Группа Хемоинформатики (молекулярной информатики)

В.н.с., д.ф.-м.н. Баскин Игорь Иосифович

С.н.с., к.х.н. Жохова Нелли Ибрагимовна

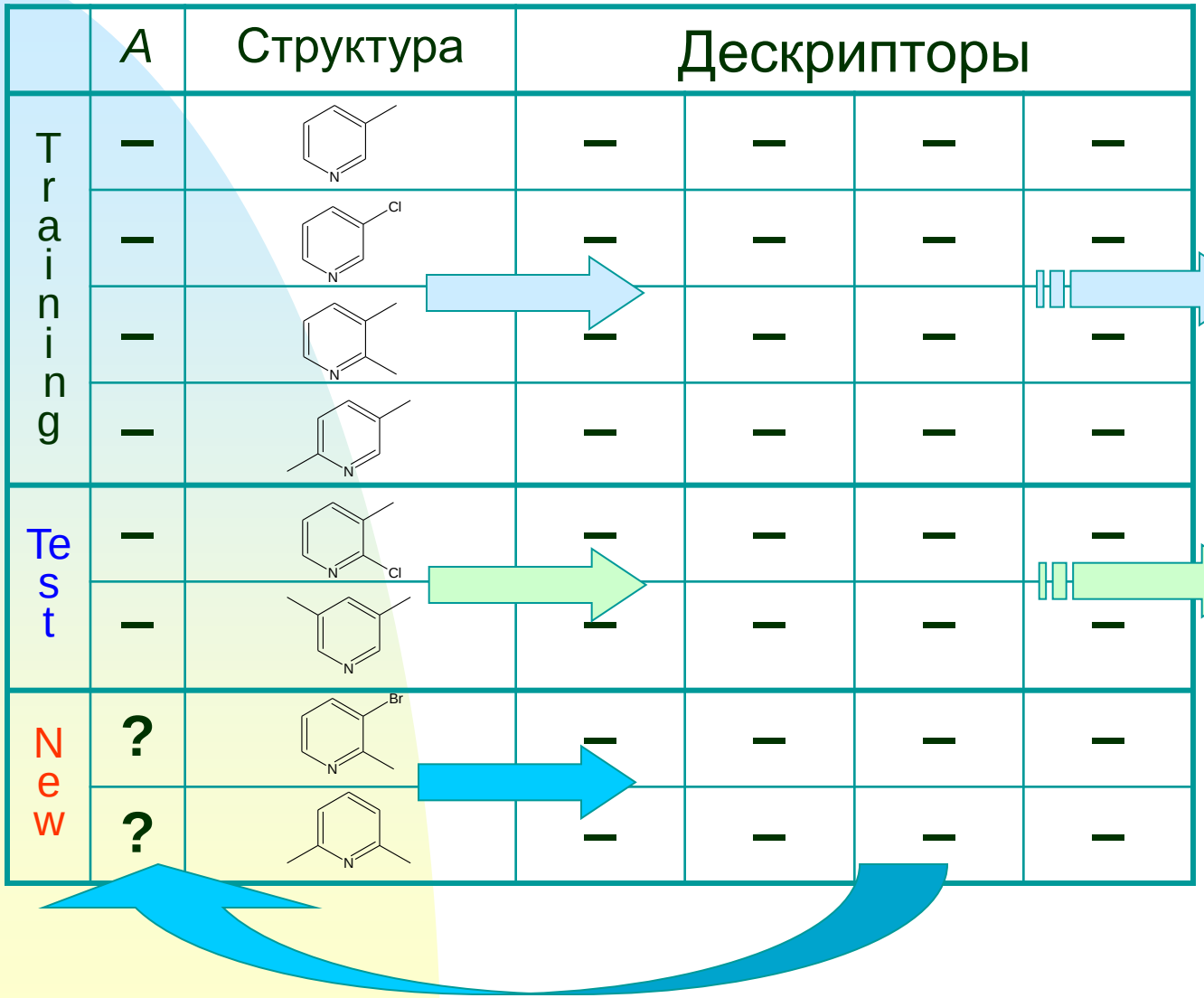
План доклада

- Что такое хемоинформатика (молекулярная информатика)
- Наш вклад
- Текущие исследования
- Возможные будущие направления работ

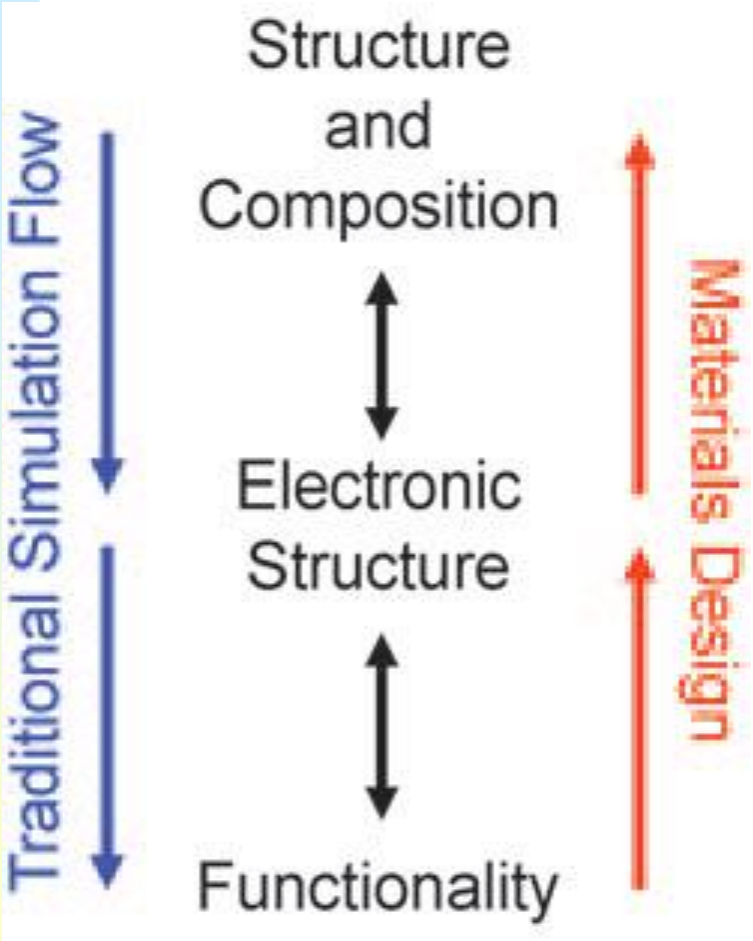
План доклада

- Что такое хемоинформатика (молекулярная информатика)
- Наш вклад
- Текущие исследования
- Возможные будущие направления работ

QSAR/QSPR: количественные соотношения структура-свойство и структура-активность



Молекулярный дизайн



Виртуальный скрининг

«Обратный» QSPR

Хемоинформатика (молекулярная информатика)

	Квантовая химия	Основанное на силовых полях молекулярное моделирование	Хемоинформатика (молекулярная информатика)
Молекуляр-ная модель	Электроны и ядра	Атомы и связи	**Химические объекты (графы)**
Механизм вывода	Дедуктивный	Дедуктивный > индуктивный	**Индуктивный**
Применя-ется к	Малочисленные объекты	Многочисленные объекты	**Базы данных объектов**
Основная концепция	Дуализм частица / волна	Классическая механика	**Химическое пространство**
Основные подходы	Уравнение Шредингера, HF, MO, DFT	Молекулярная механика и динамика, Монте-Карло, FEP	**Методы машинного обучения**
Основные задачи	Интерпретация >> прогнозирование	Интерпретация > прогнозирование	**Прогнозирование**
Набор свойств	Очень ограниченный	Ограниченный	**Потенциально любые свойства**
Виртуальный скрининг	Нет	Нет	**Да**

План доклада

- Что такое хемоинформатика (молекулярная информатика)
- **Наш вклад**
- Текущие исследования
- Возможные будущие направления работ

Основные научные достижения группы

Методология хемоинформатики

- Основы хемоинформатики как науки
- «Обратная» задача для моделей QSPR
- Фрагментные дескрипторы
- Нейросетевой подход к прогнозированию свойств химических соединений
- Методы виртуального скрининга на основе одноклассовых моделей
- Методы визуализации химического пространства на основе Байесовской статистики и дифференциальной геометрии
- Методы многозадачного обучения и многоуровневого моделирования в задаче «структура-свойство»
- Методы построение моделей QSPR для многокомпонентных смесей при варьируемых внешних условиях
- Метод непрерывных молекулярных полей

Разработка моделей QSPR

- Многочисленные физико-химические свойства химических соединений
- Спектры биологической активности

Разработка программного обеспечения

- NASAWIN – пакет программ для нейросетевого моделирования
- FRAGMENT – блок расчета фрагментных дескрипторов

Теоретические основы хемоинформатики (молекулярной информатики)

Chemoinformatics as a Theoretical Chemistry Discipline

Alexandre Varnek*[a] and Igor I. Baskin[b]

Mol. Inf. 2011, 30, 20 – 32

the most accessed in 2011

Machine Learning Methods for Property Prediction in Chemoinformatics: *Quo Vadis*?

