ПРЕДСКАЗАТЕЛЬНАЯ СИСТЕМА ДЛЯ ОПТИМИЗАЦИИ УСЛОВИЙ РЕАКЦИЙ СНЯТИЯ ЗАЩИТНЫХ ГРУПП

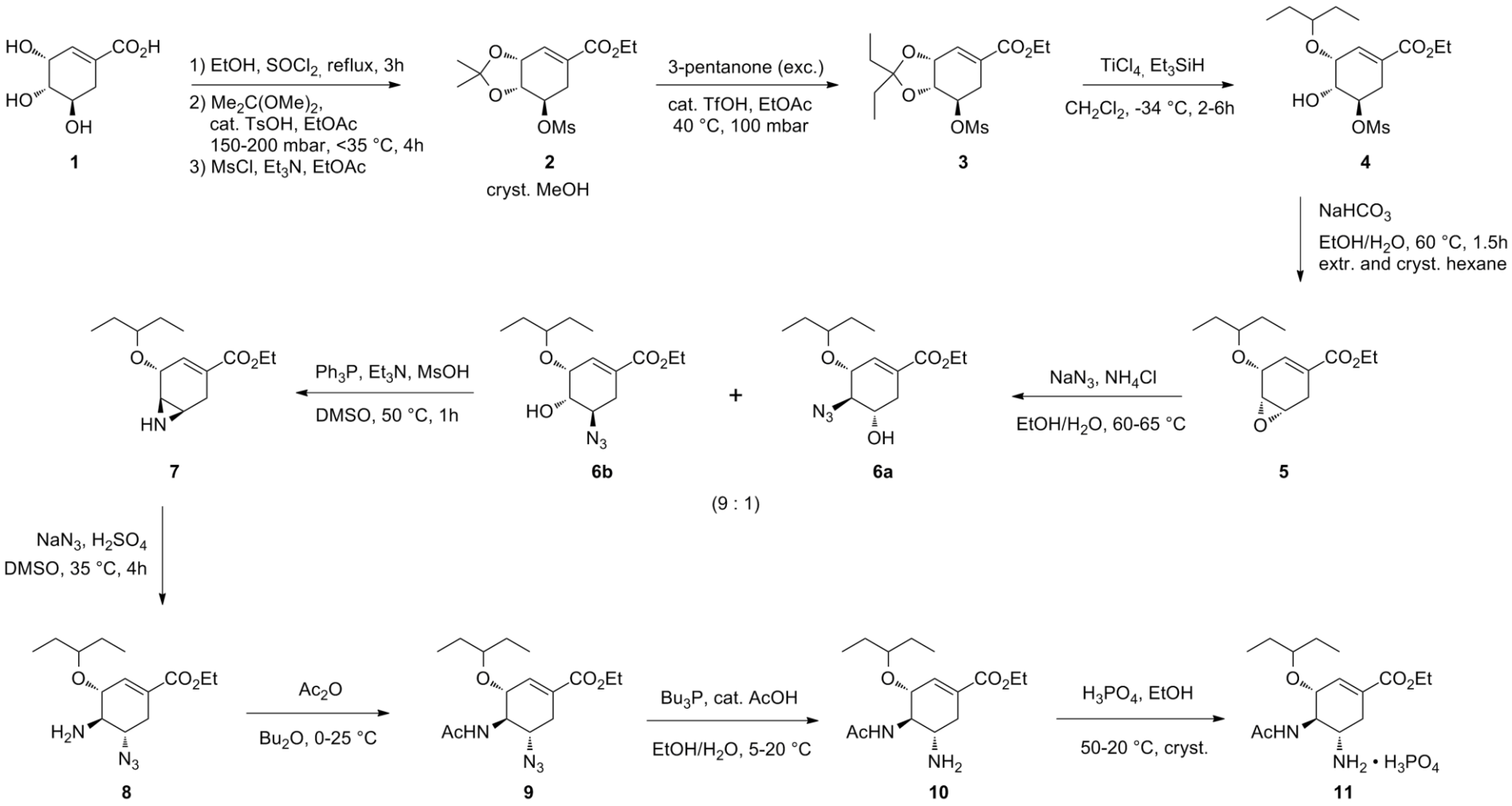
Т. Маджидов, А. Лин, В. Афонина, Р. Нугманов, И. Антипин - КФУ

O. Klimchuk, A. Varnek – University of Strasbourg

Затраты на производство лекарств

Steven M. Paul, *Nat. Review Drug Discovery* March Vol. 9 **2010**

Синтез осельтамивира



Важно знать не только ЧТО синтезировать, но и КАК синтезировать

«Большие данные» о химических реакциях

REAXYS® > 40 млн. реакций

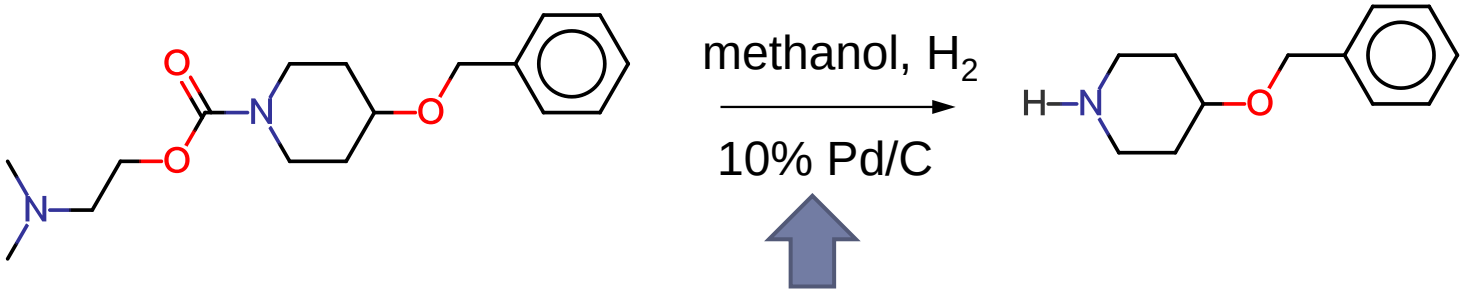
SciFinder® The choice for chemistry research.™ > 76 млн. реакций

- Около 10^8 реакций аннотировано в базах данных
- Автоматически обрабатывать информацию о химических реакциях сложно.
  Большинство попыток получить полезную информацию или модели для химических реакций, основаны на анализе небольших, собранных вручную наборов данных.

