



МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
имени М.В.ЛОМОНОСОВА

ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ

КАФЕДРА физики полимеров и кристаллов

Компьютерный дизайн высокоселективных экстрагентов для  
разделения редкоземельных элементов с помощью квантово-  
химических расчетов и моделей "структура-свойство"

магистр Карпов К.В., рук. ст.н.с., кхн Жохова Н.И.

Москва  
2018



## Краткое общее представление области исследований - **ХЕМОИНФОРМАТИКА**

**“Хемоинформатика – это применение методов информатики для решения химических проблем” - И. Гастайгер, 2003**

Основной сферой применения методов хемоинформатики является **изучение общих закономерностей, связывающих микро- и макроскопические свойства химических объектов.**

С этой целью в хемоинформатике используют **методологию SAR/QSAR/QSPR**, которая основана на построении **с помощью методов машинного обучения** статистических моделей, связывающих значения характеристик молекулярной структуры химических объектов со значениями их свойств.

**Модели “структура-свойство”** используют для расчета неизвестных свойств новых химических объектов и конструирования материалов с заданными свойствами.

Моделирование “структура-свойство” **проводят на основе баз данных**, содержащих информацию о структурах и свойствах химических объектов.

# Актуальность изучаемой проблемы

**Редкоземельные элементы** (РЗЭ) и их препараты (с содержанием основного металла 99,999%) широко востребованы в различных областях современных инновационных технологий



La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Мировое производство РЗЭ в 2015 г - 200 тыс. тонн

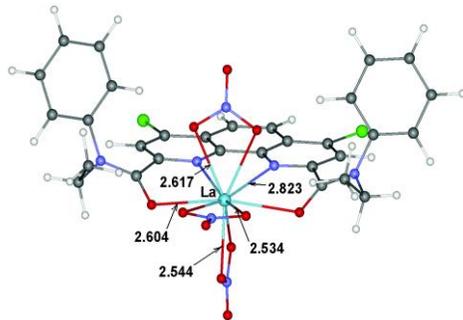
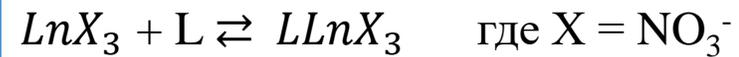
**Выделение** индивидуальных РЗЭ из их смесей представляет сложную задачу по причине близости их химических свойств вследствие заполнения 4f внутренней электронной оболочки при движении от  $\text{La}^{3+}$  к  $\text{Lu}^{3+}$   $\blacktriangleright 4f^{1-14}, 5d^1, 6s^2$

Фактор различия между РЗЭ - уменьшение R иона  $\text{Ln}^{3+}$  с ростом Z от  $1,061\text{Å}$  ( $\text{La}^{3+}$ ) до  $0,85\text{Å}$  ( $\text{Lu}^{3+}$ )

# Актуальность изучаемой проблемы

Известным промышленным способом получения индивидуальных РЗЭ из смесей их концентратов является жидкостная экстракция, которую проводят в системе водный раствор смеси солей РЗЭ - органический растворитель.

В процессе экстракции **органический лиганд захватывает ион РЗЭ, образуя при этом координационный комплекс**, и выводит его в органическую фазу. Селективность лиганда, характеризует **фактор селективности**.



$$K = \frac{[LLnX_3]}{[L][LnX_3]}$$

$$\text{фактор селективности } SF_{Ln1/La2} = \frac{K_1}{K_2}$$

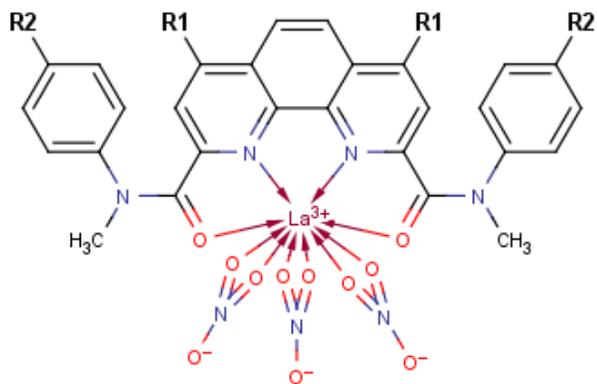
# Актуальность изучаемой проблемы.