Alexandre Varnek *† and Igor Baskin †‡

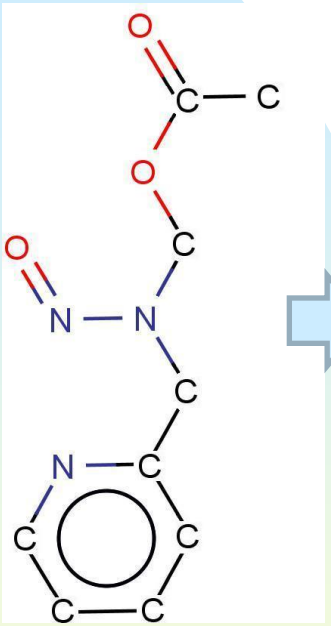
the most accessed in 2012

J. Chem. Inf. Comput. Model. 2012, 52, 1413-1437

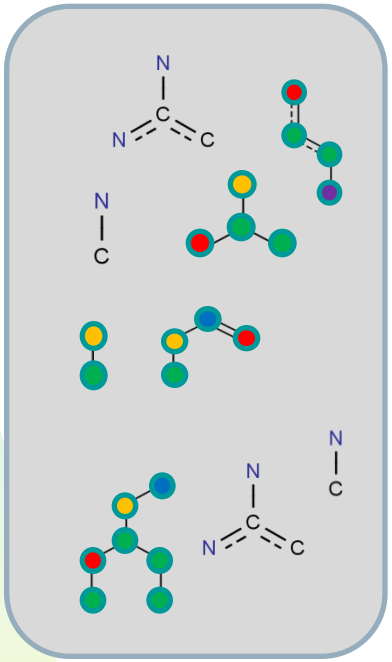
А,А.Варнек, И.И.Баскин, Т.И.Маджидов «Основы хемоинформатики»

Фрагментные дескрипторы

Молекулярный граф



Молекулярные фрагменты



Вектор фрагментных дескрипторов

Фрагмент	Число
Fr_1	N_1
Fr_2	N_2
….	…
Fr_i	N_i
….	…

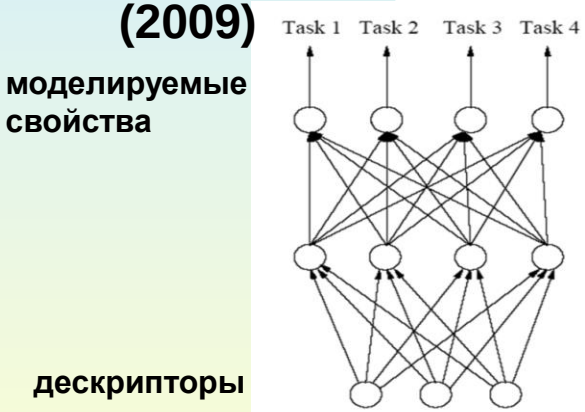
- ✓**иерархическая схема генерации фрагментных дескрипторов**
- ✓**включение атомных свойств**

Математическая основа: **теоремы о базисе инвариантов помеченных графов**

Нейросетевое моделирование

***Наш вклад*:**

- **Нейросетевые модели на основе фрагментных дескрипторов (1993)**
- **Нейросетевые модели на молекулярных графах (1993)**
- **Интерпретация нейросетевых моделей (2003)**
- **Нейросетевое моделирование многокомпонентных систем (2007)**
- **«Индуктивный перенос знаний» между нейросетевыми моделями (2009)**



«Индуктивный перенос знаний»

Одновременное построение нейросетевых моделей для нескольких свойств ведет к повышению прогнозирующей способности каждой из них

Программный комплекс NASAWIN: Разработка моделей для прогнозирования более 60 различных свойств химических соединений

Сравнение с лучшими опубликованными результатами (среднеквадратические ошибки)

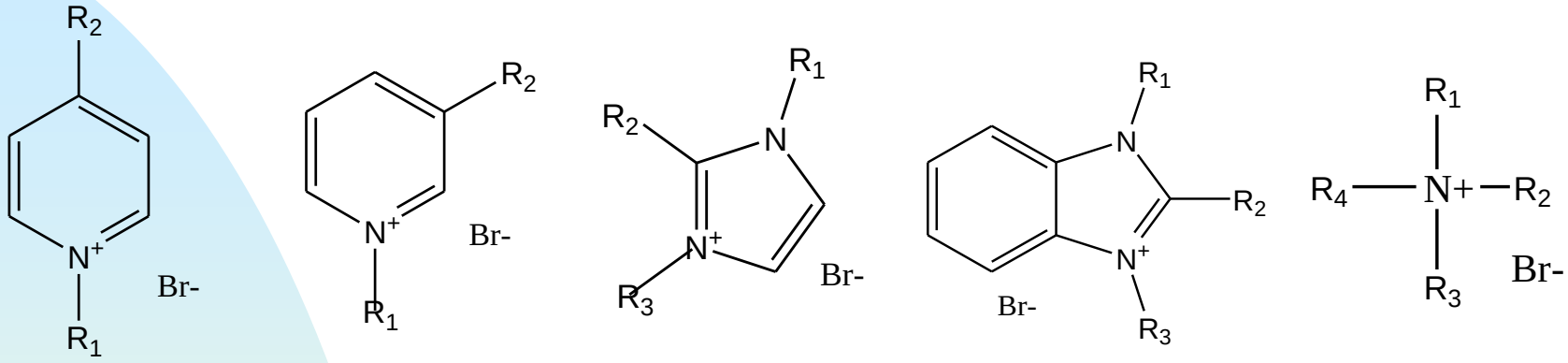
Свойство	Наши результаты	Опубликовано в литературе	Авторы
Плотность жидкостей, ст. усл., г/см^3	0.043	0.046	Katritzky et al., 2000
Вязкость жидкостей, ст. усл., lg(Па·с)	0.177	0.22	Katritzky et al., 2000
Давление насыщенных паров, 25 ^{o}C, lg(Па)	0.158	0.209	Jurs et al, 1999
Температура кипения, 1 атм., ^{o}C	16.9	19.4	Tetteh et al, 1999

Прогнозирование свойств на основе ФД и различных статистических методов

(Программы: NASAWIN, LibSVM, WEKA)

Свойство	Выборка	Ссылка
Сродство красителей к целлюлозному волокну, $-\Delta\mu^0$ (кДж·моль-1)	79 органических красителей	*ЖПХ, 2005*
Энтальпия испарения	52 соединения	ЖФХ, 2007
Стабильность комплексов органических соединений с *β*-ЦД,	218 комплексов (H_2O, 25 °C)	*Вестн. МГУ, 2007*
Стабильность комплексов включения гость-хозяин катионов Na+ и K+ с Краун-эфирами, *ΔlogK*	223 комплекса	Москва, 2007
Энтальпия образования, ΔH^0_f (ккал/моль)	163 нитросоединения	*Colorado, 2008*
Чувствительность к искре, E	73 нитросоединения	*Pardubice, 2008*

Прогнозирование температуры плавления ионных жидкостей



	PYR	IMZ	QUAT	FULL
ИНС	26.2	32.4	30.3	31.5
БПМЛР	34.8	36.2	36.1	33.7

Расчет энтальпии сублимации методом SVR с применением фрагментного подхода

ЭНТАЛЬПИЯ СУБЛИМАЦИИ –
энтальпия перехода вещества из твердого состояния непосредственно (без плавления) в газообразное.

ПРОБЛЕМЫ –
экологические (перенос органических загрязнителей в атмосфере);
диспергирование красителей;
выцветание материалов и др.

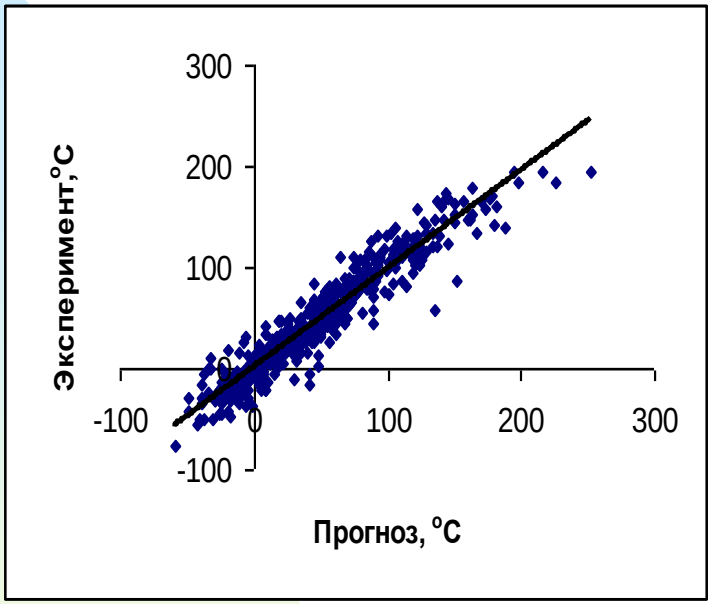
Статистические параметры модели:
(10-кратный скользящий контроль)
Параметр Q^2 = 0.801
Среднеквадратичное отклонение, *RMSE* = 3.33 ккал/моль

Параметры метода SVR:
Nu = 0.72, Log2(c) = 12.6, Log2(g)= -18.7

БАЗА ДАННЫХ –
104 органических соединений различных классов с известной кристаллической структурой

$$q^2 = 1 - \frac{PRESS}{SS} = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i^{pred})^2}{\sum_{i=1}^{n}(y_i - y)^2}$$

Расчет температуры вспышки методом SVR с применением фрагментного подхода



ТЕМПЕРАТУРА ВСПЫШКИ –
Ключевая характеристика горючих свойств органических материалов

ВАЖНА для понимания закономерностей процессов горения

Показывает насколько БЕЗОПАСНО работать с веществом и его транспортировать

БАЗА ДАННЫХ –
515 органических соединений различных классов

Статистические параметры модели:
(10-кратный скользящий контроль)

Параметр Q^2 = 0.907
Среднеквадратичное отклонение,
RMSE = 15.82, ^{0}C

$$q^2 = 1 - \frac{PRESS}{SS} = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i^{pred})^2}{\sum_{i=1}^{n}(y_i - y)^2}$$

Прогнозирование физических свойств аморфных полимеров

Свойство	Только ФД			ФД с добавлением ПФД		
	Q^2_{DCV}	$RMSE_{DCV}$	MAE_{DCV}	Q^2_{DCV}	$RMSE_{DCV}$	MAE_{DCV}
n	**0.782**	**0.033**	**0.021**	**0.872**	**0.026**	**0.015**
T_g	**0.849**	**45.0**	**32.0**	**0.864**	**42.7**	**28.0**
ρ	**0.474**	**0.159**	**0.096**	**0.910**	**0.066**	**0.043**

где: *n* – показатель преломления при 298K; *Tg* – температура стеклования (в градусах Кельвина); *ρ* – плотность в аморфном состоянии (г/см3, 298K).

Жохова Н.И.., Баскин И.И. и др. // ДАН, 2010, Т. 430, № 5, С. 635-638.