Struebing, H. et al. *Nat. Chem.* **2013**, *5* (11), 952.
Marcou, G. et al. *J. Chem. Inf. Model.* **2015**, *55* (2), 239.
Madzhidov, T.I. et al. *J. Struct. Chem.* **2015**, *56*,1227
Nugmanov, R.I. et al. *J. Struct. Chem.* **2015**, *55*,1026

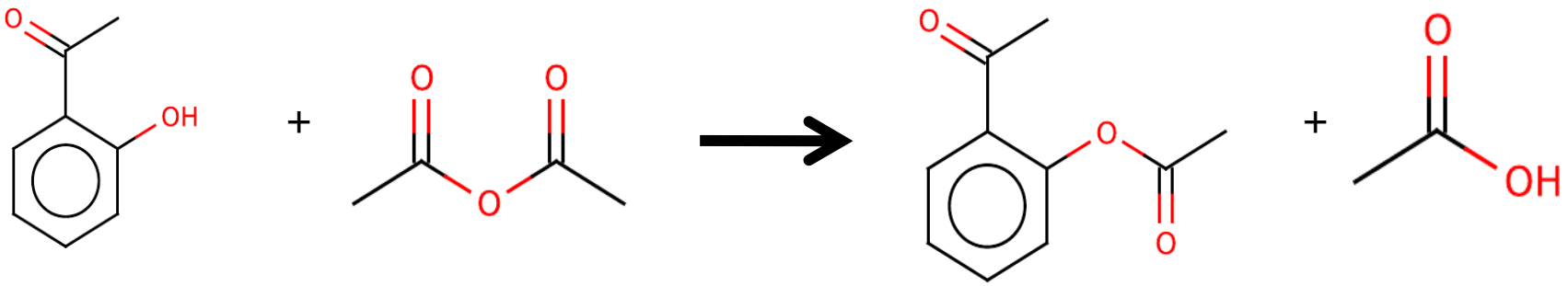
«Большие данные» о химических реакциях: Шум

- Большинство реакций стехиометрически неуравновешены
- Для некоторых реакций указаны конкурирующие продукты
- Информация о выходах, условиях или иных важных параметрах может отсутствовать или некачественно извлеячена из первичной литературы
- Нет стандартных имен веществ, растворителей, катализаторов



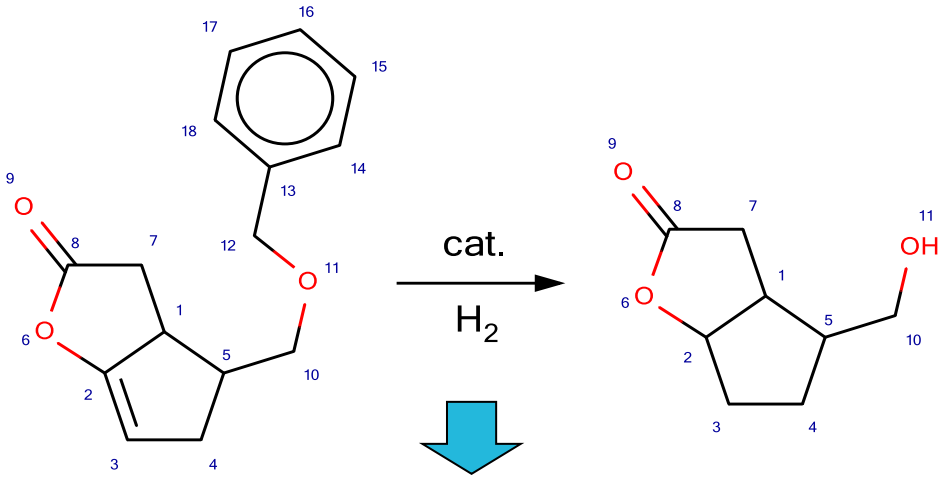
- 10% Pd on carbon(eggshell)
- 10% palladium on charcoal
- 10% palladium/C

«Большие данные» о химических реакциях: Сложность

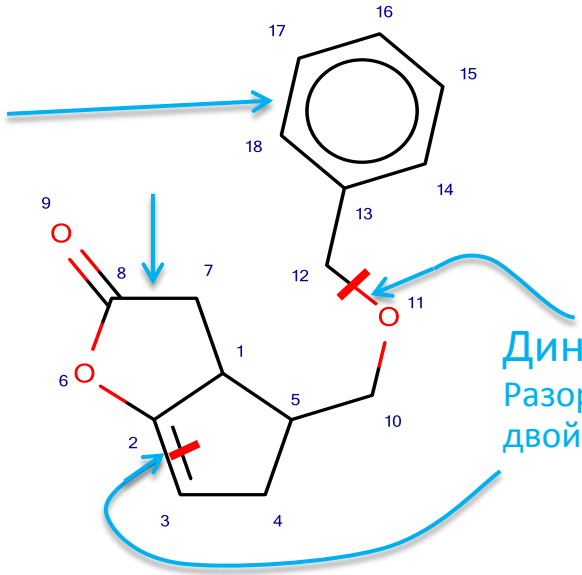


- Соединения двух типов: реагенты и продукты
- Реакции со множеством стадий
- Зависимость выходов от условий (катализаторы, растворители, и др.)