Задача - создание высокоселективных экстрагентов для разделения РЗЭ, америция и кюрия

Используемые на сегодняшний день в промышленности экстрагенты не обладают высокой селективностью по отношению к РЗЭ, что приводит к необходимости использования последовательных экстракционных каскадов, состоящих из сотен экстракторов.

**Важной стратегической задачей является создание высокоселективных экстрагентов для разделения РЗЭ.**

**Такие экстрагенты эффективны для разделения элементов, в том числе америция и кюрия, при переработке отходов замкнутого ядерного цикла,**



1,10-фенантролин-2,9-дикарбоксиамиды

Из исследований известно, что одним из классов соединений, наиболее подходящих на роль экстрагентов, являются производные N-донорных гетероциклов, содержащие “жесткие” кислородные и “мягкие” азотные донорные центры - атомы кислорода и азота с неподеленными парами электронов, которые образуют координационные связи с катионами РЗЭ с образованием комплексов.

# Актуальность проблемы.

Задача - создание высокоселективных экстрагентов для разделения РЗЭ, америция и кюрия

**Поиск новых экстрагентов**, основанный на экспериментальном синтезе и тестировании, требует дорогостоящих и времязатратных процедур. Для многих производных целевой группы соединений методы синтеза не разработаны.

Поэтому для расчета потенциальных свойств новых экстрагентов, которые влияют на образование комплексов, **мы применили методологию, успешно развиваемую в мире в последние 5 лет для прогнозирования свойств новых материалов.**

Она заключается в исследовании корреляций между микро- и макроскопическими свойствами химических объектов на основе построения статистических моделей “структура-свойство” (QSPR) с помощью **методов машинного обучения.**

**Модели QSPR** строят с использованием **баз данных**, содержащих наборы структур химических объектов и значений свойств, рассчитанных для этих объектов с помощью **квантово-химических методов.** При построении модели для представления структуры используют **молекулярные дескрипторы.**

**Полученные QSPR модели используют для расчета неизвестных значений свойств для ранее не исследованных объектов.**

# Метод. Построение статистических моделей “структура – свойство” на основе данных квантово-химических расчетов



\* Chemaxon  
[www.chemaxon.com](http://www.chemaxon.com)

\*\* Полуэмпирический метод PM7 - MOPAC16  
[www.openmopac.net](http://www.openmopac.net)  
Метод функционала плотности - SPARTAN14  
[www.wavefun.com](http://www.wavefun.com)

\*\*\* ISIDA/QSPR  
[www.vpsolovev.ru](http://www.vpsolovev.ru)

# Метод. Построение консенсусных моделей “структура - свойство”

## Программа ISIDA-QSPR ([www.vpsolovev.ru](http://www.vpsolovev.ru))

### 1. Виды функциональной зависимости МЛР:

$$Y = \sum_i A_i X_i$$

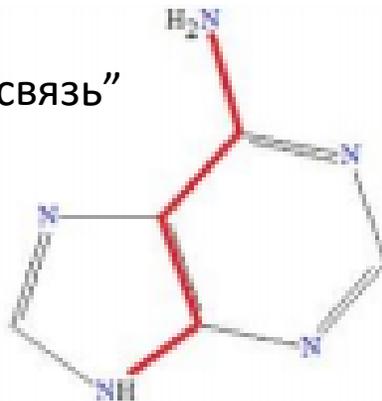
$$Y = A_0 + \sum_i A_i X_i$$

Внутренний 5-кратный скользящий контроль

### 2. Дескрипторы молекулярной структуры – подструктурные 2D фрагменты

Тип фрагментных дескрипторов - топологические пути в виде “атом - связь”

Длину топологического пути - 2-3 мин., 6-15 макс.



