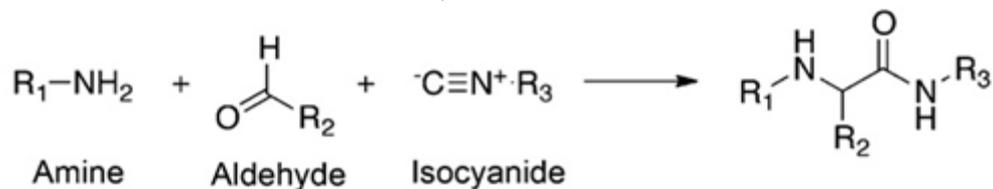
Chemical similarity	Mol. weight	LogP	Rotatable bonds	Aromatic rings	Heavy atoms
	A	341.4	5.23	4	4
B	463.5	4.43	4	5	35

Molecular similarity		
2D similarity	A	B

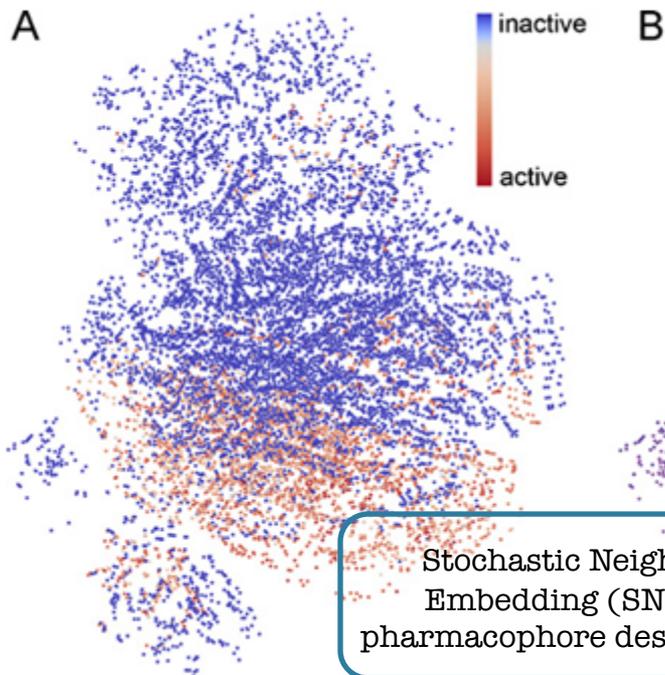
3D similarity		
	A	B

Химическая информатика: взаимосвязь структура-свойство

15,840 продуктов (-aminoacyl amide derivatives) трехкомпонентной конденсации Уги (взаимодействие аминов, изоцианидов, карбонильных соединений, органических/неорганических кислот).

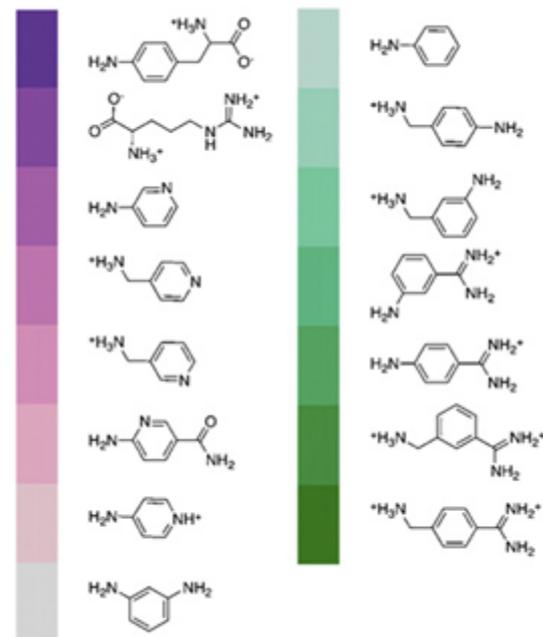


Распределение данных по значению биологической активности (ингибирование триптазы)



Распределение данных по подструктуре

amine building block R1



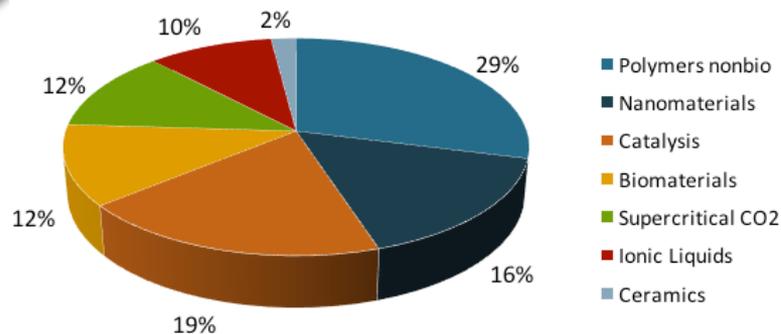
Химическая информатика: химическое пространство данных

Термин, относящийся к оценке количества соединений при виртуальном скрининге, направленном на решение конкретной задачи. В общем случае, совокупность химических соединений, описываемых набором параметров, устанавливающих взаимосвязь структура-свойство



Оценка числа соединений зависит от размера, состава, учета стереоизомерии:

- ❖ $6.2 \cdot 10^{13}$ ациклические алканы без учета стереоизомеров (<40 атомов)
- ❖ 10^{23} соединений (C, N, O, S, P, Se, Si, Hal; <36 атомов)
- ❖ 10^{180} соединений (C, N, O, S, P; <1000 Da, стереоизомеры)
- ❖ 166 млрд «виртуальных» соединений (C, N, O, S, Hal; < 17 атомов (GDB-17 Database))



Chem. Rev. 2012, 112, 2889–2919

Химическая информатика: синтез свойств

Описание и поиск химических структур и данных

Анализ и работа с химическими базами данных

Молекулярный дизайн новых химических соединений/материалов, обладающих заданными свойствами

Компьютерный синтез / дизайн эксперимента

Автоматическая идентификация ошибочных данных

Доэкспериментальное обоснование гипотез

Алгоритмы для работы с большими данными

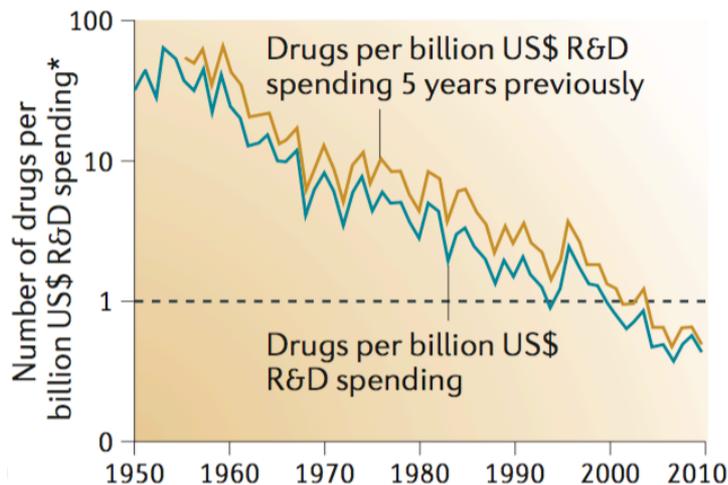
Комплементарное применение химической информатики с другими дисциплинами теоретической химии