Конденсированный граф реакции (CGR)



Обычные химические связи:
одинарные, двойные
ароматические, …

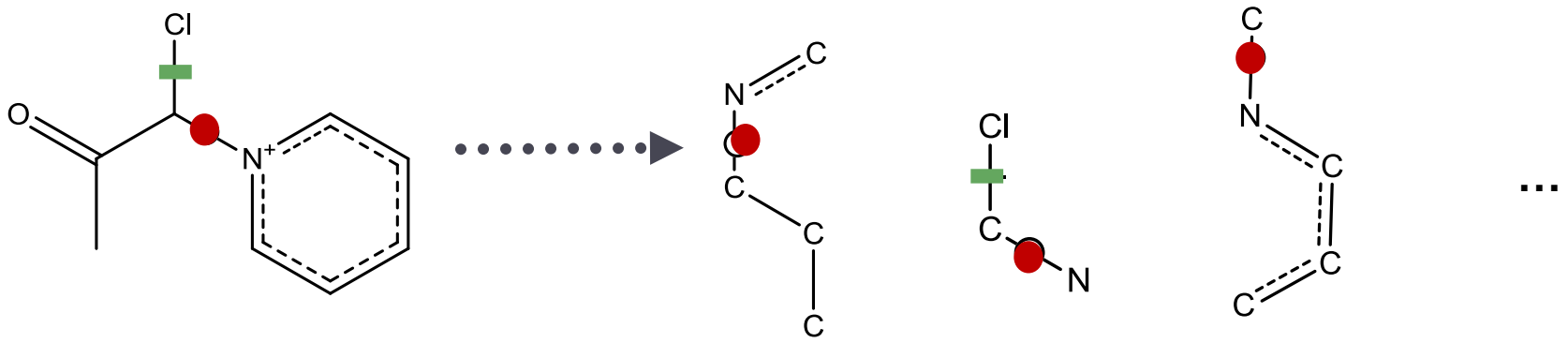


Динамические связи:
Разорванная одинарная, превращение двойной в одинарную, …

Дескрипторы химических связей

Конденсированный граф реакции

Фрагментные дескрипторы ISIDA



2	1	2	…

Реакции могут быть закодированы в виде строки дескрипторов которые могут быть использованы в моделировании «структура-свойство», поиске по схожести, кластеризации объектов, и др.

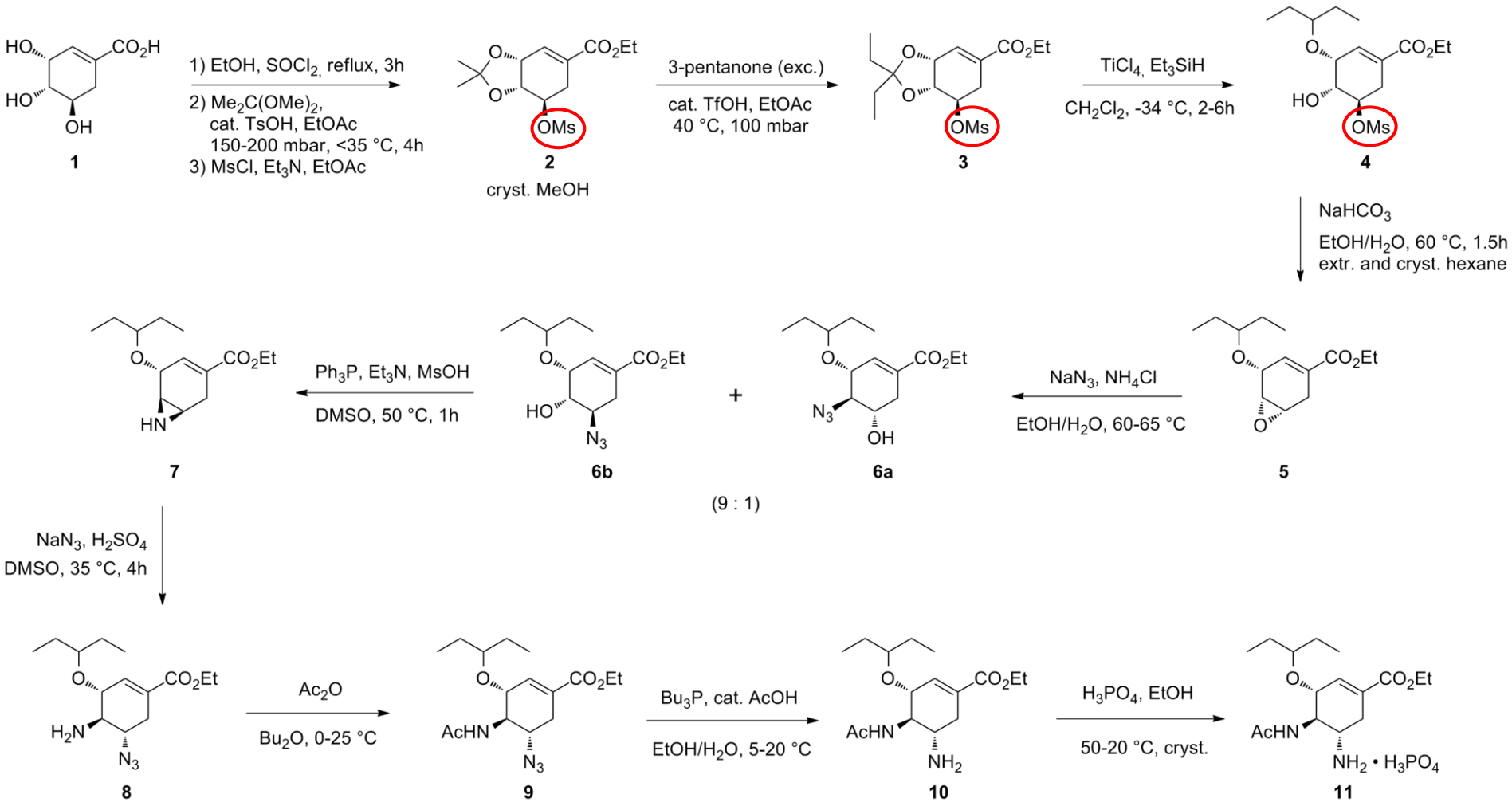
A. Varnek In: "*Chemoinformatics and Computational Chemical Biology*", J. Bajorath, Ed., Springer, 2010