### Параметры качества моделей:

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right]^{1/2}$$

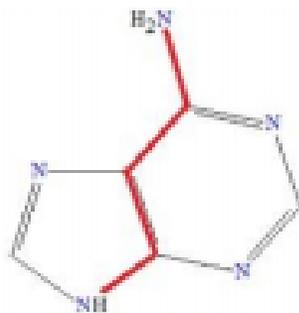
$$R_{det}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \langle y \rangle)^2}$$

$y_i$  и  $\hat{y}_i$  - значения моделируемого свойства, рассчитанные квантово-химически и предсказанные по модели.

$\langle y \rangle$  - среднее значение величин рассчитанного свойства

# Построение статистических моделей “структура - свойство” методом множественной линейной регрессии

1. Структурные молекулярные фрагменты



2. Процесс прямого и обратного отбора дескрипторов

3. Построение индивидуальных моделей

$$y = \sum_i a_i x_i + \Gamma \quad y = a_0 + \sum_i a_i x_i + \Gamma$$

Generated Individual QSPR models

$$Q^2 > Q_{lim}^2 \\ (R^2 - Q^2) < \Delta_{lim}$$

Acceptable QSPR models

Applicability domain (AD)

Bounding Box  
Fragment Control  
Quorum Control

Applicable individual models

Excluding of outlying predictions

Thompson's rule

$$\langle Y_{pred} \rangle = 1/m \sum_{i=1}^m Y_{pred, i}$$

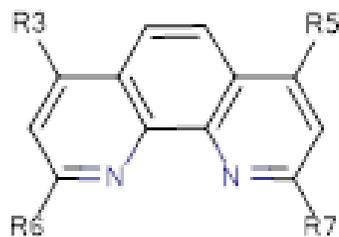
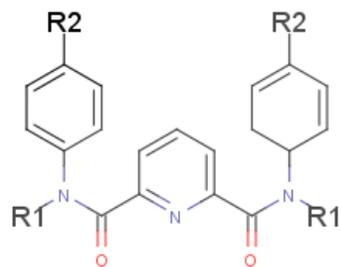
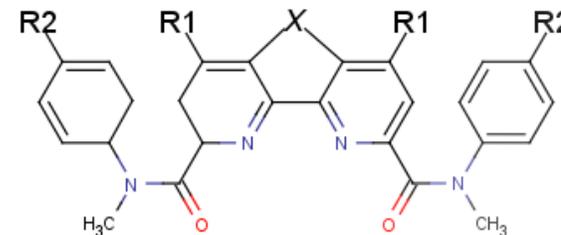
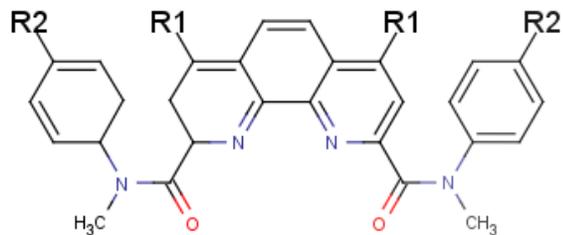
Consensus model

# Построение консенсусных моделей “структура - свойство”

<b>DATA</b> Structure Data File _CURRENT.SDF Modeling Property Mask File 2016-12-09_20-52-28_AllLigFukuiHLD.MSK	<b>MODEL</b> Equation <input checked="" type="checkbox"/> $Y = \text{SUM}(A_i \cdot X_i)$ <input checked="" type="checkbox"/> $Y = A_0 + \text{SUM}(A_i \cdot X_i)$  Variable Selection <input checked="" type="checkbox"/> FVS methods: 3 Ry,i    2,3    60 Ryi > 0,001    Rij < 0,990 <input checked="" type="checkbox"/> backward variable selection t-test: 1.96    eps: 1.0E-12 exact: 0 <input type="checkbox"/> m > n  CONSENSUS MODEL Q2 lim > 0.50    R2-Q2 < 0.30
<b>DESCRIPTORS</b> SMF <input type="checkbox"/> Paths: Atoms <input type="checkbox"/> Paths: Bonds <input checked="" type="checkbox"/> Paths: Atoms-Bonds <input type="checkbox"/> Augmented Atoms <input type="checkbox"/> Augmented Bonds <input type="checkbox"/> Augmented Atoms-Bonds <input type="checkbox"/> Atomic Hybridization  <input type="checkbox"/> add XTR [ ]  Path Length min: 2 max: 15  SMF Modifications <input type="checkbox"/> terminal groups <input type="checkbox"/> equal length paths <input type="checkbox"/> Benson's scheme <input type="checkbox"/> Marked atoms 1 <input type="checkbox"/> Marked atoms 2 <input type="checkbox"/> Marked atoms 3 <input type="checkbox"/> normalize  min count: 3    min comps: 1	<b>VALIDATION</b> <input checked="" type="checkbox"/> LOO <input checked="" type="checkbox"/> rapid External 1 -fold CV  <input checked="" type="checkbox"/> Appl.Domain1 <input checked="" type="checkbox"/> Appl.Domain2 <input checked="" type="checkbox"/> Appl.Domain3 : > 15 % models
<b>Select Directory</b> for output files: D:\SIDA_QSPR\	<input type="checkbox"/> Save Models: <b>Select Directory</b>