Создание механизмов для свободного обмена информацией и эффективной ее обработки

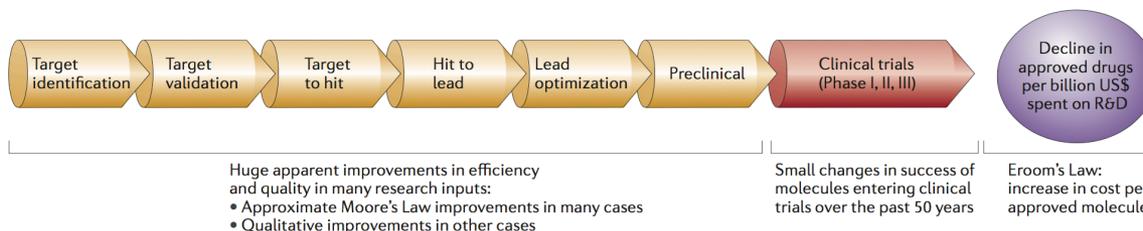
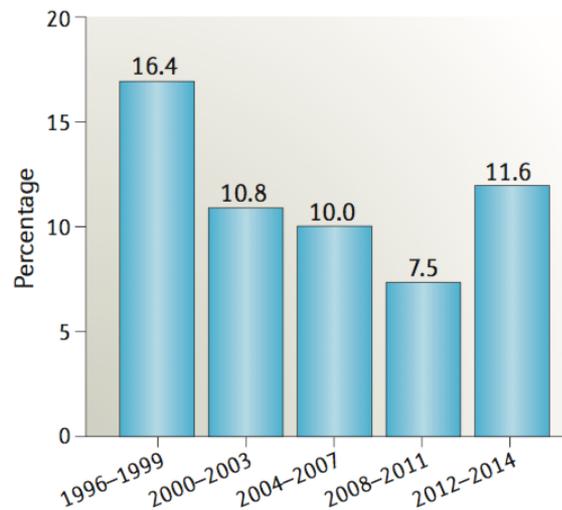
Дизайн лекарств



«BIG PHARMA»



Percentage likelihood of moving from Phase I to launch



Возможные
причины
недостаточной
эффективности

Проблема "Лучше, чем Beatles"

Проблема "Осторожный чиновник"

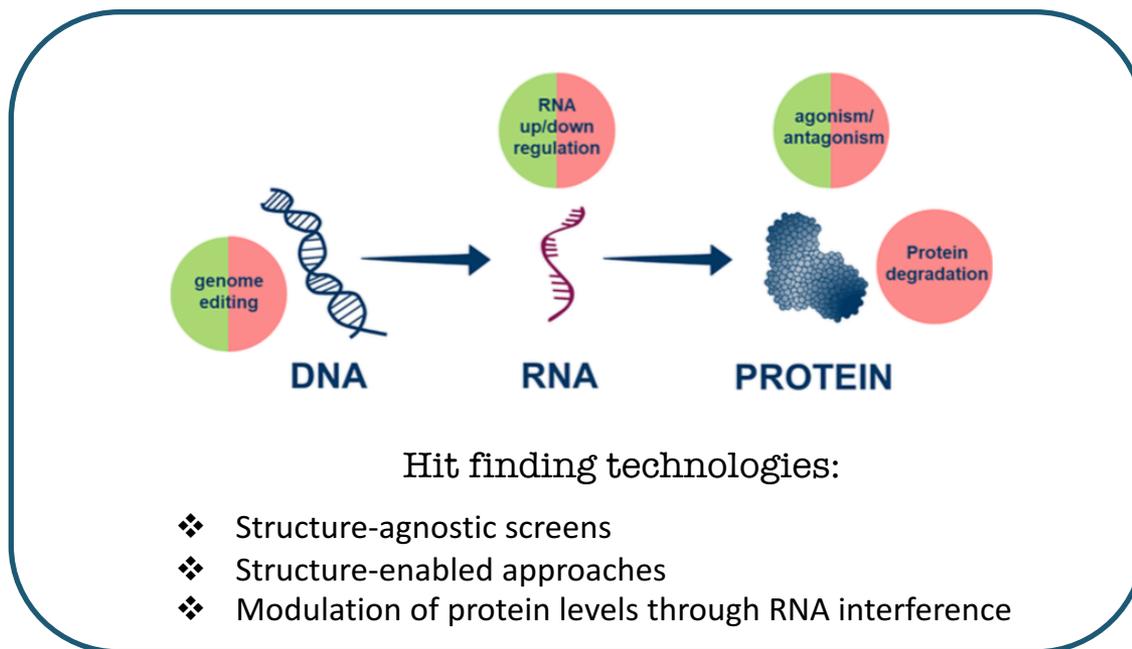
Инерционность, консерватизм

Scannell et al Nature Reviews Drug Discovery 11, 191-200 (March 2012)

McGilvray Fresh hunting ground Nature. 2016 May 12;533(7602):S65-7.

Smetana Trends in clinical success rates Nature Reviews Drug Discovery 15, 379-380 (2016)

Механизм действия (Mode-of-Action) как основа определения стратегии скрининга новых соединений



Examples of possible cheminformatics support:

- ❖ Aid new technologies (e.g. DNA-encoded libraries DEL)
- ❖ Data sharing and analysis
- ❖ Integrating new machine learning technologies (Deep learning, one-shot-learning, machine learning for dynamic systems, etc)

«AI-powered drug discovery captures pharma interest»

AI company/ location	Technology	Announced partner/ location	Indication(s)	Deal date
Atomwise	Deep-learning screening from molecular structure data	Merck	Malaria	2015
BenevolentAI	Deep-learning and natural language processing of research literature	Janssen Pharmaceutica (Johnson & Johnson), Beerse, Belgium	Multiple	November 8, 2016
Berg, Framingham, Massachusetts	Deep-learning screening of biomarkers from patient data	None	Multiple	N/A
Exscientia	Bispecific compounds via Bayesian models of ligand activity from drug discovery data	Sanofi	Metabolic diseases	May 9, 2017
GNS Healthcare	Bayesian probabilistic inference for investigating efficacy	Genentech	Oncology	June 19, 2017
Insilico Medicine	Deep-learning screening from drug and disease databases	None	Age-related diseases	N/A
Numerate	Deep learning from phenotypic data	Takeda	Oncology, gastroenterology and central nervous system disorders	June 12, 2017
Recursion, Salt Lake City, Utah	Cellular phenotyping via image analysis	Sanofi	Rare genetic diseases	April 25, 2016
twoXAR, Palo Alto, California	Deep-learning screening from literature and assay data	Santen Pharmaceuticals, Osaka, Japan	Glaucoma	February 23, 2017

Виртуальный скрининг: бесконечная лестница?

OPINION

Virtual screening: an endless staircase?