Цели

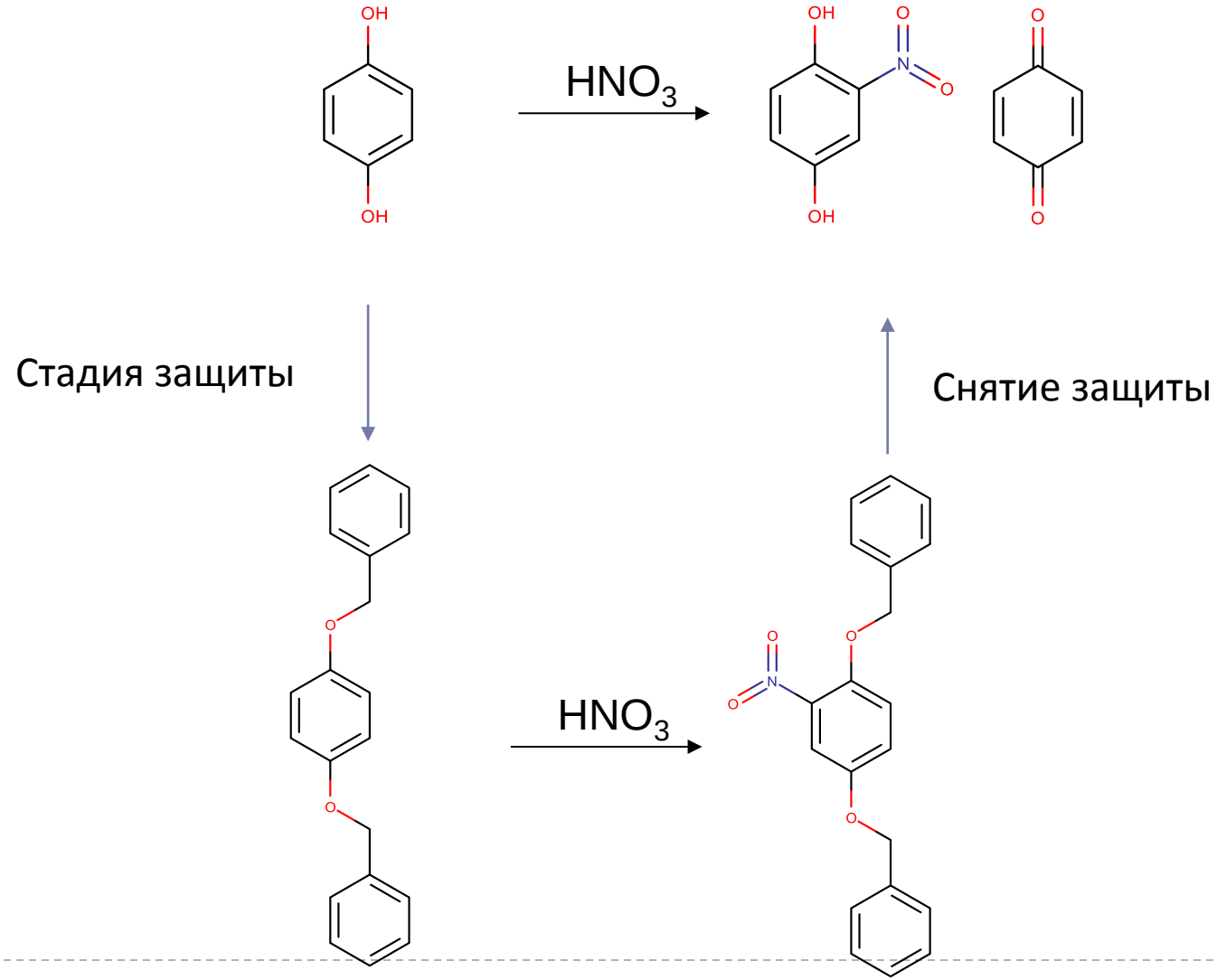
Выходы и селективность химических реакций зависит от условий (катализатор, растворитель, добавки, температура, и др.)

- Целью работы явилась разработка подхода, способного предсказывать оптимальные условия проведения химических реакций.
- Разработанный подход был применен для химии защитных групп.

Синтез осельтамивира

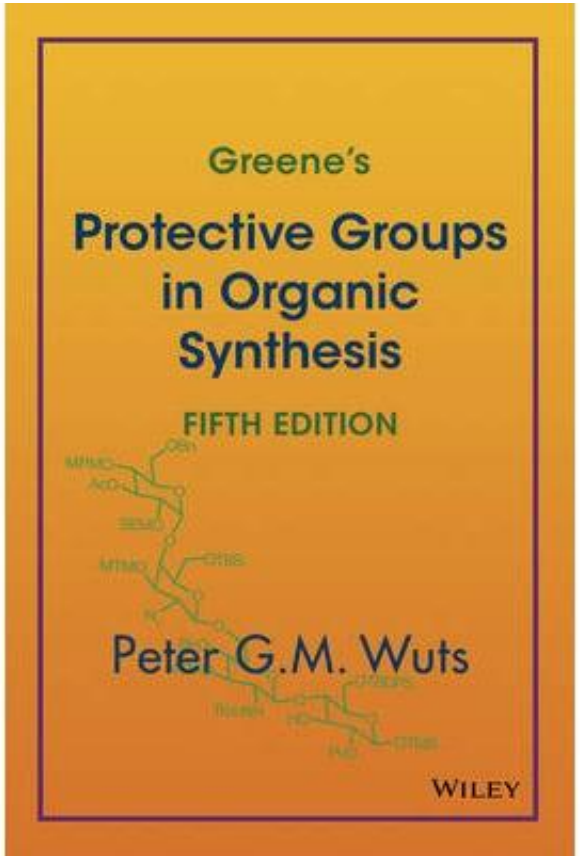


Защитные группы



Pain, Ch., L. Pain. *How to choose the path of organic substance synthesis*, 1973

Квинтэссенция знаний в химии защитных групп

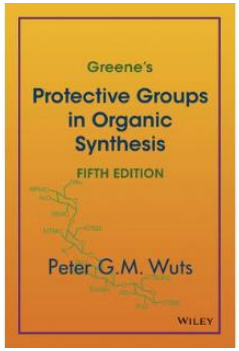


Theodora W. Greene
(1931-2005)