**Объекты исследования** – молекулярная структура и свойства перспективных экстрагентов класса N донорных гетероциклов и их комплексов с ионами  $\text{La}^{3+}$ ,  $\text{Nd}^{3+}$ ,  $\text{Eu}^{3+}$  и  $\text{Lu}^{3+}$ .

Сгенерированный набор виртуальных и экспериментальных структур потенциальных экстрагентов (Программа Chemaxon)



**R1, R2** = -CH<sub>3</sub>, -C<sub>6</sub>H<sub>5</sub>, -CH<sub>2</sub>CH<sub>3</sub>, -CH<sub>2</sub>CH<sub>2</sub>OH

**R3** = -H, -Cl, -F, -CH<sub>3</sub>, -NO<sub>2</sub>, -OCH<sub>3</sub>, -CH<sub>2</sub>OCH<sub>3</sub>, -OCH<sub>2</sub>CH<sub>3</sub>

**R3, R5** = H, Cl, F, Me, NO<sub>2</sub>, OMe, OEt, CH<sub>2</sub>OMe

**R6, R7** = CH<sub>2</sub>OH, CMe<sub>2</sub>OH, C(O)Me, C(O)CF<sub>3</sub>, C(O)CHMe<sub>2</sub>, C(O)CHMeEt, C(O)CHEt<sub>2</sub>

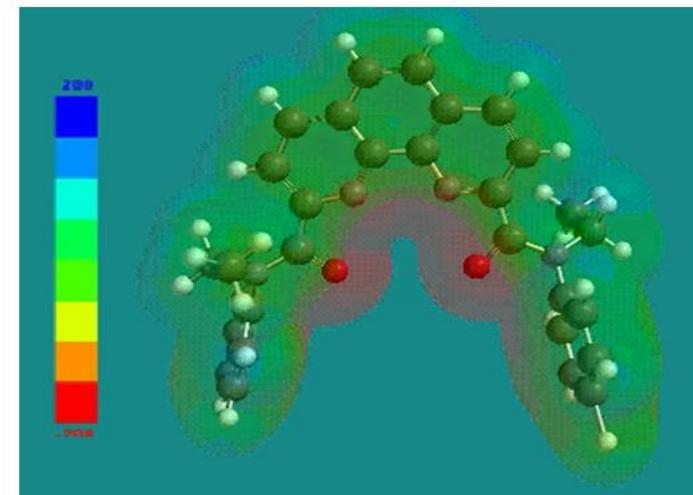
# Моделируемые свойства перспективных экстрагентов

На селективность экстрагента оказывают влияние 2 фактора:

1. **Основной вклад** в связь между Ln и экстрагентом вносит **электростатическая составляющая** - как это видно по карте распределения электростатического потенциала по молекуле.

Введение электронодонорных и электроноакцептных заместителей в гетероциклические ядра и к амидным атомам азота может в широких пределах менять **распределение электронной плотности**, и тем самым влиять на селективность экстрагента.