Gisbert Schneider

Abstract | Computational chemistry — in particular, virtual screening — can provide valuable contributions in hit- and lead-compound discovery. Numerous software tools have been developed for this purpose. However, despite the applicability of virtual screening technology being well established, it seems that there are relatively few examples of drug discovery projects in which virtual screening has been the key contributor. Has virtual screening reached its peak? If not, what aspects are limiting its potential at present, and how can significant progress be made in the future?

Необходимость индивидуального подбора метода виртуального скрининга для решения конкретной задачи, разработка новых методов

Достаточная интеграция методов виртуального скрининга в процесс разработки

Акцент на получение новых фундаментальных знаний

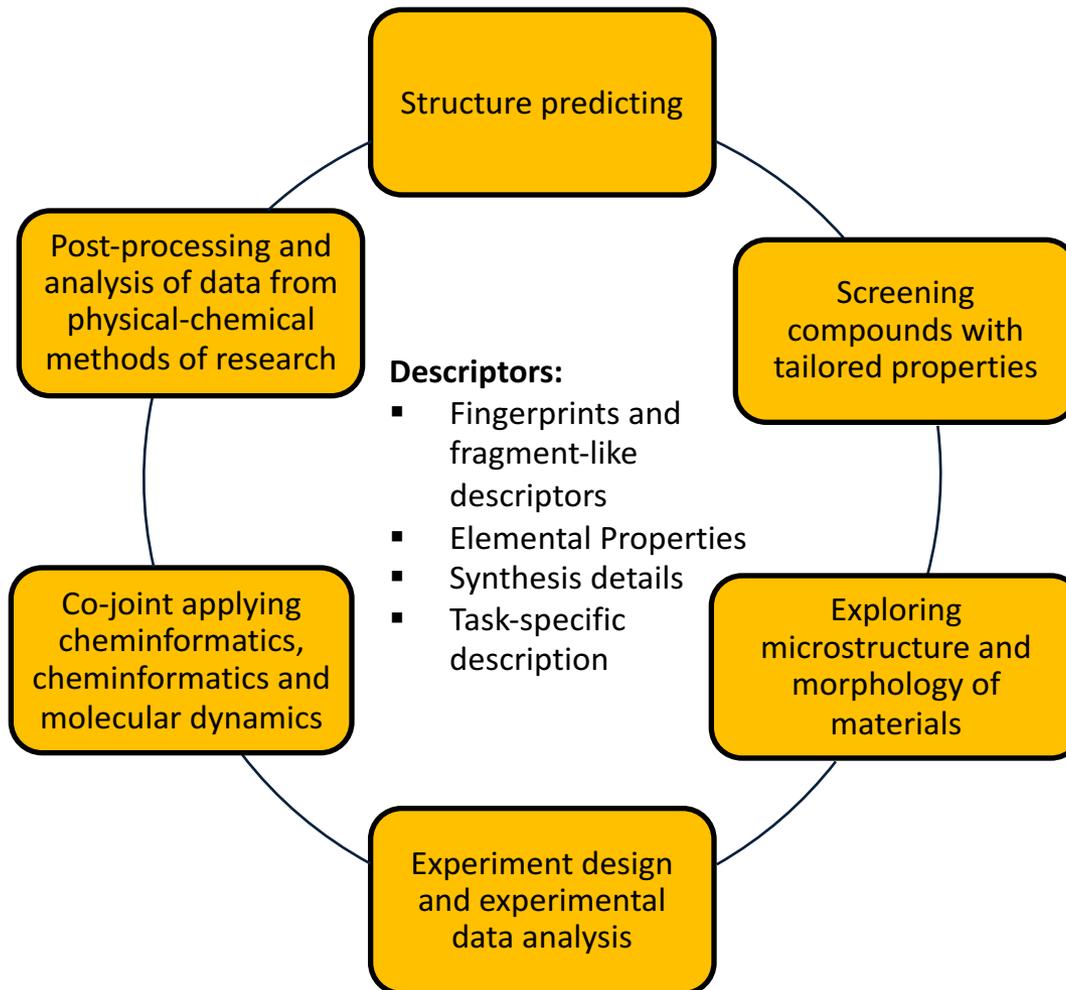
Обмен данными

Kogej, T., N. Blomberg, et al. (2013). "Big pharma screening collections: more of the same or unique libraries? The AstraZeneca-Bayer Pharma AG case." *Drug discovery today* **18**(19-20): 1014-1024.

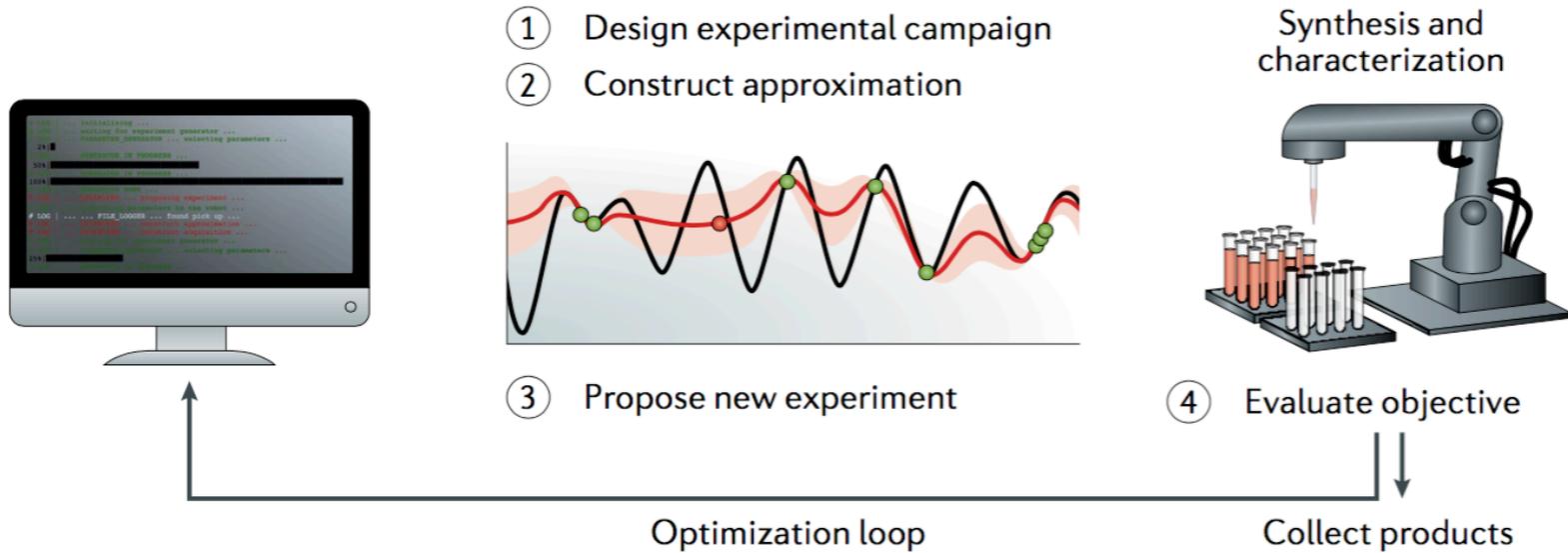
Информатика материалов (Materials informatics)