Некоторые методологические разработки группы молекулярной информатики/ хемоинформатики (*2009-2012г*)

1. РАЗРАБОТКА

подходов QSAR/QSPR *(поиск соотношений ” структура-активность/структура-свойство)* на основе классических методов статистического анализа и методов машинного обучения

2. ПРИМЕНЕНИЕ ЭТИХ ПОДХОДОВ

для исследования и прогнозирования свойств органических соединений и материалов, а также выработки рекомендаций по созданию новых материалов с заранее заданными свойствами

1. Многоуровневый подход к построению QSAR/QSPR моделей

Решение проблемы недостатка объема экспериментальных данных в реальных базах структур соединений (“проблема малых выборок”)

Оценка качества модели: процедура двойного скользящего контроля

Параметры моделей:

Q2=(SS-PSS)/SS, квадрат коэффициента детерминации
$RMSE_{DCV}$, среднеквадратичная ошибка прогнозирования
MAE_{DCV} , средняя абсолютная ошибка прогнозирования

где: PSS - сумма квадратов ошибок прогноза свойства, SS - сумма квадратов отклонения свойства от среднего значения для усредненных спрогнозированных значений

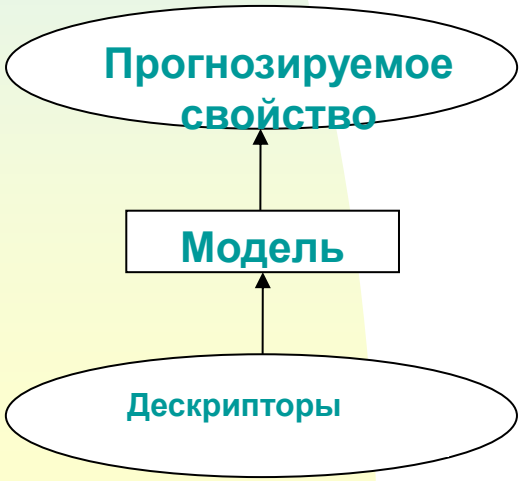
Методы : *Быстрой Пошаговой Множественной Линейной Регрессии (БПМЛР), Искусственные нейронные сети (ИНН)*

BaskinI.I., Zhokhova N.I.,etc./Conf.,Italy,2006

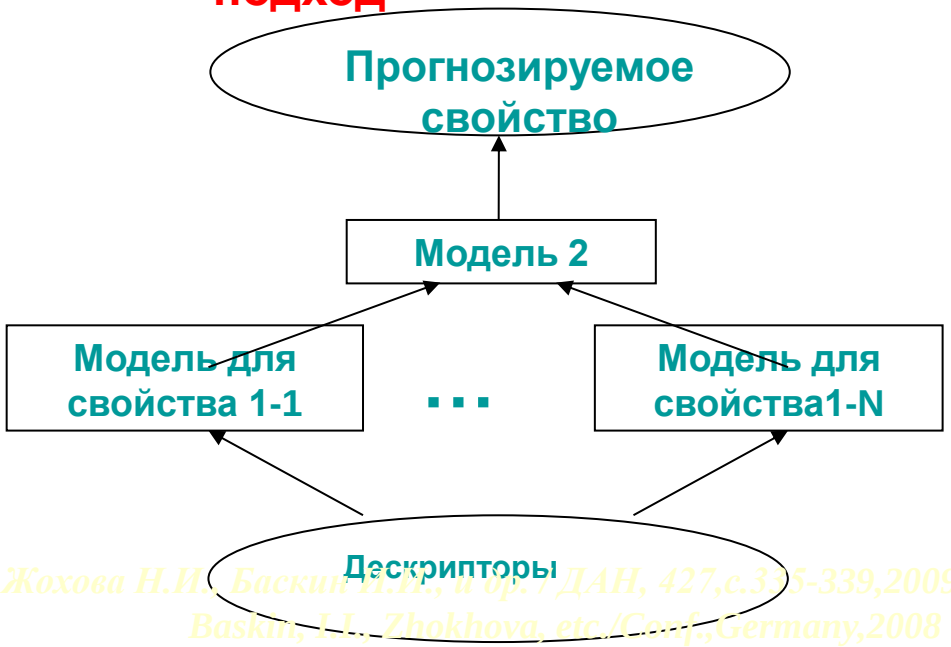
Принцип
объединения моделей на основе многоуровнего обучения

Для небольшой выборки структур соединений подход обеспечивает улучшение прогнозирующей способности целевых моделей за счет интеграции разнородных экспериментальных данных, имеющихся для каждого из связанных друг с другом свойств при использовании соответствующих выборок большего размера.

Одноуровневый подход



Многоуровневый подход



Жохова Н.И., Баскин И.И., и др. // ДАН, 427,с.335-339,2009
Baskin, I.I., Zhokhova, etc.//Conf.,Germany,2008

Сравнение одноуровнего и многоуровнего подходов

при построении моделей для прогнозирования коэффициента сорбции органических соединений в почве и растворимости фуллерена С60 в качестве входных параметров использованы значения липофильности log P и четырех констант Абрахама A, B, E и S, спрогнозированные на основе ИНН моделей первого уровня, построенных для соответствующих обучающих выборок структур органических соединений.

Свойство	Одноуровневый подход		Многоуровне-вый подход	
	Q^2_{DCV}	$RMSE_{DCV}$	Q^2_{DCV}	$RMSE_{DCV}$
Коэффициент сорбции 568 органических соединений в почве, lg K_{oc} **(распределение в экосистеме H_2O/седименты)**	0.598	0.76	0.800	0.53
Растворимость фуллерена С60 в 165 органических растворителях, lg S	0.448	0.91	0.637	0.74

Статистические характеристики ИНН моделей «структура- свойство» первого уровня для прогнозирования липофильности и констант Абрахама для наборов структур органических соединений

Свойство	Число соединений в выборке	R	$RMSE_t$	$RMSE_v$
Log P	7805	0.980	0.345	0.395
Абрахам A (кислотность по отношению к образованию Н-связей)	457	0.983	0.051	0.058
Абрахам B (основность по отношению к образованию Н-связей)	457	0.971	0.066	0.081
Абрахам E (избыточная молярная рефракция)	457	0.997	0.040	0.074
Абрахам S (диполярность/ поляризуемость)	457	0.987	0.072	0.137

2. Фрагментные дескрипторы с “выделенными” атомами для построения QSAR/QSPR моделей

↓

►► Выделение меткой в структуре молекулы атомов, которые играют специфическую роль в природе моделируемого свойства

►► Построение моделей с использованием в качестве дескрипторов фрагментов структуры с выделенными атомами

►► Модели наглядно показывают пути влияния на изучаемое свойство отдельных атомов или групп внутри молекулы

Предлагаемый прием обеспечивает использование при построении моделей наиболее важных по смыслу фрагментных дескрипторов.

Эффективен при прогнозировании:

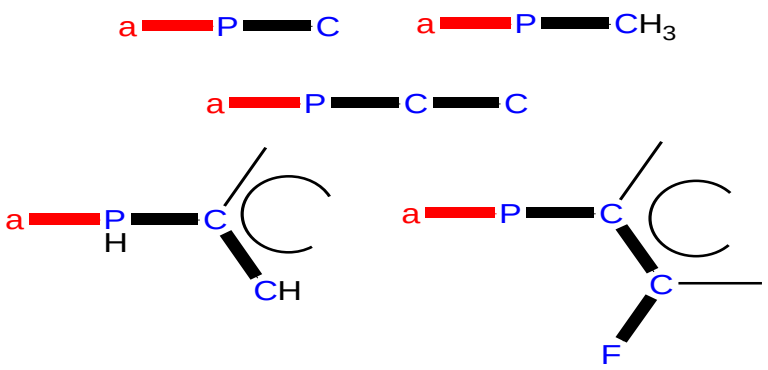
А. физических свойств полимеров (за счет добавления специальных меток к атомам, принадлежащим основной цепи полимера) ,
а также при

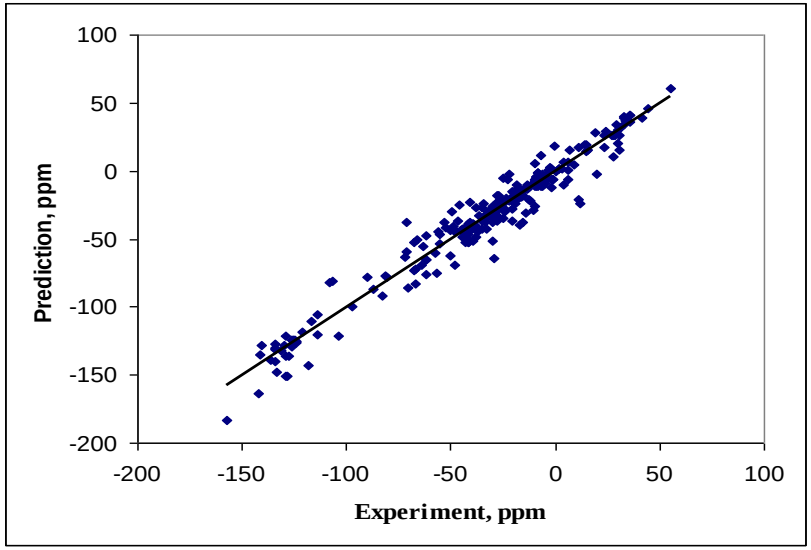
Б. Прогнозировании “локальных” свойств, связанных с присутствием в структуре атома

Химический сдвиг ^{31}P ЯМР в фосфинах.

В структурах выборки моно-втор-и трет-замещенных фосфинов меткой выделен атом Р. Приведены наиболее важные фрагменты с выделенным атомом Р, вошедшие в модель. Первые три отражают σ-индукционное влияние алкильных заместителей на атом фосфора, четвертый – эффект сопряжения с ароматическим ядром, пятый – влияние расположенного в орто-положении атома фтора.