2. Варьируя заместители в положении **X** и **изменяя размер координационной полости**, можно менять **степень переноса электронной плотности с экстрагента на металл и на селективность образования комплекса**.



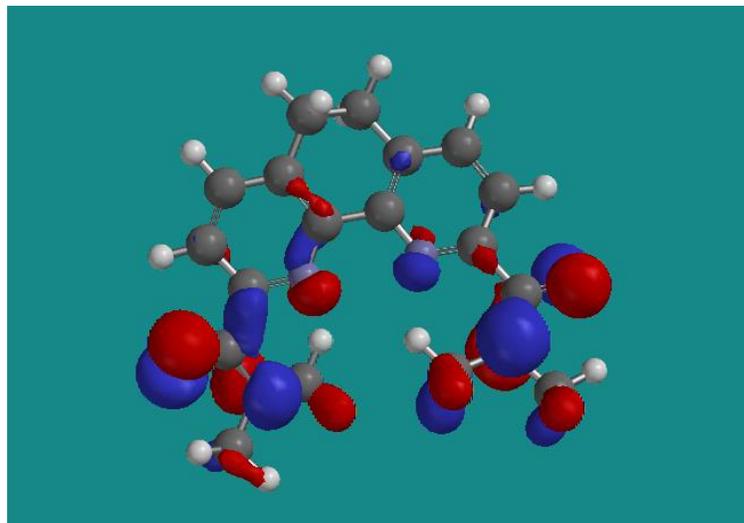
В качестве характеристик, определяющих свойства экстрагентов и комплексов, осуществляли расчет и моделирование:

1. энергии связывания экстрагента с катионом лантанида
2. Энергии граничных орбиталей (ВЗМО и НСМО) молекул экстрагентов .

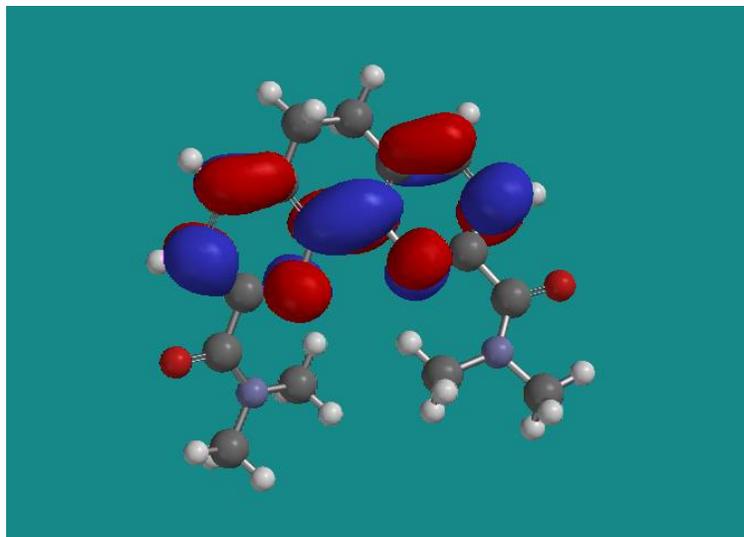
X	Расстояние между атомами N-N, Å
-CH=CH-	2.763
-CH <sub>2</sub> CH <sub>2</sub> -	2.766
>C=O	3.083
>Cl <sub>2</sub>	3.070
-O-	3.222