MATERIALS INFORMATICS: OBJECTS AND APPLICATIONS



Autonomous experimentation



Поиск химических данных в патентах и других документах



SureChEMBL: 17 миллионов соединений, извлеченных из более чем 14 млн патентов

SureChEMBL Query Help | Quick Reference Guide | Patent Number Search | Clear form | Field Search

SELECT STRUCTURE SEARCH

- Substructure
- Similarity
- Identical
- Basic
- Major Match

FILTER BY MOLECULAR WEIGHT

0 to 800

SEARCH FOR STRUCTURE IN DOC SECTION(S)

- All
- Title or Abstract
- Claims
- Description
- Images

PATENT AUTHORITIES

- All chemically annotated authorities (?)
- US Applications
- US Granted
- EP Applications
- EP Granted
- WO
- JP
- All authorities (inc. DocDB) (?)

PUBLICATION DATE

Example: YYYYMMDD; YYYY; YYYYMMDD TO YYYYMMDD; YYYY TO YYYY

Search

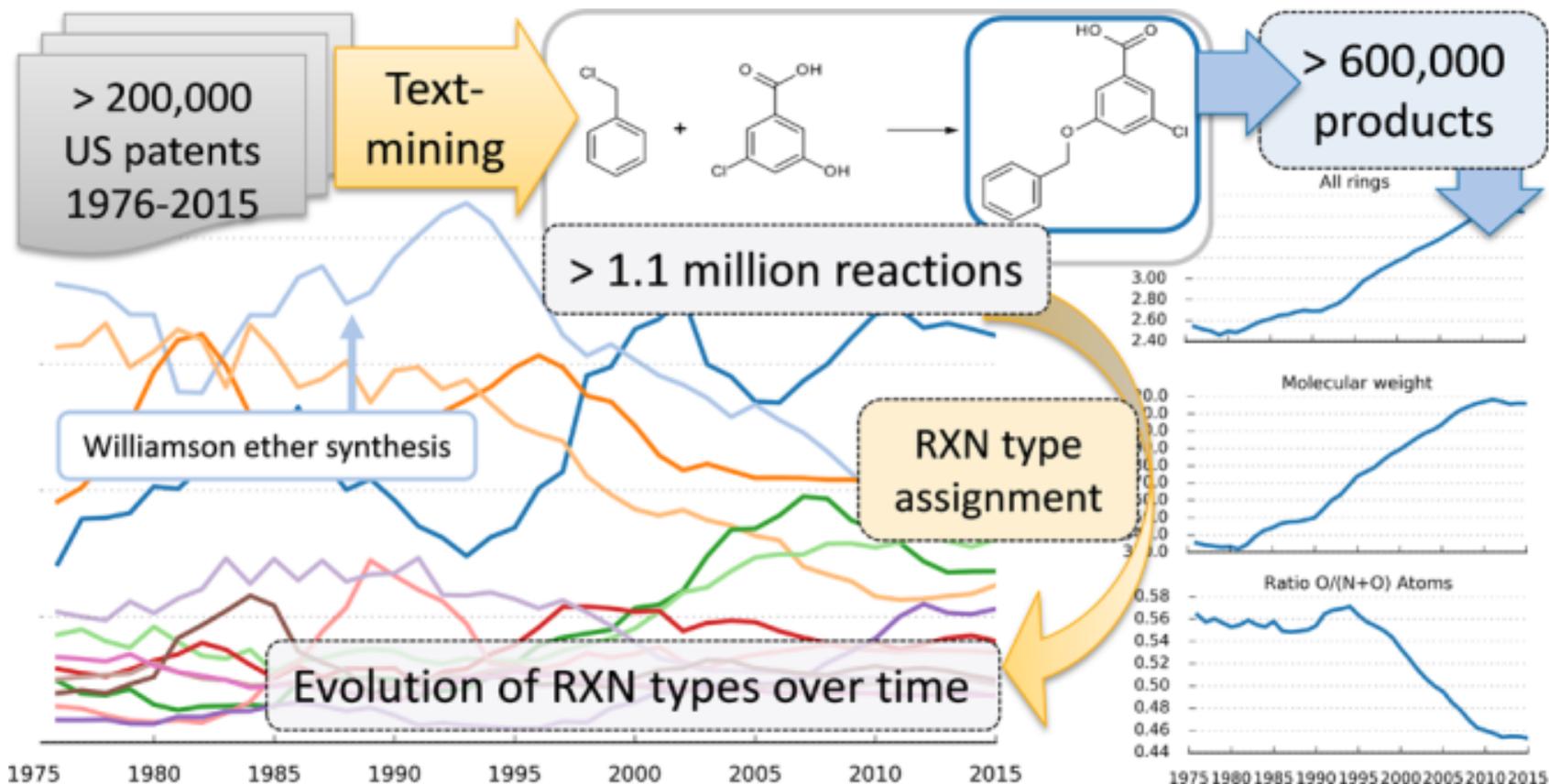
Our Chemistry Annotation Coverage
Chemistry annotations for US, EP, WO full text and JP abstracts are now available as follows:

Structures from text: from Jan 1, 1976 to Sep 11, 2017

Structures from images: from **Jan 1, 2007** to Sep 11, 2017

<https://www.surechembl.org/search/>

Big Data from Pharmaceutical Patents: A Computational Analysis of Medicinal Chemists' Bread and Butter

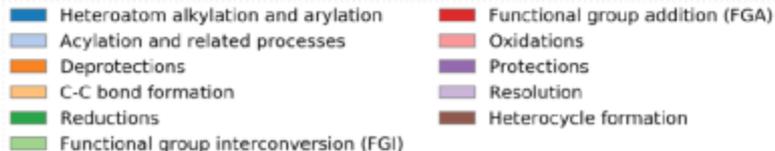
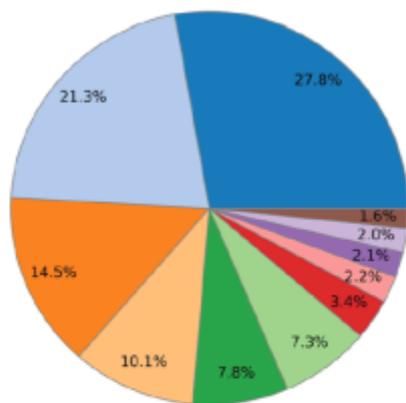


Big Data from Pharmaceutical Patents: A Computational Analysis of Medicinal Chemists' Bread and Butter