Наиболее важные фрагменты:





База : $PH_{3-n} – R_n$

(PH_2-R, PH-R_2, P-R_3)

n=291 соединения

Q^2_{DCV} = 0.9560, $RMSE_{DCV}$=9.1 ppm, MAE_{DCV}=6.1 ppm

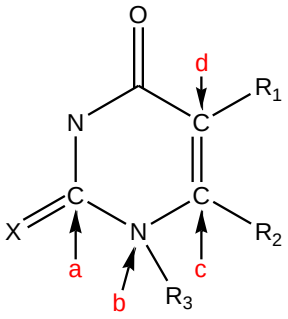
Жохова Н.И., Баскин И.И.,etc.//ДАН,417,с.639-641,2007

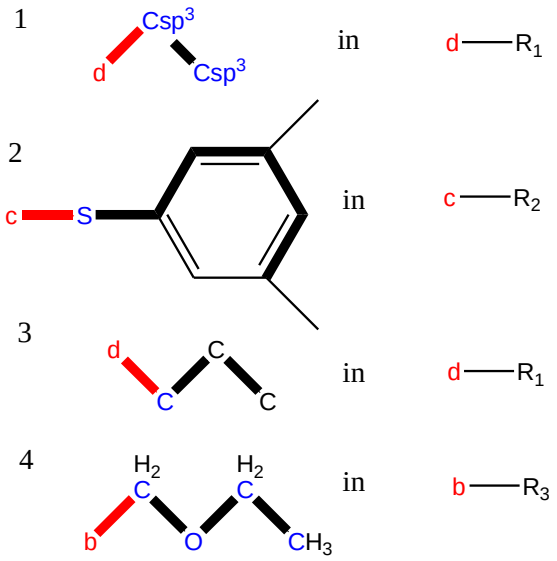
B. Прогнозирование биологической активности в узких рядах органических соединений с одинаковым общим скелетом

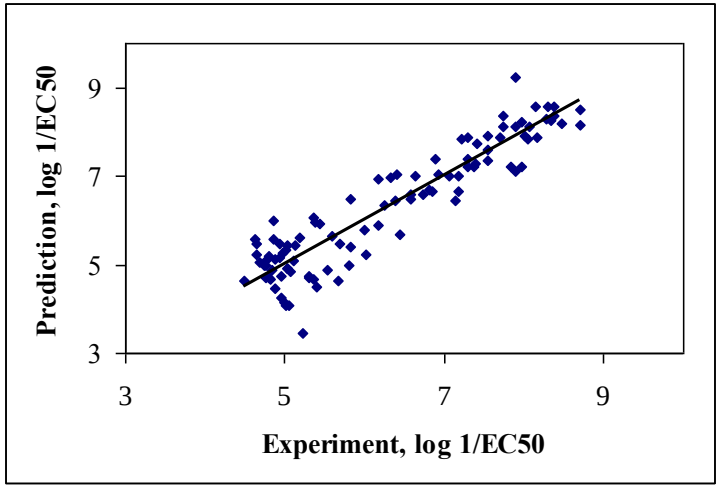
При прогнозировании биологической активности в узких рядах органических структур меткой выделяют позиции присоединения заместителей к общему скелету молекул.

Способность производных 1-[(2-гидроксиэтокси)-метил]–6(фенилтио)тимина (HEPT) ингибировать обратную транскриптазу вируса ВИЧ-1 (log (1/ E_{50})

Наиболее важные фрагменты:







База: n=101 соед.

$$Q^2_{DCV} = 0.856,\ RMSE_{DCV}=0.5,\ MAE_{DCV}=0.4$$

Жохова Н.И., Баскин И.И.,etc.//ДАН,417,с.639-641,2007

Г. Прогнозировании кинетических параметров органических реакций одного типа

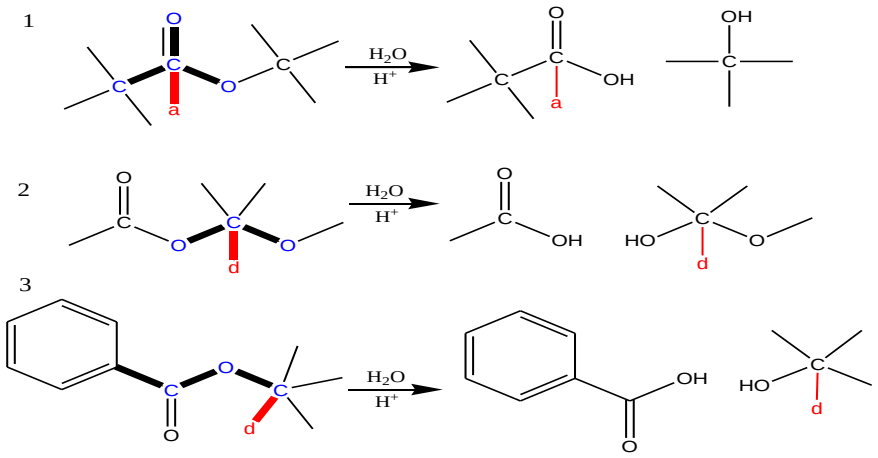
Константа скорости реакции кислотного гидролиза сложных эфиров

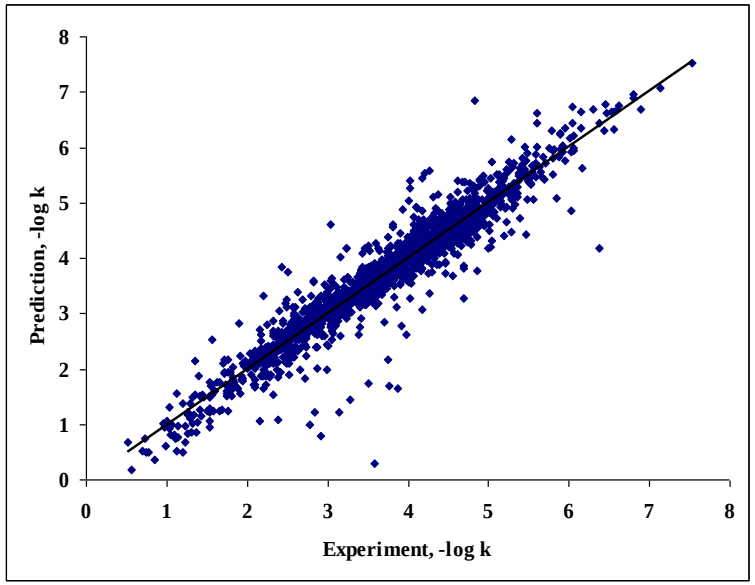
В качестве «выделенных» отмечены атомы углерода, входящие в состав реакционных центров на одной из стадий реакции кислотного гидролиза сложных эфиров.

Показаны три наиболее важных фрагмента из вошедших в построенную модель. Первый описывает стерическое влияние заместителей при α-углеродном атоме карбоновой кислоты, второй – электронное влияние расположенного в уходящей группе атома кислорода, несущего неподеленные электронные пары, третий – влияние фенильной группы при карбоксиле.

Данный прием удобен для автоматизированного извлечения из огромной массы экспериментальных данных основных факторов, влияющих на протекание органических реакций.

Наиболее важные фрагменты:





База: n=2092
T 0-154°C
H_2O/раст.0-98%

$$Q^2_{DCV} = 0.9162,\ RMSE_{DCV}=0.31,\ MAE_{DCV}=0.19$$

Жохова Н.И.., Баскин И.И., etc. // ДАН, 2007,

3. Метод прогнозирования принадлежности органических соединений к фармакологическим группам

на основе двухклассовой классификации,

фрагментных дескрипторов и метода опорных векторов

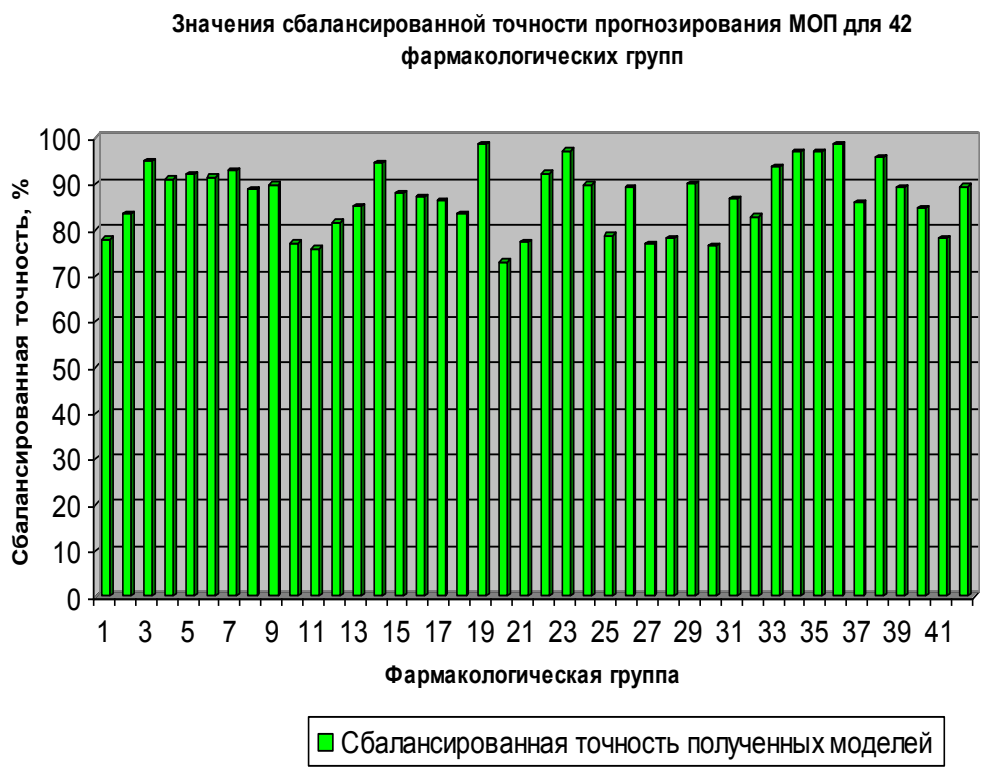
Построенные модели “структура-активность” позволяют успешно соотносить структуру органического соединения с ее принадлежностью к потенциальной фармакологической группе и прогнозировать возможную активность потенциальных лекарственных препаратов.