1054 защитные группы (PG)

11249 статей

Green's Reactivity Charts



	H_2/Raney (Ni)	H_2/ Pt, pH 2-4	H_2/Pd	H_2/Lindlar	H_2/Rh
PG	Catalytic Reduction				
Me	L	L	L	L	L
MOM	L	M	L	L	L
THP	L	L	L	L	L
t-Butyl	L	L	L	L	L
Bn	H	H	H	L	L
TPM	H	H	H	L	L

Катализатор

Метод снятия защиты

Наблюдения

H – уходящая PG; ***L*** – остающаяся PG; ***M*** – нельзя сделать четкого заключения

Недостатки книги Greene

- *Reactivity Charts* получаются ручным анализом относительно небольшого объема данных и, по этой причине, могут иметь место некоторые ошибки или предвзятость
- Не ясно в соответствии с какими количественными критериями – выход, % снятых и оставшихся групп – были присвоены метки реакционной способности групп (*H, L или M*);

- ***Reactivity Charts* не принимают во внимание окружение защитной группы и изменения по этой причине реакционной способности защитных групп**

Анализ реакционной способности защитных групп

1. Статистический анализ реакционной способности защитных групп и сравнение полученных результатов с Greene’s Reactivity Charts.

2. Разработка альтернативного подхода для анализа реакционной способности защитных групп как функции их химического окружения

Данные

Набор из **142111** ***реакций каталитического гидрирования,*** извлеченные из базы данных ***Reaxys*** (2012)

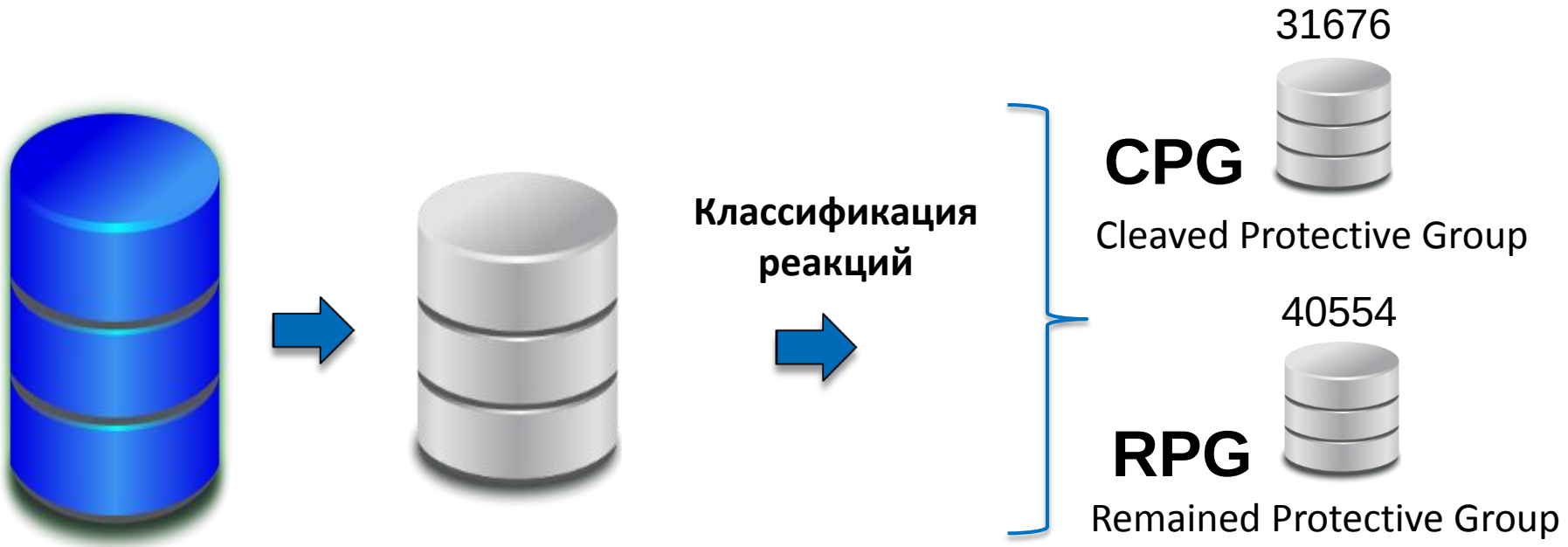
Катализатор или реагент	T	время	P	выход	растворитель	**Вся**
95.6	45.1	57.6	33.5	67.8	83.7	**10.9**

% реакций для которых известна информация о температуре (T), давлении (P), времени проведения (t), выходе, растворителе, катализаторе или реагенте, либо все указанные параметры

Процедура обработки данных

72230 реакций для 86 защитных групп

Классификация реакций



Сырые данные

142 111 реакций
>271 000 условий

Подготовленные данные

72 230 реакций

Отобранные данные

72 230 реакций

Оценка частоты снятия защитной группы

- ***C**leavage **R**ate* $CR = \frac{CPG}{(CPG+RPG)} * 100, \%$

Если CR ≥ 80% → Защитная группа легко снимается (“*H*”)

Если CR ≤ 20% → Защитная группа не снимается (“*L*”)

Иначе → Нельзя сделать четкого вывода (“*M*”)

Сравнение с книгой Greene (защита спиртов)

Защитная группа	Raney (Ni)	Pt, pH 2-4	Pd/C	Lindlar	Rh/C or Rh/Al2O3
Me	L	L	L	L	L
MOM	L	M	L	L	L
MEM	L	M	L	L	L
Cy	L	L	L	L	L
t-Bu	L	L	L	L	L
Bn	H	H	H	L	H
TBDMS	L	H	L	L	L
Ac	L	M	L	L	L
piv	L	L	L	L	L
Bz	L	L	L	L	L
Ms	R	L	L	L	L

51% - Согласие

7% - Противоречие

42% - Недостаточно данных (≤ 10 реакций)