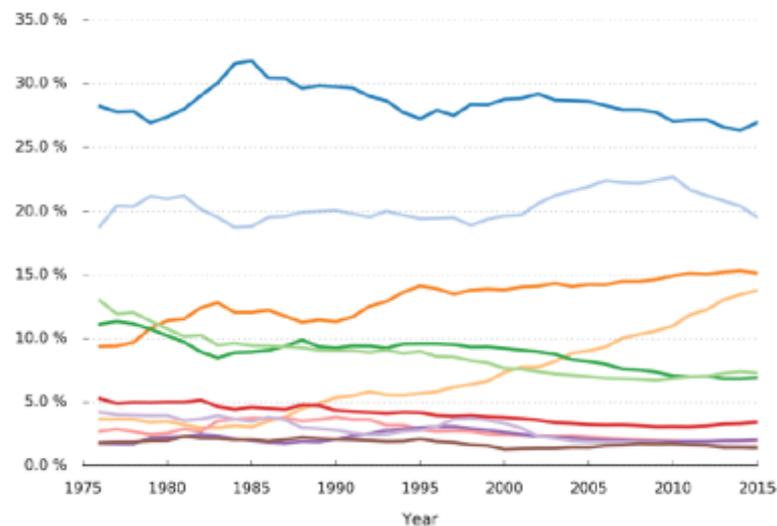
Table 1. Overview of Recent Studies Investigating Reactions and Molecules

	This Work	Carey, 2006	Leeson, 2007	Walters, 2011	Roughley, 2011	Brown, 2015
Date Range	1976–2015	2005–2006	1964–2007	1959–2009	2008–2009	1984–2014
Source	Granted U.S. Patents and Applications	In-house synthesis data GSK, AZ, Pfizer	FDA, Prous, GVKBio, Cerep	GVK, ChEMBL, Journals: <i>J. Med. Chem.</i>	Bellstein, SciFinder, Journals: <i>J. Med. Chem.</i> , <i>Bioorg. Med. Chem.</i> , <i>Bioorg. Med. Chem. Lett.</i>	Journals: <i>J. Med. Chem.</i> , <i>J. Am. Chem. Soc.</i> , <i>Angew. Chem., Int. Ed.</i>
Documents	126 435	unk	>5000	19 299	139	280
Extraction method	Automated	Manual	Automated	Automated	Automated + Manual	Manual
Sampling over time	Continuous	Snapshot	Continuous	Continuous	Snapshot	Snapshot
Focus	Med-chem patent reactions	Process chem R&D	Oral drugs and development cmpds	Med-chem literature	Med-chem publications from GSK, Pfizer, AZ	Med-chem and natural products publications
Reactions	1.15 million unique	1039			7315	unk
Compounds	628 200	128	>120 000	415 284	3566	unk
Highlights	New frameworks, physicochem. props.	Purchased vs installed chirality, large-scale "practical reactions"	Comparison of various companies' data	Physicochemical properties, complexity, common frameworks	Heterocycle analysis, fluorine	PMI analysis, comparison against natural products literature

B



Evolution of major reaction classes over time



Патентный химический поиск



WORLDWIDE RESEARCH & DEVELOPMENT
Medicinal Chemistry

**Structure Clipper – an interactive tool for
extracting chemical structures from patents**

Патентный химический поиск

Желательные возможности:

- Автоматическое извлечение химических структур из исходных PDF документов
- Совместимость с основными типами патентов и статьями
- Требования к возможностям работы с графикой и изображениями
- Ассоциирование извлеченной информации с исходными документами (интегрированная возможность перехода к исходному тексту)
- Удобный пользовательский интерфейс для отображения результатов поиска с подсветкой задействованной информации
- Возможность внесения изменений в извлеченную информацию (модификация химических структур; исправление ошибок в именах соединений, в структурах, преобразованных из изображений)
- Интеграция с программами молекулярного дизайна

Патентный химический поиск

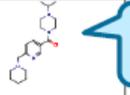
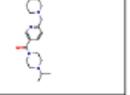
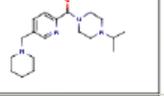
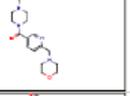
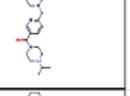
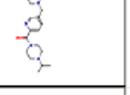
PCAT v5.1: US2005222151.patx

File Edit Select View Properties Analysis Visualization Prototypes ExternalApps Help

Access to *in-silico* models and other chemoinformatic tools

Compound Table

Table Menu | cmpds: 400 loaded 0 selected 77 visible | Hide Show Show Only Show All | < Prev Next > | Page: 13

Id.	Structure	PCAT_DOCUMENT	PCAT_PAGE	PCAT_LOCATION	SC_Type	Chemical Name	N25 Message
compound_12-1							
compound_12-3		US20052221:	12	1337.2691.65	N25	(4-isopropyl	Ok
compound_13-1		US20052221:	13	1349.708.661	OSRA		
compound_13-2		US20052221:	13	428.2407.661	OSRA		
compound_13-8		US20052221:	13	405			
compound_13-9		US20052221:	13	1337.1119.65	N25	(4-isopropyl	Ok

Extracted chemical structures

PDF page viewer highlights location of extracted chemical structures

US 2005/0222151 A1

13

Oct. 6, 2005

Chromatography of the residue (SiO₂; 0.3% EtOH/DCM) gave the title compound as a solid (0.506 g, 46%).

[0125] Step B. 6-(piperidin-1-ylmethyl)-nicotinic acid methyl ester

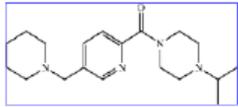
[0126] To a solution of 6-(benzyl)-nicotinic acid methyl ester (0.200 g, 1.21 mmol) and piperidine (0.14 mL, 1.33 mmol) in DCM (15 mL) was added NaH(OAc)₂H (0.380 g, 1.80 mmol). After 18 h, the reaction was diluted with 1 N NaOH (10 mL) and extracted with DCM (2x50 mL). The organic layers were combined, dried (Na₂SO₄), and concentrated. Chromatography of the residue (SiO₂; 1-3% 2 M NH₃ in MeOH/DCM) gave the title compound as an oil (0.210 g, 74%).

[0127] Step C. (4-(4-isopropyl-piperazin-1-yl)-(5-(piperidin-1-yl)methyl)pyridin-3-yl)methanone

[0128] A solution of 6-(piperidin-1-yl)methyl-nicotinic acid methyl ester (0.300 g, 1.28 mmol), and MgI₂·OEt₂ (0.900 g, 3.84 mmol) in THF (15 mL) was stirred for 15 min. A solution of 1-(4-isopropyl-piperazine) (0.325 g, 2.56 mmol) in THF (15 mL) was then added to the reaction drop-wise and the reaction mixture was heated at reflux for 48 h. The reaction mixture was cooled to rt, concentrated, treated with 1 N aq. NaOH (50 mL), and extracted with ethyl acetate (3x50 mL). The organic layers were combined, dried (Na₂SO₄), and concentrated. Chromatography of the residue (SiO₂; 1-3% 2 M NH₃ in MeOH/DCM) gave the title compound as a solid (0.297 g, 70%). MS (ESI⁺): exact mass calcd. for C₂₂H₃₂N₆O, 330.2; m/z found, 331.2 [M+H]⁺. ¹H NMR (400 MHz, CDCl₃): 8.55 (d, J=2.2, 1H), 7.87 (dd, J=7.8, 2.0, 1H), 7.62 (d, J=7.6, 1H), 3.78 (br s, 2H), 3.67 (s, 2H), 3.48 (br s, 2H), 2.77-2.73 (m, 1H), 2.65-2.48 (m, 8H), 1.65-1.59 (m, 4H), 1.49-1.48 (m, 2H), 1.09 (d, J=6.6, 6H).