Программа LIBSVM (

Моделирование принадлежности органических соединений к фармакологическим группам на основе метода опорных векторов

База KEGG DRUG

(6000 действующих структур лекарственных препаратов, 120 фармакологических групп)



1.Расчет ФД
2.Определение активности 0-1
3.Выбор 42 фарм. Групп
3. Создание метода расчета кернелов
5.Оптимизация модификации и параметров метода (C,
6.Процедура ресэмплинга для несбалансированных выборок

Параметры моделей:

Селективность= ПА / (ПА + ЛН)

Специфичность = ПН / (ПН + ЛА)

Сбалансированная точность = (Селект. + Специф.) / 2

План доклада

- Что такое хемоинформатика (молекулярная информатика)
- Наш вклад
- **Текущие исследования**
- Возможные будущие направления работ

The continuous molecular fields approach to building 3D-QSAR models

Igor I. Baskin & Nelly I. Zhokhova

Journal of Computer-Aided Molecular Design
Incorporating Perspectives in Drug Discovery and Design

ISSN 0920-654X
Volume 27
Number 5

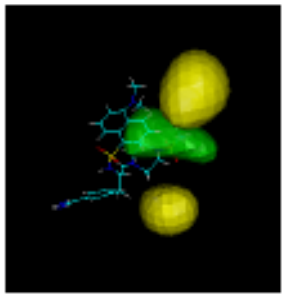
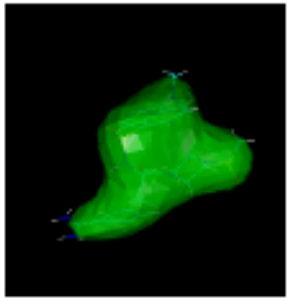
J Comput Aided Mol Des (2013)
27:427-442
DOI 10.1007/s10822-013-9656-4

Springer

Метод непрерывных молекулярных полей

Традиционный QSAR: $Активность = F(X) = \sum c_i x_i$

CMF: $Активность = F[X(\mathbf{r})] = \int C(\mathbf{r}) X(\mathbf{r}) d\mathbf{r}$

$Активность = \int$  .  $d\mathbf{r}$

C(r) **X(r)**

Поля регрессионных коэффициентов

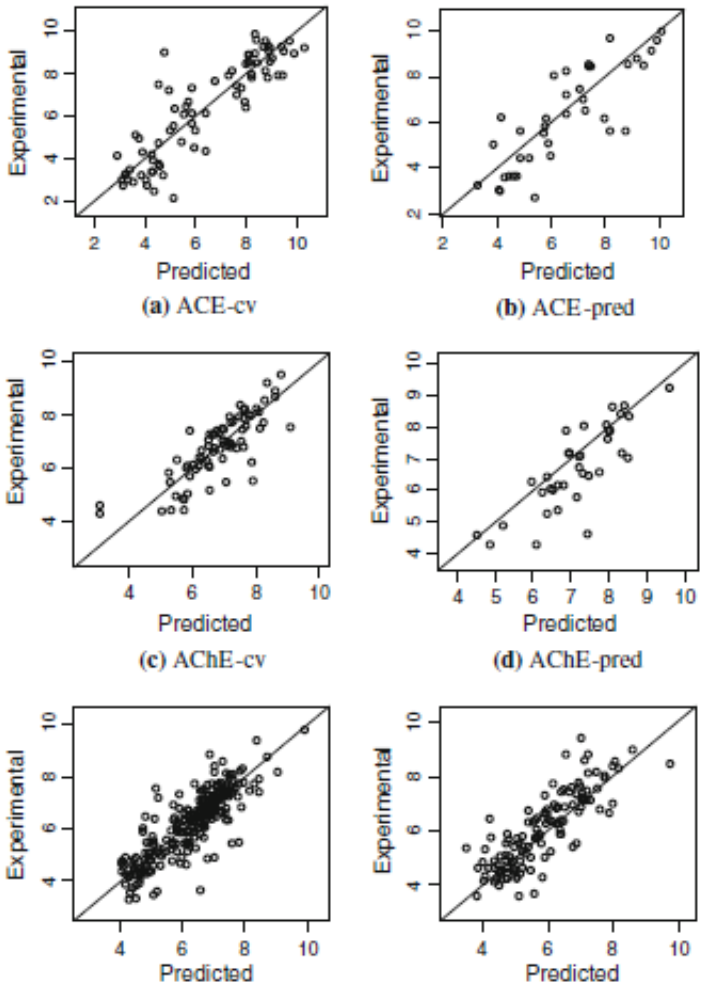
Молекулярные поля, аппроксимированные Гауссовыми функциями

Типы физико-химических полей: электростатическое, стерическое, липофильное, донорное и акцепторное поля по отношению к образованию водородных связей

Метод непрерывных молекулярных полей

Table 1 QSARDataSets

Ligand data set	Training set	Test set
Angiotensin converting enzyme (ACE) inhibitors	76	38
Acetylcholinesterase (AchE) inhibitors	74	37
Ligands for benzodiazepine receptors (BZR)	98	49
Cyclooxygenase-2 (COX-2) inhibitors	188	94
Dihydrofolatereductase (DHFR) inhibitors	237	124
Glycogen phosphorylase b (GPB) inhibitors	44	22
Thermolysin (THER) inhibitors	51	25
Thrombine (THR) inhibitors	59	29



Continuous Molecular Fields

Search this site

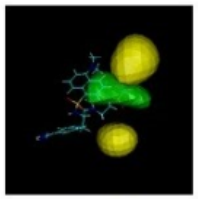
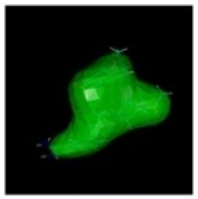
Home

Home

The project belongs to the scientific area of chemoinformatics. The aim of the project is to develop a new universal approach to predict various properties of chemical compounds, which uses description of molecules by means of continuous fields (such as electrostatic, steric, electron density functions, etc). The essence of the proposed Continuous Molecular Fields (CMF) approach consists in performing statistical analysis of functional molecular data by means of joint application of kernel machine learning methods and special kernels which compare molecules by computing overlap integrals of their molecular fields. The principal novelty of this approach is the ability to conduct a statistical analysis of chemical data presented in the form of continuous molecular fields, i.e. an infinite number of attributes organized in a functional form. This approach is an alternative to traditional methods of building "structure-property" models based on the use of fixed sets of molecular descriptors. The project deals with possible ways of the development of this approach and its applications in various areas of chemistry.

Traditional QSAR $Activity = F(X) = \sum_i c_i x_i$

CMF $Activity = F[X(\mathbf{r})] = \int C(\mathbf{r}) X(\mathbf{r}) d\mathbf{r}$

$Activity = \int$  $\cdot$  $d\mathbf{r}$

$C(\mathbf{r})$ $X(\mathbf{r})$

References of CMF:

- I.I.Baskin, N.I.Zhokhova. The continuous molecular fields approach to building 3D-QSAR models. *J. Comput. Aided Mol. Des.*, **2013**, Vol. 27, No. 5, pp. 427-442.

The CMF approach is implemented in the CMF package of R scripts.

The following distributions are currently available:

- Supplementary material for JCAMD article (a set of R scipts with 8 examples necessary to reproduce results discussed in the article);
- The last version of package CMF (R scripts implementing the CMF approach) with manual and changelog .

To download them, go to the Files page.

http://sites.google.com/site/conmolfields/home

Гранты РФФИ 10-07-00201 и 13-07-00511

Непрерывные индикаторные поля (НИП)

Индикаторное поле $\theta_t(r)$ показывает степень принадлежности точки ***r*** атому, относящемуся к типу ***t***

Согласно методу CMF: $$X_f(\vec{r}) = \sum_i w_{fi} e^{-\alpha(\vec{r}-\vec{r}_i)^2}$$

w_{fi} – вклад атома i в поле типа f *(напр., для электростатического поля – частичный заряд на атоме)*

Предполагается, что значения параметров w_{fi} являются переносимыми между атомами одного типа (хим. элемент, гибридизация, окружение в параметризации силового поля Tripos)

$$X_f(\vec{r}) = \sum_t c_{ft}\theta_t(\vec{r}) \qquad \theta_t(\vec{r}) = \sum_i \delta_{t,T(i)} e^{-\alpha(\vec{r}-\vec{r}_i)^2}$$

$\delta_{t,T(i)} = 1$ если атом i имеет тип $T(i)$

База комплексонов катионов Am^{+3}/Eu^{+3}

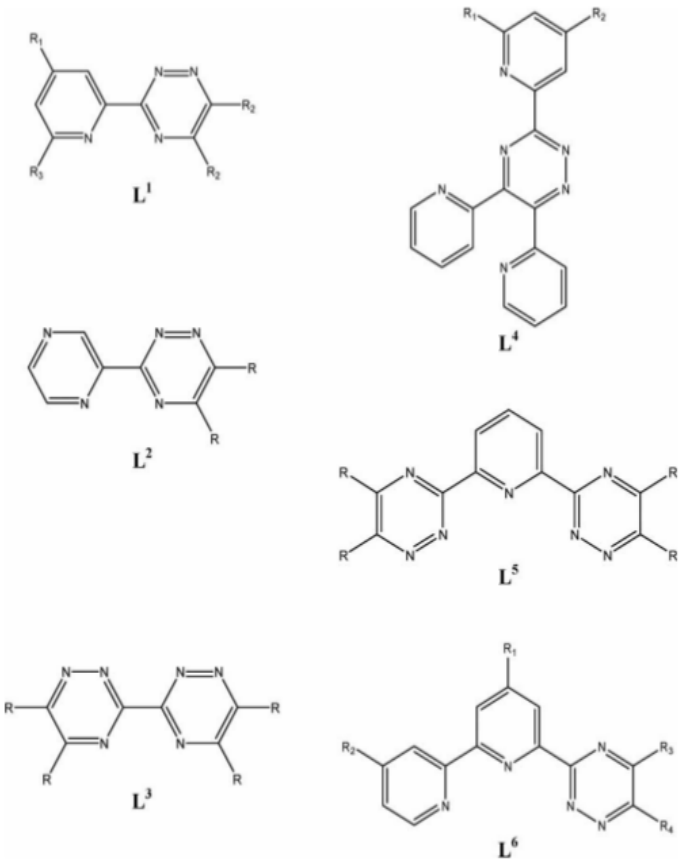
SF - фактор разделения Am^{+3}/Eu^{+3}

Am ($5f^7 7s^2$)	**Eu** ($4f^7 6s^2$)
0.097 нм*	0.100 нм

Значения SF были получены путём экстракции катионов металлов (Am^{+3}, Eu^{+3}) из водного раствора азотной кислоты в 1,1,2,2-тетрахлороэтановую фазу, содержащую α-бромкаприновую кислоту в качестве соэстрагента