- Нет данных в Reaxys DB

Больше, чем Green's Reactivity Charts

Влияние добавок

PG	Greene's annotation	Pd/C		
		pure	acid	poison
Bz phenol	H	92	94 ↑	73 ↓
Bz aliph.	H	86	93 ↑	21 ↓
TBDMS aliph.	L	3	29 ↑	4 ↑
Bz carbamate	H	93	91 ↓	60 ↓

Больше, чем Green's Reactivity Charts

Влияние растворителя

PG	Greene's annotation	Pd/C		
		any	polar	non-polar
Bz phenol	H	92	92 ↑	72 ↓
Bz aliph.	H	84	85 ↑	67 ↓
TBDMS aliph.	L	4	4 ↑	0 ↓
Bz carbamate	H	93	94↑	25 ↓

Оценка реакционной способности защитной группы на основе принципа схожести

Главная концепция:

Похожие реакции идут в схожих условиях

Реализация:

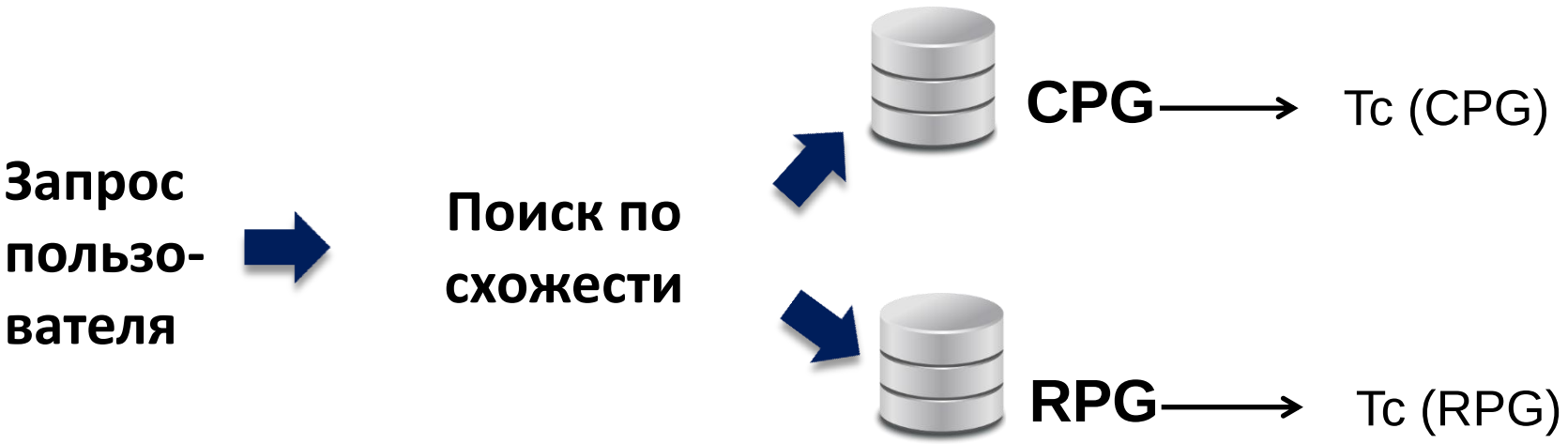
Для данной пользователем реакции, программа проводит поиск наиболее похожей реакции в базе данных и возвращает ее условия проведения

Оценка схожести:

Для Конденсированного графа реакции рассчитывается с помощью индекса Танимото

$$Tc = \frac{c}{a + b - c}$$

Оценка реакционной способности защитной группы на основе принципа схожести



$\Delta Tc = Tc(CPG) - Tc(RPG)$

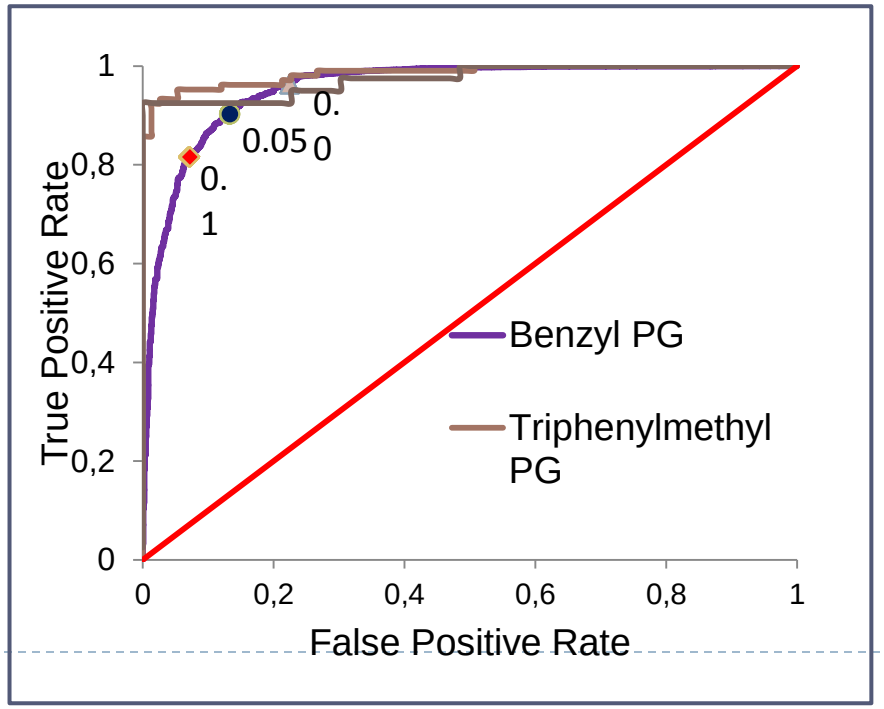
If $\Delta Tc \geq T_0$ ➡ Группа снимается