Example 3

[0131]



(4-isopropyl-piperazin-1-yl)-(5-(piperidin-1-yl)methyl)pyridin-2-yl)methanone

[0132] Step A. 6-(4-isopropyl-piperazine-1-carboxyl)-nicotinic acid methyl ester

[0133] To a solution of pyridine-2,5-dicarboxylic acid 5-methyl ester (1.00 g, 5.50 mmol) and 1-(4-isopropyl-piperazine) dihydrochloride (1.20 g, 6.10 mmol) in DCM (100 mL) was added 1-hydroxybenzotriazole hydrate (HOBt, 1.10 g, 8.30 mmol), 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride (EDC, 1.60 g, 8.30 mmol), and N-methyl morpholine (2.9 mL, 27.0 mmol). After 18 h, the reaction mixture was quenched with 1 N aq. NaHCO₃ (50 mL) and extracted with DCM (3x50 mL). The organic layers were combined, dried (Na₂SO₄), and concentrated. Chromatography of the residue (SiO₂; 2-5% 2 M NH₃ in MeOH/DCM) gave the title compound (1.20 g, 74%).

[0134] B. (5-(Hydroxymethyl)pyridin-2-yl)-(4-isopropyl-piperazin-1-yl)methanone

[0135] A solution of 6-(4-isopropyl-piperazine-1-carboxyl)-nicotinic acid methyl ester (0.500 g, 1.72 mmol) in THF (15 mL) was cooled to -78° C. in a dry ice bath. A solution of lithium tri-tert-butylaluminumhydride (1 M in THF, 3.44 mL) was then added drop-wise to the reaction mixture. The

Патентный химический поиск

PCAT v5.1: US2005222151.pcatx

File Edit Select View Properties Analysis Visualization Prototypes ExternalApps Help

Detail Compound Table Structure Clipper

Table Menu cmpds: 400 loaded 1 selected 77 visible

Id.	Structure	PCAT_DOCUMENT	PCAT_PAGE	PCAT_LOCATION	SC_Type	Chemical Name	N2S Message
compound 12-1		US20052221	12	1349.2282.66	OSRA		
compound 12-3		US20052221	12	1337.2691.65			
compound 13-1		US20052221	13	1349.708.661			
compound 13-2		US2005222	13	428.2407.64	OSRA		
compound 13-8		US20052221	13	409.2821.712	N2S	(4-isopropyl	Ok
compound 13-9		US20052221	13	1337.1119.65	N2S	(4-isopropyl	Ok

Property Value

PCAT_CMP... compound
PCAT_DOC... US2005222
PCAT_PAGE 13
PCAT_LOC... 428.2407.64
SC_Type OSRA
Chemical ...
N2S Message
cdk_path_fp 0,5,13,15,1
Identity EXAMPLE

Selecting a compound in the PCAT table will display the source location in the PDF page viewer.

(4-isopropyl-piperazin-1-yl)-(1-morpholin-4-ylmethyl)-pyridin-3-yl-methanone

[0130] The title compound was synthesized in a similar manner as Example 1 substituting morpholine for piperidine in step B. MS (ESI): exact mass calcd. for $C_{24}H_{32}N_6O_2$, 352.2; *m/z* found, 353.2 [M+H]⁺. ¹H NMR (400 MHz, CDCl₃): 8.60 (d, J=2.3, 1H), 7.72 (dd, J=7.8, 2.3, 1H), 7.48 (d, J=7.8, 1H), 3.79-3.72 (m, 6H), 3.67 (s, 2H), 3.44 (br s, 2H), 2.75-2.72 (m, 1H), 2.60 (br s, 2H), 2.52-2.48 (m, 6H), 1.05 (d, J=6.6, 6H).

[0131] A solution of 6-(4-isopropyl-piperazine-1-carboxyl)-nicotinic acid methyl ester (0.500 g, 1.72 mmol) in THF (15 mL) was cooled to -78° C. in a dry ice bath. A solution of lithium tri-tert-butoxyaluminum hydride (1 M in THF, 3.44 mL) was then added drop-wise to the reaction mixture. The resulting solution was allowed to come to rt and was stirred for 18 h. The reaction was quenched with sat. aq. potassium sodium tartrate (Rochelle's salt, 15 mL) and extracted with DCM (3x50 mL). The organic layers were combined, dried (Na₂SO₄), and concentrated to give the title compound (0.275 g, 61%).

[0136] Step C. 6-(4-isopropyl-piperazine-1-carboxyl)-pyridine-3-carbaldehyde.

[0137] To a solution of (5-hydroxy-methyl-pyridin-2-yl)-(4-isopropyl-piperazin-1-yl)-methanone (0.275 g, 1.10 mmol) in DCM (30 mL) was added MeOH (0.400 g, 5.20 mmol). The reaction was stirred for 6 h, filtered through a pad of diatomaceous earth, and concentrated to give the desired aldehyde (0.250 g, 95%).

[0138] Step D. (4-isopropyl-piperazine-1-yl)-(5-piperidin-1-ylmethyl-pyridin-3-yl)-methanone.

[0139] To a solution of 6-(4-isopropyl-piperazine-1-carboxyl)-pyridine-3-carbaldehyde (0.250 g, 0.95 mmol) and piperidine (0.11 mL, 1.10 mmol) in DCM (20 mL) was added NaB(OAc)₃H (0.300 g, 1.44 mmol). After 18 h, 1 N NaOH (15 mL) was added and the mixture was extracted with DCM (3x25 mL). The organic layers were combined, dried (Na₂SO₄), and concentrated. Chromatography of the