48 полиазагетероциклических соединений

Соотношение лиганд : металл = 2 : 1

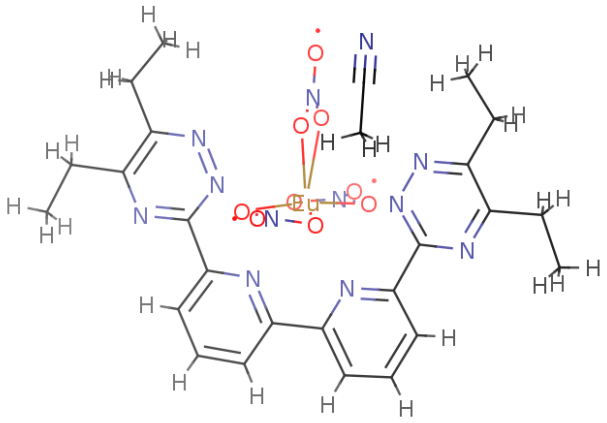


Базовые структуры полиазагетероциклических соединений**

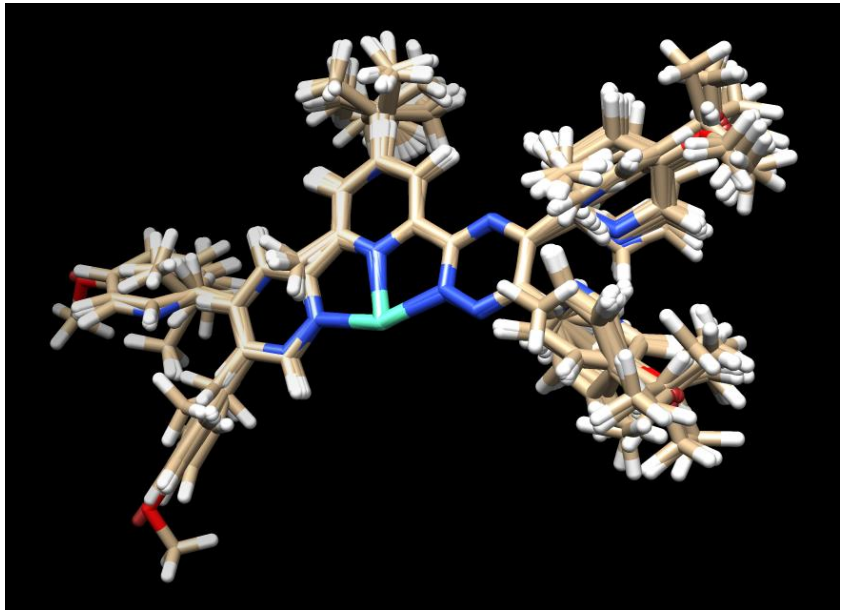
** Varnek A. et al. // Solvent Extr. Ion. Exch. 2007. V.25. P.1–26

* Справочник химика. под. Ред. Б.П. Никольского. М-Л.: Химия. 1982

Выравнивание базы лигандов



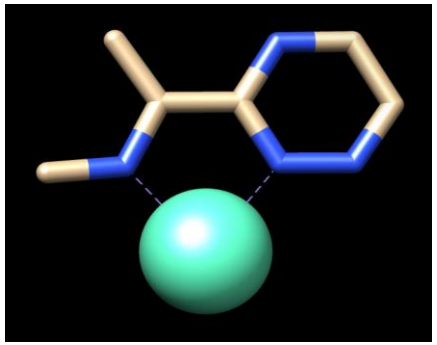
Структура комплекса 5,6-диэтил-3-пиридин-1,2,4-триазина с катионом Eu^{+3} по данным рентгеноструктурного анализа



Выравненная база лигандов

Построение 3D геометрии на ChemAxon

алгоритм совмещения структур с использованием шаблона по методу наименьших квадратов

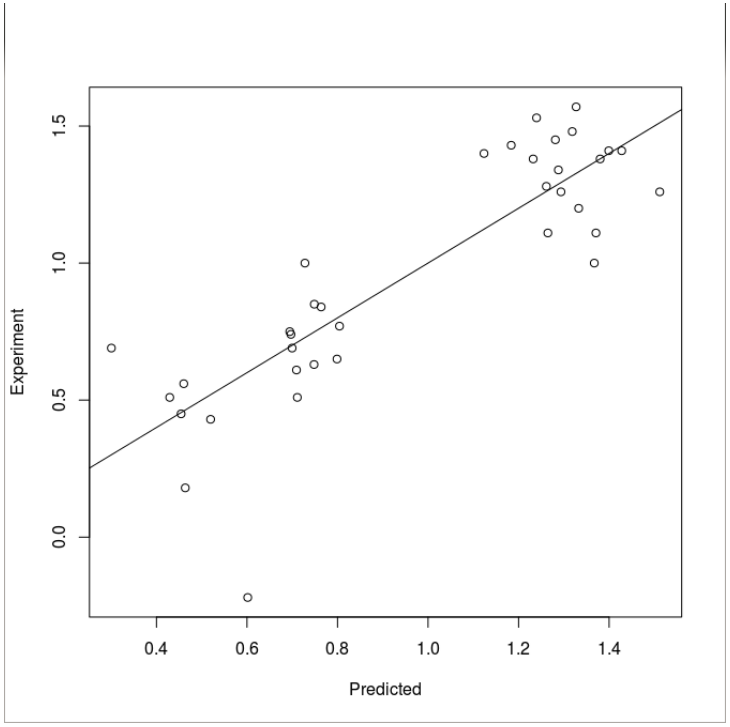


Структура шаблона

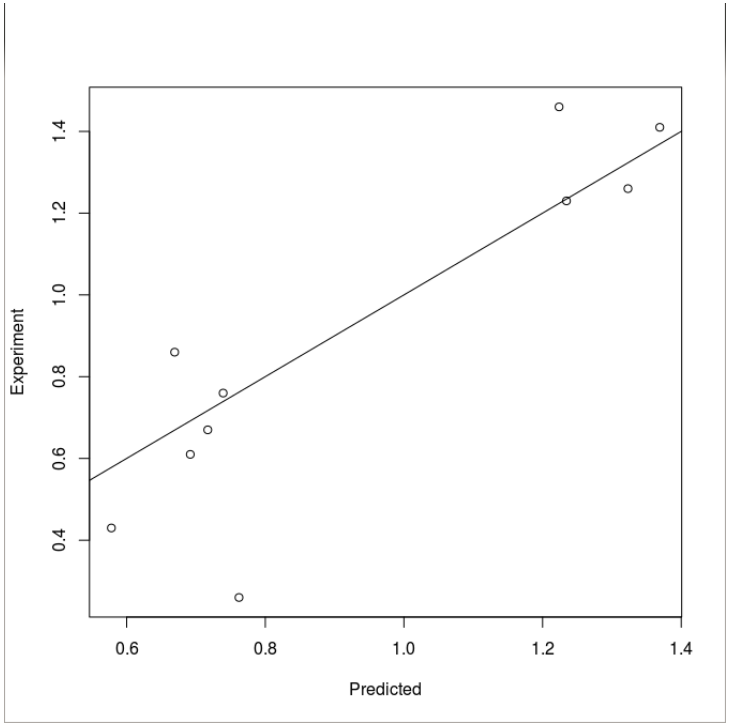
Статистические параметры моделей, полученных методами 3D QSPR и 2D QSPR

3D QSPR (CMF) с использованием непрерывных индикаторных полей

q^2	$RMSE_{cv}$	R^2_{pred}	$RMSE_{pred}$
0.73	0.22	**0.76**	0.20



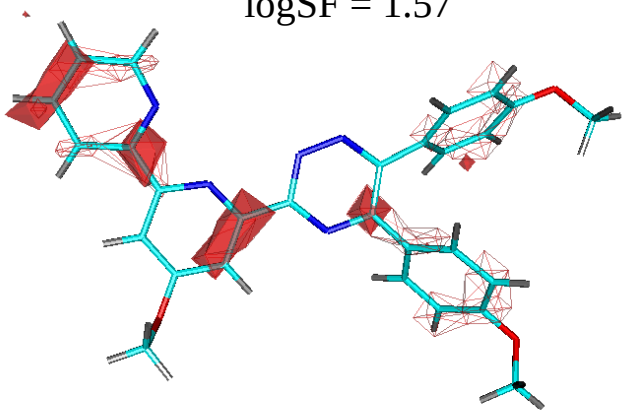
Скользящий контроль



Независимая выборка

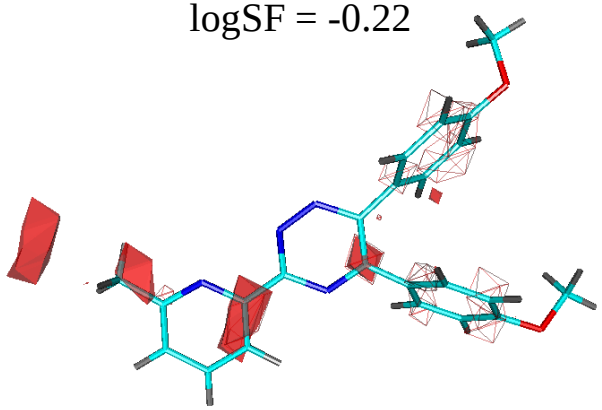
Визуализация перекрывания индикаторных молекулярных полей и полей регрессионных коэффициентов модели для типов атомов C (sp^2) и C (sp^3)