$\Delta Tc \leq -T_0$ ➡ Группа остается

Проверка принципа схожести для реакций

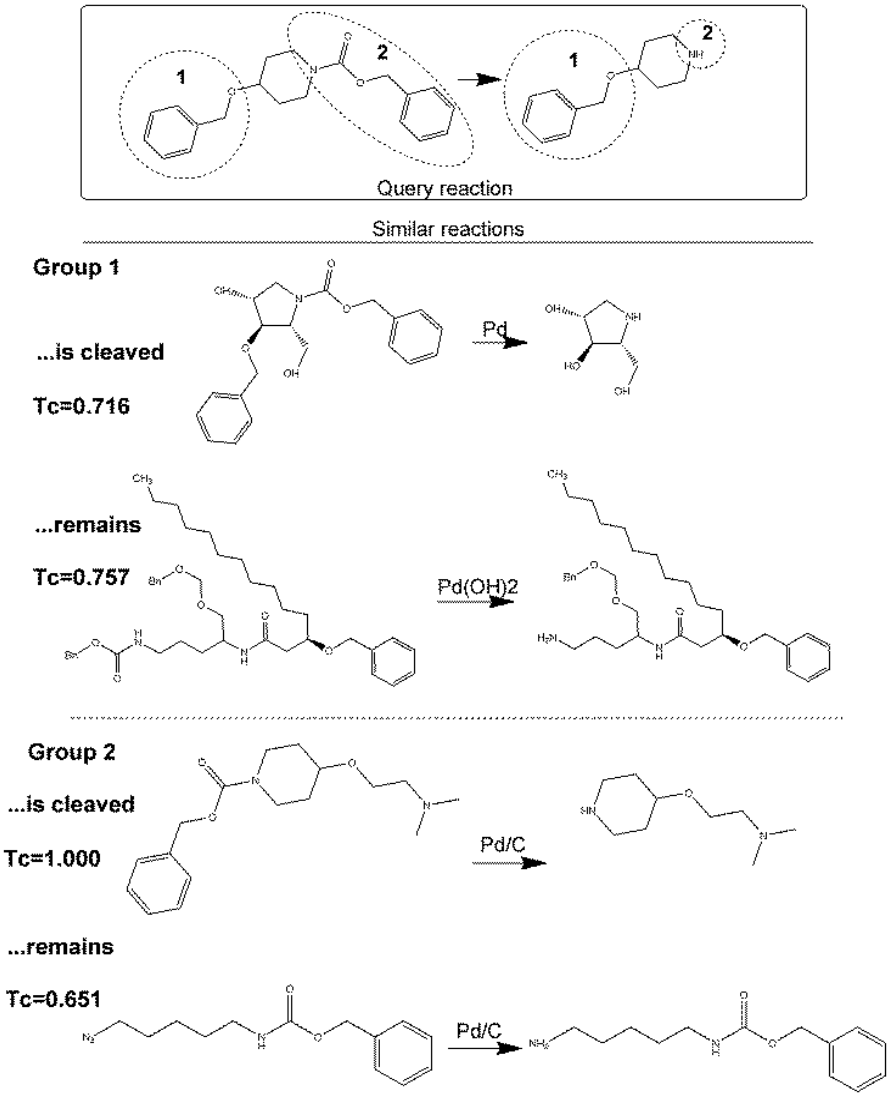
В перекрестном контроле по одному (Leave One Out cross-validation) для 27356 реакций (>40 реакций для данного катализатора)

- ROC AUC = 0.94 – 0.98
- Balanced Accuracy = 0.86 – 0.96 at T_0=0.05

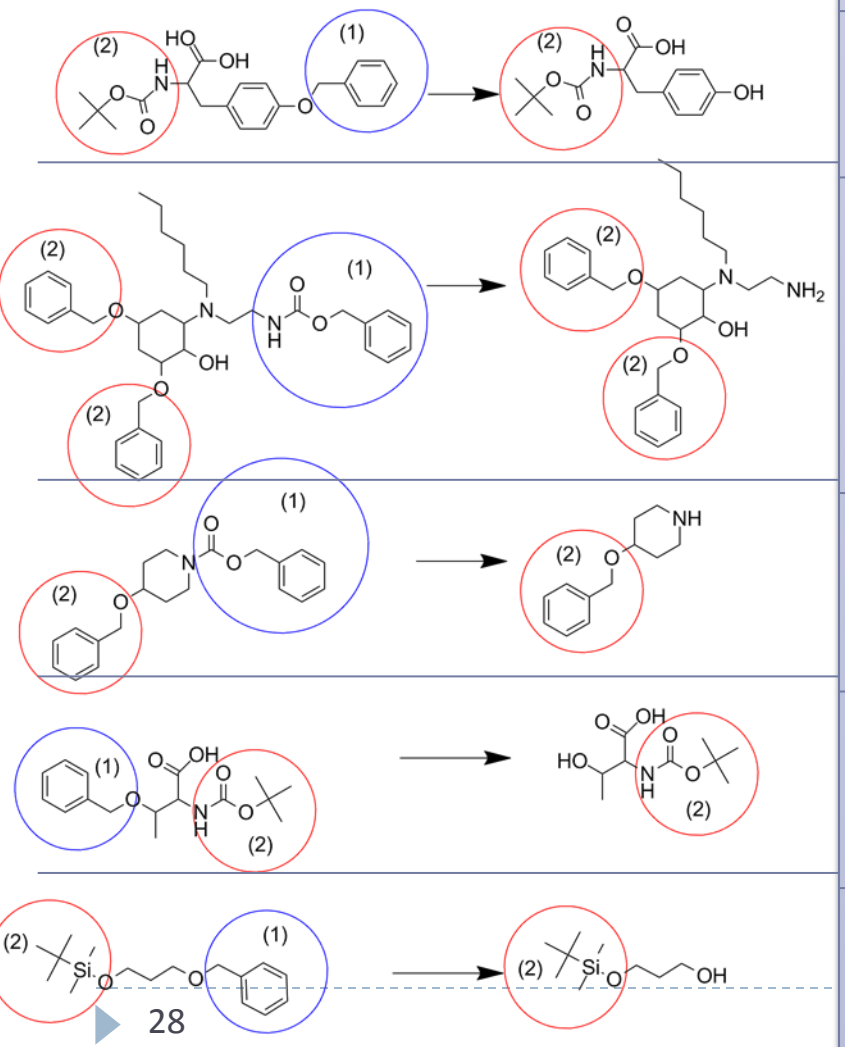


Независимая валидация

- 7 субстратов, содержащих **одну** защитную группу **- 5 корректно предсказанных**
- 5 субстратов содержащих **две** защитные группы - **все корректно предсказаны**