Компьютерный синтез / Дизайн эксперимента

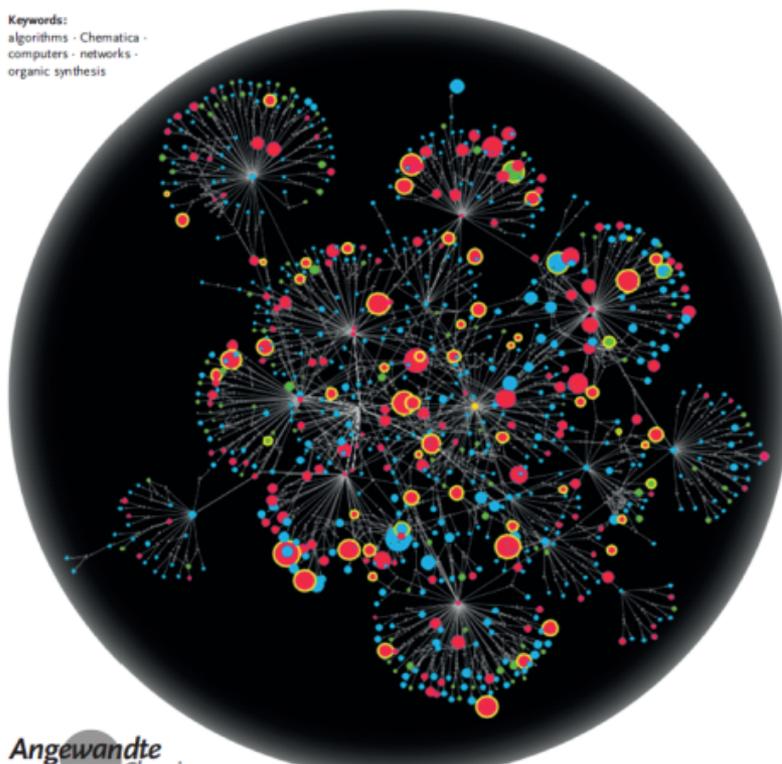
Computer-Aided Synthetic Planning

International Edition: DOI: 10.1002/anie.201506101
German Edition: DOI: 10.1002/ange.201506101

Computer-Assisted Synthetic Planning: The End of the Beginning

Sara Szymkuć, Ewa P. Gajewska, Tomasz Klucznik, Karol Molga, Piotr Dittwald, Michał Startek, Michał Bajczyk, and Bartosz A. Grzybowski*

Keywords:
algorithms · Chemica ·
computers · networks ·
organic synthesis



Компьютерный синтез – совокупность методов анализа данных, специализированных алгоритмов и программного обеспечения для планирования синтеза, прогнозирования протекания и результатов химических реакций.

Компьютерный синтез / Дизайн эксперимента: наиболее известное программное обеспечение

Основанные на имеющихся данных

- LHASA (Logic and Heuristic Applied to Synthesis Analysis) и SECS (Simulation and Evaluation of Chemical Synthesis) - независимые ветви первого программного обеспечения OCSS (Organic Chemical Simulation and Synthesis) - 1960
- SYNLMA
- SYNCHEM
- SYNGEN +FORWARD
- WODCA (оценка фундаментальных свойств связей - полярности, резонанса, поляризуемости) - Johann Gasteiger group 1990th

Неэмпирический компьютерный синтез

- IGOR и IGOR2 (анализ потенциальных реакций, основанных на перераспределении электронов атомных оболочек, расчет энтальпий реакций) - Ivar Ugi 1980th-1990th
- EROS (Elaboration of Reaction for Organic Synthesis)
- TOSCA (Topological Synthesis design by Computer Application)
- COMPASS (COMPUter-ASSisted organic Synthesis)
- CAMEO (Computer Assisted Mechanistic Evaluation of Organic reactions)
- ARChem (использование совокупности «правил»)

Возможности компьютерного синтеза, интегрированные в базы данных

REAXYS Setting up AutoPlan

Substances (Grid) Substances (Table) Citations

Limit to Exclude Output Print Zoom In Zoom out Hide Sort by No of References

Structure Structure/Compound Data

Synthesize | Show Details

- Manually
- by AutoPlan
- by AutoPlan (with options)

Manually
by AutoPlan
by AutoPlan (with options)

REAXYS Setting up AutoPlan

AutoPlan options

No of plans to create 10

Max. alternative branches 10

Max. steps 10

All starting materials commercially available

Default yield for reactions without a given yield: 50 %

Behavior of Synthesize Hyperlink

- On click offer option to select between simple search and AutoPlan
- On click only search for the reactions of the given compound as product
- On click start AutoPlan with the given compound

Hits per page Show 9 results per page

Query Results **Synthesis Plans** History Report My Alerts My Settings Help

Synthesis 1 Synthesis 2 Synthesis 3 Synthesis 4 Synthesis 5 Synthesis 6 Synthesis 7 Synthesis 8

New Undo Open Save Rename Duplicate Output Print Left Right Top Resize Thumbnail

8 Details 57.00% Add Remove

7 Details

6 Details

5 Details 95.00% Add Remove

4 Details

3 Details

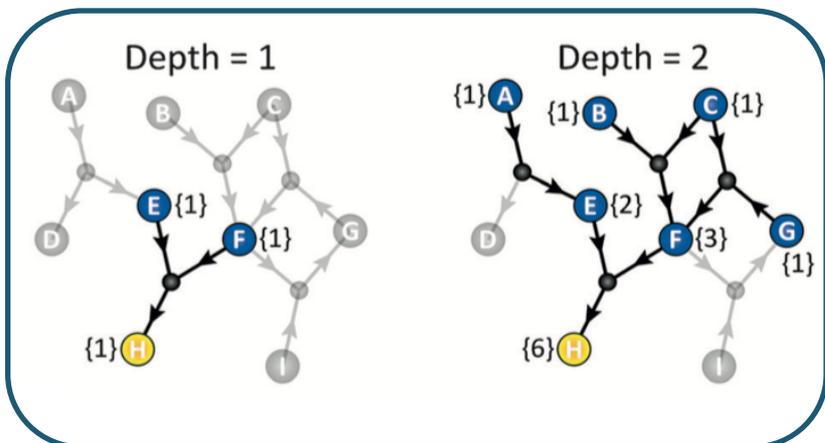
2 Details 81.00% Add Remove

1 Details

19

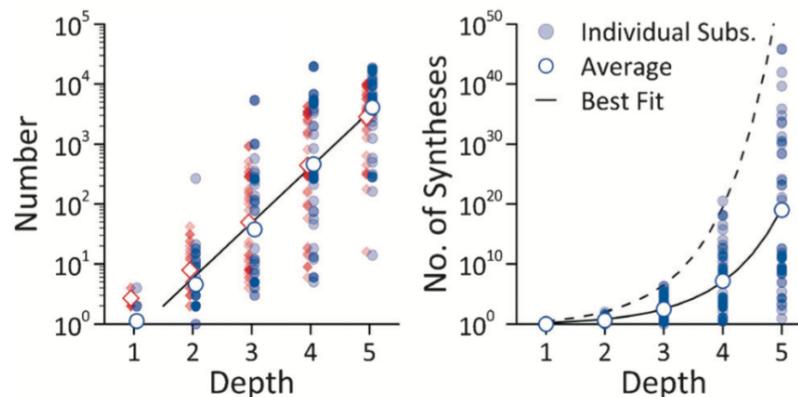
98%	T=20°C; 0.16667h;	Tetrahedron, 2001 , vol. 57, # 27, p. 5851 - 5854 Title/Abstract Full Text View citing articles Show Details
Show Next 20 Details Show All Remaining Details (40)		
14 Reaxys	With bromine T=500°C; Leiten ueber Bimstein;	den Hertog; Wibaut Recueil des Travaux Chimiques des Pays-Bas, 1932 , vol. 51, p. 381,386, 949 Full Text Show Details
	With bromine T=500°C; Leiten ueber Glas;	McElvain; Goese Journal of the American Chemical Society, 1943 , vol. 65, p. 2233,2235 Full Text Show Details
15 Reaxys	With pyridine; sodium carbonate in hydrogenchloride; water; toluene Show Experimental Procedure	PMC Corporation Patent: US5028731 A1, 1991 ; Title/Abstract Full Text Show Details
16 Reaxys	With hydrogen; Cu/ZnO(Al ₂ O ₃) T=205 - 215°C; P=30003 - 37503.8 Torr; Product distribution / selectivity; Show Experimental Procedure	DALIAN INSTITUTE OF CHEMICAL PHYSICS, CHINESE ACADEMY OF SCIENCES; Patent: W0200714487 A1, 2007 ; Location in patent: Page/Page column 7-10 ; Title/Abstract Full Text Show Details
	With hydrogen; chromium; potassium; zinc	Ilna, G. V.; Denisov, A. V.; Davydov, A. A.; Kurina, L. N. Russian Journal of Physical Chemistry, 1989 , vol. 63, # 1 p. 73 - 75 Zhurnal Fizicheskoi Khimii, 1989 , vol. 63, p. 140 - 145 Title/Abstract Full Text Show Details