logSF = 1.57

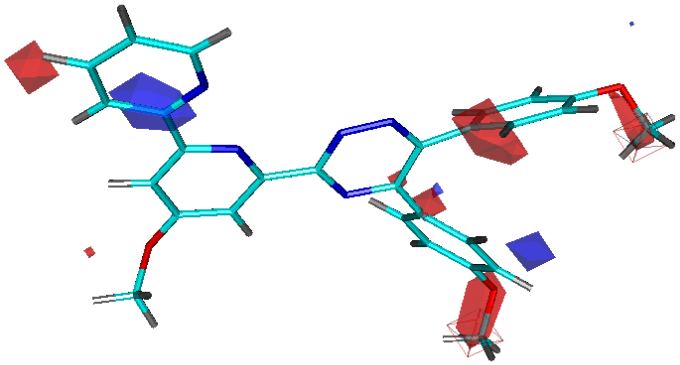


C (sp^2) для наиболее активного соединения

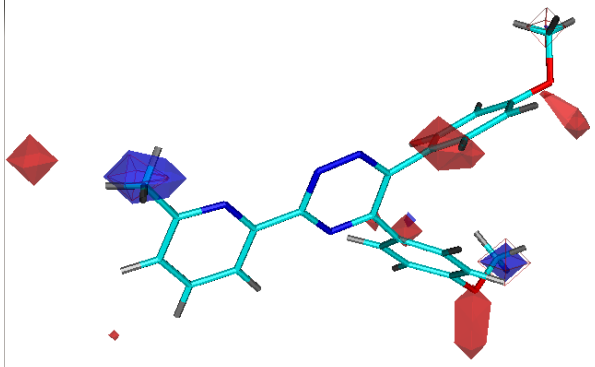
logSF = -0.22



C (sp^2) для наименее активного соединения



C (sp^3) для наиболее активного соединения

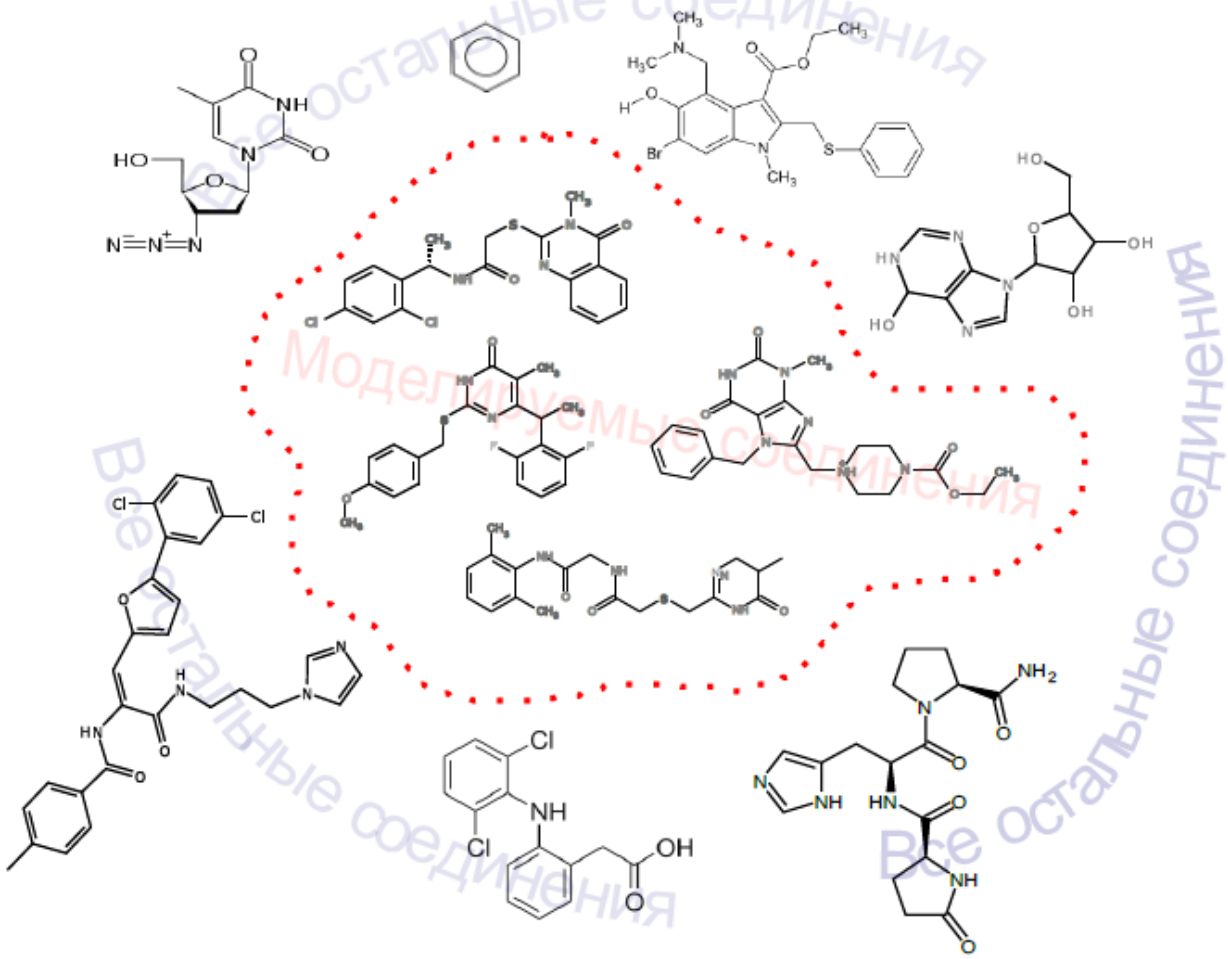


C (sp^3) для наименее активного соединения

Сплошная – поля регрессионных коэффициентов

Сетка – молекулярные поля индивидуальных молекул

Одноклассовая классификация



Одноклассовая классификация – это группа методов машинного обучения, которая позволяет отличать объекты одного класса, часть из которых использовалась для обучения, от объектов всех остальных классов

Одноклассовая классификация 1-SVM

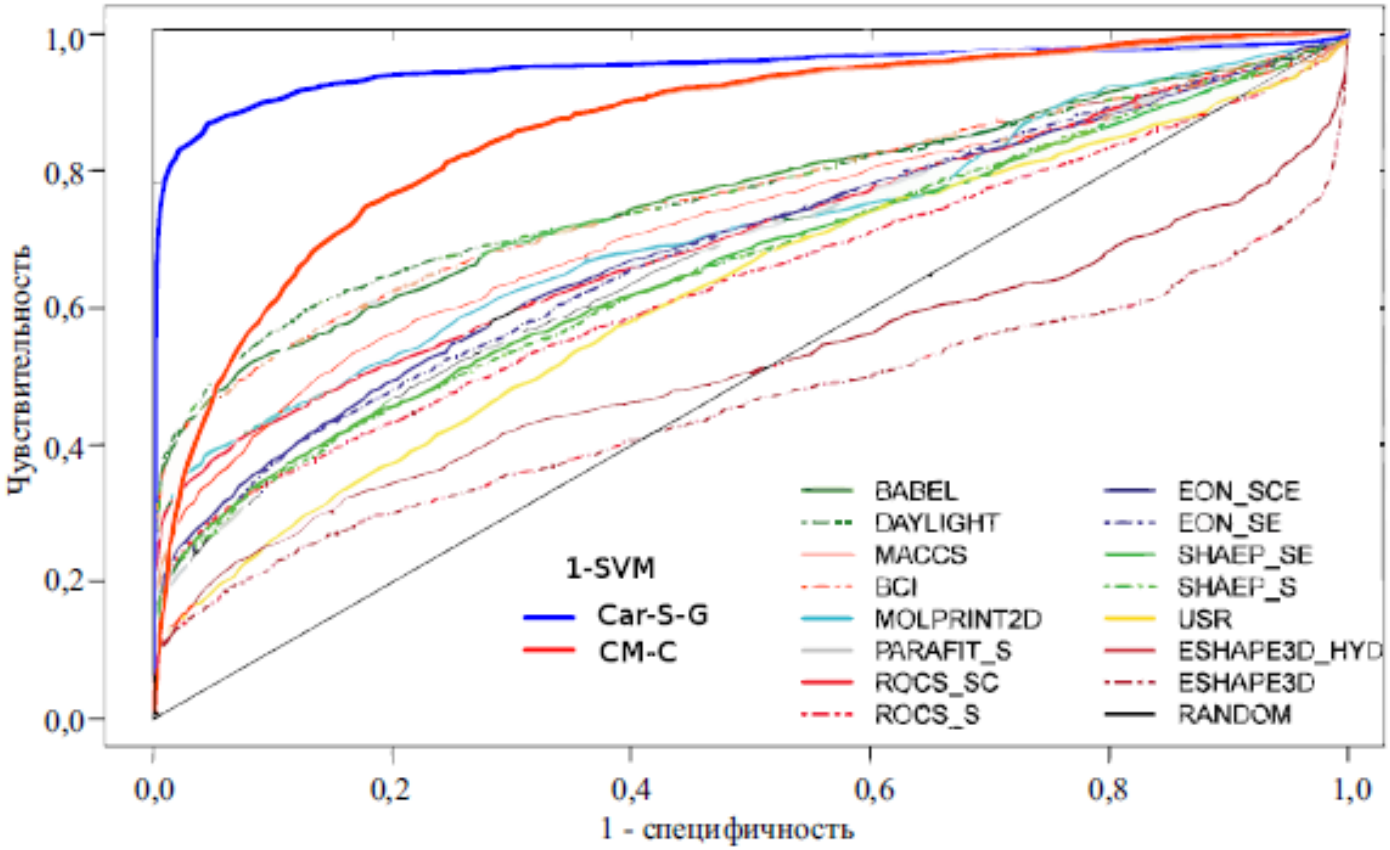
В методе 1-SVM ищется плоскость, максимально отделяющая образцы моделируемого класса в пространстве признаков от начала координат:

Параметры классификатора:

- параметр ν — верхняя граница доли неверно классифицированных образцов;
- параметры ядер скалярного произведения.

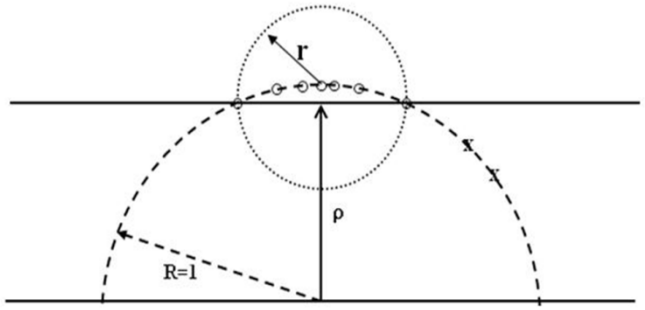
Варьируя положение разделяющей плоскости, можно получать разную эффективность классификатора.

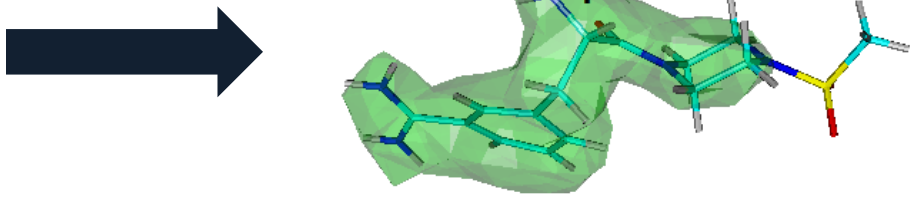
Эффективность одноклассового подхода



Эффективность одноклассовых моделей выше, чем эффективность большинства современных подходов к LBVS. CAR-S-G — 1-SVM модель, ядро Гаусса, дескрипторы Кархарта; CM-C — комбинированное поле. Остальные кривые взяты из работы Venkatraman V., Perez-Nueno V., Mavridis L., Ritchie D. // J. Chem. Inf. Model. 2010. V. 50. P. 2079–2093.