Независимая валидация



Experimental conditions	Group	Greene's Reactivity Charts	Expert system recommendation
Pd/C, Methanol	(1)	to be cleaved (H)	Pd-catalyst [Pd/C]
	(2)	remain (L)	
Pd/C, Methanol	(1)	to be cleaved (H)	Pd-catalyst [Pd/C]
	(2)	to be cleaved (H)	
Pd/C, Methanol	(1)	to be cleaved (H)	Pd-catalyst [Pd/C] Ni-catalyst [Raney Ni]
	(2)	to be cleaved (H)	
Pd/C, Ethanol	(1)	to be cleaved (H)	Pd-catalyst [Pd/C]
	(2)	remain (L)	
Pd/C, Ethyl acetate	(1)	to be cleaved (H)	Pd-catalyst [Pd/C] Ni-catalyst [Raney Ni] Lindlar [Lindlar]
	(2)	remain (L)	

Выводы

- Подход Конденсированного графа реакции предоставляет возможности для создания эффективной техники обработки реакционной информации, который может быть применен для стехиометрически несбалансированных реакций

- Анализ реакционной способности защитных групп на большом наборе данных выявил несогласия с вручную проведенным анализом, приведенном в Green's Reactivity Charts.

- Подход для оценки реакционной способности защитных групп на основе принципа схожести является достататочно точным

Статья в J. Chem. Inf. Model.

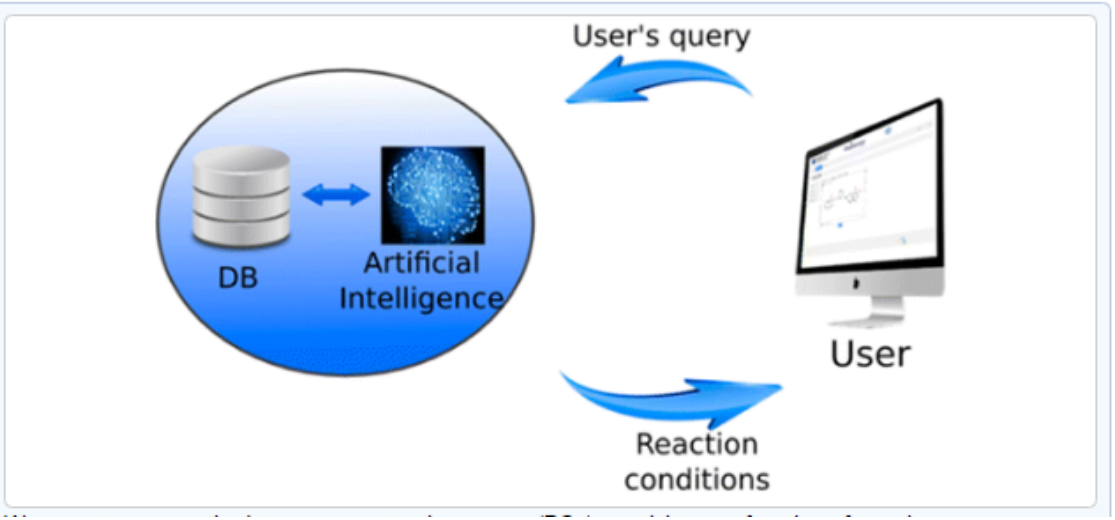
Automatized Assessment of Protective Group Reactivity: A Step Toward Big Reaction Data Analysis

Arkadii I. Lin[†‡], Timur I. Madzhidov[†], Olga Klimchuk[‡], Ramil I. Nugmanov[†], Igor S. Antipin[†], and Alexandre Varnek[*†‡]
[†] Laboratory of Chemoinformatics and Molecular Modeling, Department of Organic Chemistry, A.M. Butlerov Institute of Chemistry, Kazan Federal University, Kremlyovskaya Str. 18, Kazan, Russia, 420008
[‡] Laboratory of Chemoinformatics, Faculty of Chemistry, University of Strasbourg, rue Blaise Pascal 1, Strasbourg, France, 67000

J. Chem. Inf. Model., Article ASAP
DOI: 10.1021/acs.jcim.6b00319
Publication Date (Web): October 26, 2016

*E-mail: varnek@unistra.fr.

Abstract



Article Options

ACS ActiveView PDF
Hi-Res Print, Annotate, Reference QuickView
PDF (3108 KB)
PDF w/ Links (507 KB)
Full Text HTML

Abstract
Supporting Info
Figures
References

Add to ACS ChemWorx

Add to Favorites
Download Citation
Email a Colleague
Order Reprints
Rights & Permissions
Citation Alerts

Metrics

Received 4 June 2016
Published online 26 October 2016

SciFinder A CAS SOLUTION Sign in

Благодарности

Российский научный фонд

Arkadii Lin (KFU, UniStra)

Ramil Nugmanov (KFU)

Olga Klimchuk (UniStra)

Reaxys®

ChemAxon

Prof. Igor Antipin (Kazan)

Prof. Alexandre Varnek (UniStra)

Acknowledgements:
Gilles Marcou (UniStra)
Dragos Horvath (UniStra)
Timur Gimadiev (KFU, UniStra)
Pavel Sidorov (UniStra)
Sergey Neklyudov (KFU)

Kazan Summer School on Chemoinformatics

July 5-7, 2017
Kazan, Russia
cimm.kpfu.ru/kssci2017

Confirmed speakers:

- Alex TROPSHA (UNC, USA)
- Peter ERTL (Novartis, Switzerland)
- Alexandre VARNEK (UniStra, France)
- Valery TKACHENKO (ChemDataSoftware, USA)
- Hanoch SENDEROWITZ (Bar-Ilan University, Israel)
- Vladimir POROIKOV (IBMC, Russia)
- Igor BASKIN (MSU, Russia)
- Pavel POLISHCHUK (Olomouc University, Czech Rep.)
- Pavel YAKOVLEV (BIOCAD, Russia)
- Dragos HORVATH (CNRS, France)

… ...