Компьютерный синтез / Дизайн эксперимента



Depth – стадийность синтеза

Для 51 синтезируемого соединения



- Индивидуальные реакции
- Соединения, значимые для синтеза

SCUBIDOO (Screenable Chemical Universe Based on Intuitive Data OrganizatiOn): DATABASE OF COMPUTATIONALLY GENERATED SYNTHETIC TRACTABLE COMPOUNDS

~18000
«строительных
блоков»

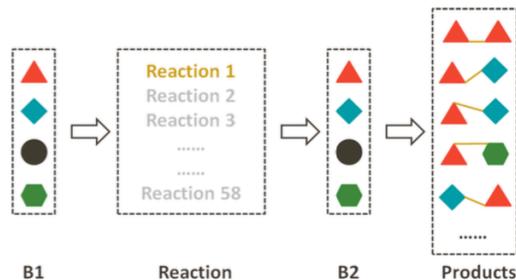
58 наиболее
распространенных
в медицинской
химии реакций

21 миллион новых
соединений

Предварительная фильтрация:

- ❖ Удаление дубликатов
- ❖ Удаление противоионов
- ❖ Контроль «молекулярной сложности»:

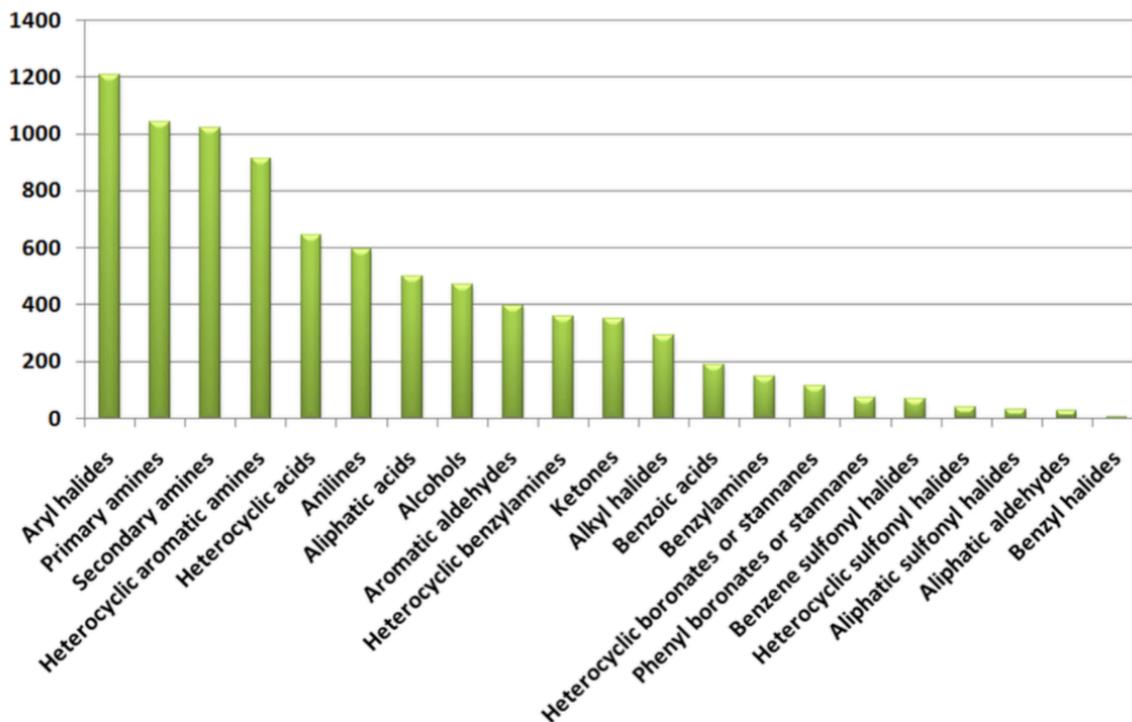
- $MW \leq 250$ Da. (для продуктов с большей вероятностью $MW \leq 500$ Da).
- Количество конформационных связей ≤ 2 . (для продуктов $\leq 6 \Rightarrow$ возможность применения методов молекулярного докинга) Количество хиральных центров ≤ 1 (для продуктов $\leq 3 \Rightarrow$ упрощает синтез)



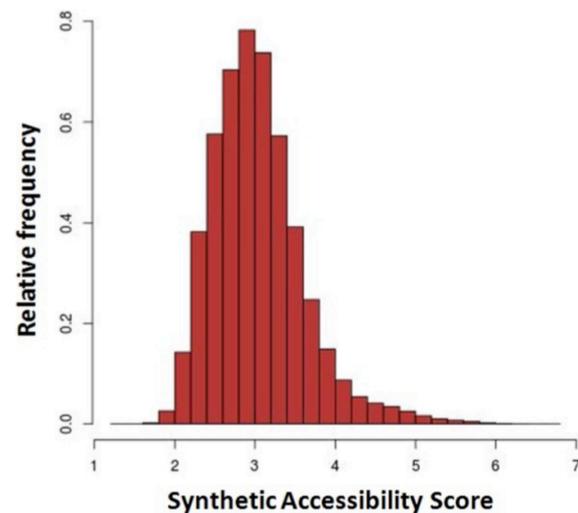
Для каждого соединения в базе данных представлена информация по методу синтеза, потенциальным побочным реакциям, и альтернативным способам получения

SCUBIDOO (Screenable Chemical Universe Based on Intuitive Data OrganizatiOn): DATABASE OF COMPUTATIONALLY GENERATED SYNTHETIC TRACTABLE COMPOUNDS

Классы реагентов («строительные блоки»)



Синтетическая доступность



SCUBIDOO: A Large yet Screenable and Easily Searchable Database of Computationally Created Chemical Compounds Optimized toward High Likelihood of Synthetic Tractability

F. Chevillard et al *Journal of Chemical Information and Modeling* 2015 55 (9), 1824-1835 DOI: 10.1021/acs.jcim.5b00203



Вопросы?