Физическая интерпретация одноклассовых моделей на основе непрерывных молекулярных полей





Поля коэффициентов одноклассовых моделей образованы направляющими косинусами нормали к раздедяющей гиперплоскости. В физическом пространстве они образуют описание «идеальной» конфигурации молекулярных полей (т.е. молекулярной формы) лиганда, которая в процессе виртуального скрининга сравнивается с полями тестовых молекул. Чем выше их сходство, тем выше шансы того, что тестовая молекула принадлежит моделируемому классу.

Одноклассовые модели на основе непрерывных молекулярных полей

ISSN 0012-5008, Doklady Chemistry, 2011, Vol. 440, Part 2, pp. 263–265. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © P.V. Karpov, I.I. Baskin, N.I. Zhokhova, N.S. Zefirov, 2011, published in Doklady Akademii Nauk, 2011, Vol. 440, No. 4, pp. 480–483.

CHEMISTRY

Method of Continuous Molecular Fields in the One-Class Classification Task

P. V. Karpov, I. I. Baskin, N. I. Zhokhova, and Academician N. S. Zefirov

Russian Chemical Bulletin, International Edition, Vol. 60, No. 11, pp. 2418—2424, November, 2011

One-class approach: models for virtual screening of non-nucleoside HIV-1 reverse transcriptase inhibitors based on the concept of continuous molecular fields*

P. V. Karpov,[a] I. I. Baskin,[a]* N. I. Zhokhova,[a] M. B. Nawrozkij,[b] A. N. Zefirov,[a]
A. S. Yablokov,[b] I. A. Novakov,[b] and N. S. Zefirov[a]

План доклада

- Что такое хемоинформатика (молекулярная информатика)
- Наш вклад
- Текущие исследования
- **Возможные будущие направления работ**

Информатика материалов

CHEMICAL REVIEWS

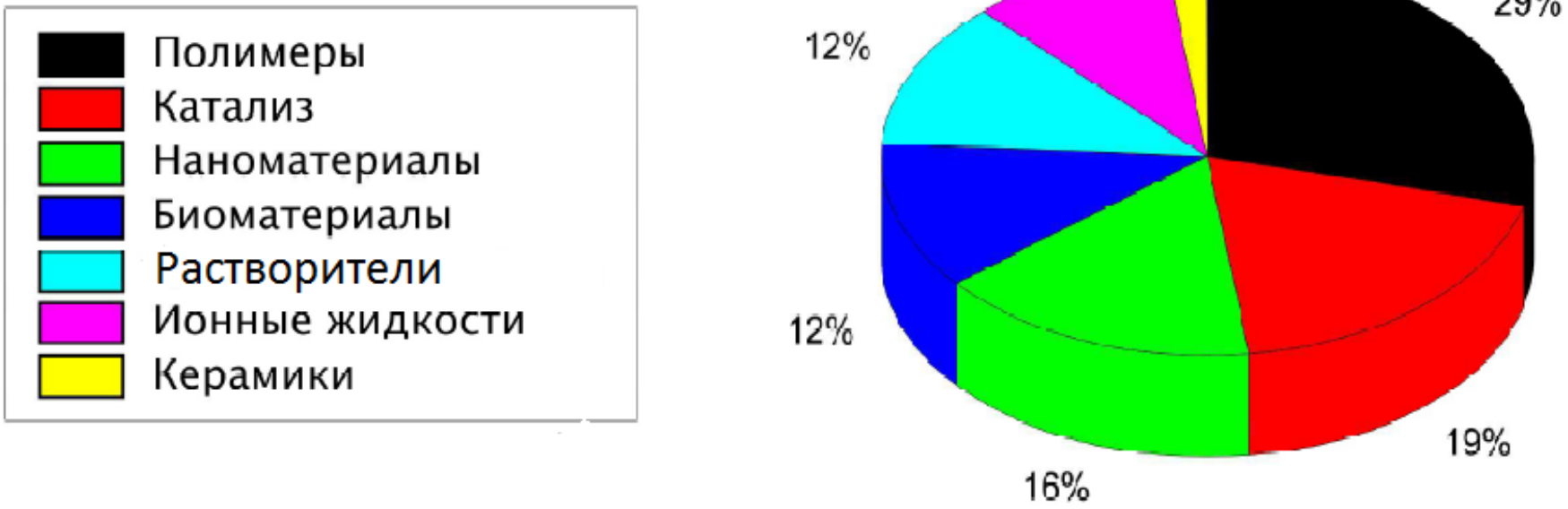
REVIEW

pubs.acs.org/CR

Quantitative Structure–Property Relationship Modeling of Diverse Materials Properties

Tu Le,[†] V. Chandana Epa,[‡] Frank R. Burden,[†] and David A. Winkler*[,†,§]

Chem. Rev. 2012, 112, 2889–2919



Adv Polym Sci
DOI:10.1007/12_2009_18

Polymer Informatics

Nico Adams

Информатика полимеров

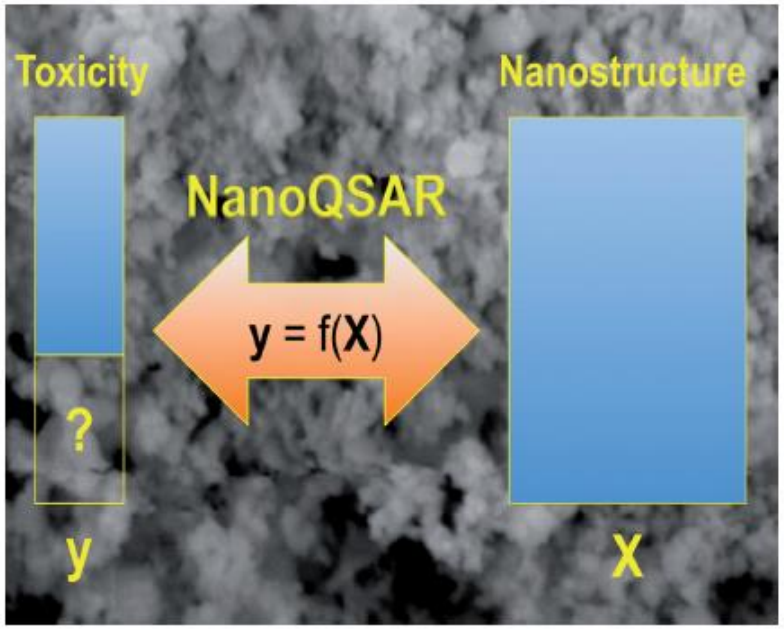
Компьютерное представление и хранение в базах данных информации о полимерах

Использование информации о полимерах:
- **Хемометрический анализ**
- **Прогнозирование свойств полимеров при помощи моделей QSPR**
- **Дизайн новых полимеров с заданными свойствами**

Прогнозируемые свойства: температура стеклования, индекс рефракции, нижняя критическая температура растворения, характеристическая вязкость, плотность в аморфном состоянии, разные виды биологической активности

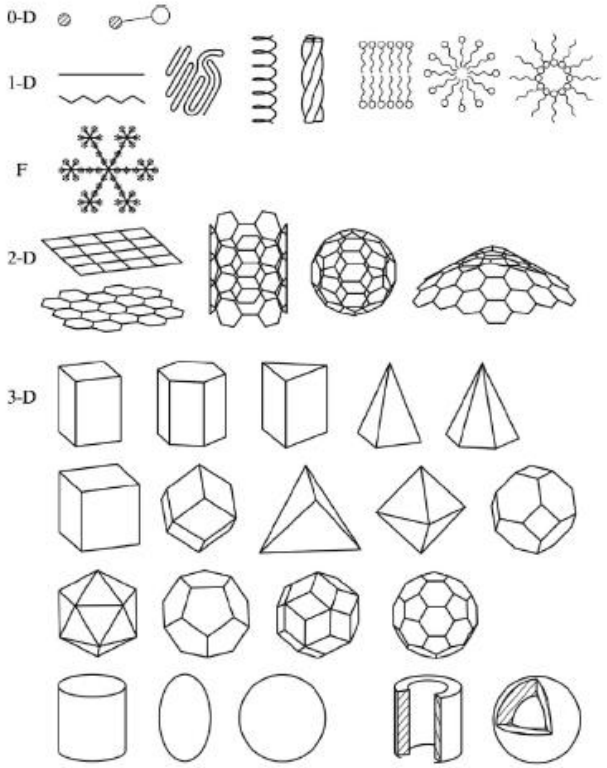
Toward the Development of "Nano-QSARs": Advances and Challenges

*Tomasz Puzyn, Danuta Leszczynska, and Jerzy Leszczynski**



small 2009, *5,* No. 22, 2494–2509

- size distribution
- agglomeration state
- shape
- porosity
- surface area
- chemical composition
- structure-dependent electronic configuration
- surface chemistry
- surface charge
- crystal structure



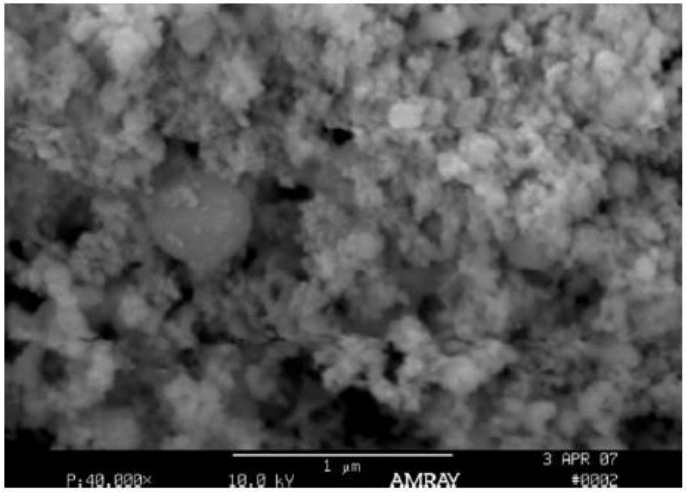


Figure 2. SEM image of nanometer-sized La_2O_3.

Возможные будущие направления работ

- Продолжение работ моделированию и дизайну комплексонов и более сложных супрамолекулярных комплексов на основе метода непрерывных полей.
- «Супрамолекулярный асссемблер» - генератор пространственных структур супрамолекулярных соединений и материалов (включая полимерные и кристаллические)
- Распространение метода непрерывных молекулярных полей на описание
  - как биологических так и синтетических сополимеров
  - кристаллов
  - наноматериалов
- Участие в совместных проектах по дизайну новых материалов по вышеперечисленным направлениям