# Прогнозирование энергии граничных орбиталей

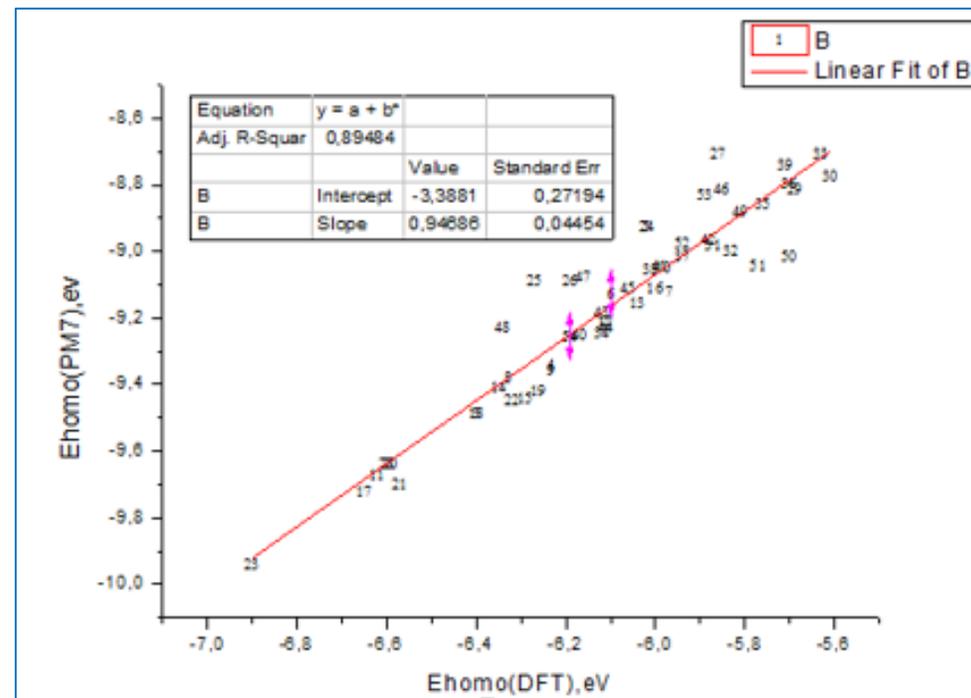
## Сопоставление результатов расчета *ab initio* и метода PM7



ВЗМО



НСМО



Зависимость  $E_{\text{взмо}}(\text{PM7})/E_{\text{взмо}}(\text{DFT})$

Квадрат коэффициента корреляции  $R_{\text{corr}} = 0.89$   
N = 54 соединения

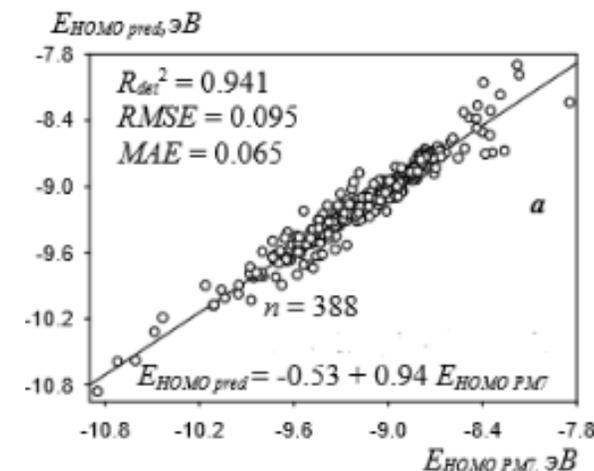
# Консенсусные модели “структура-свойство” для прогнозирования энергий ВЗМО

Построены QSPR модели с высокой прогнозирующей способностью.

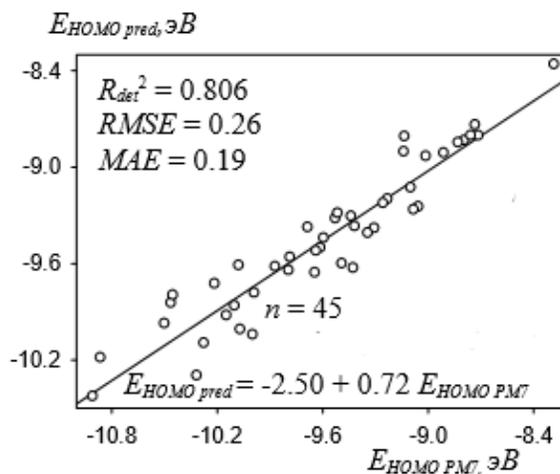
## База данных:

388 экстрагентов,  
результаты расчета  $E(\text{ВЗМО})$   
полуэмпирическим методом  
PM7

Сходство структур обучающей  
и контрольной выборок по  
Индексу Танимото **0.48**



Обучающая выборка 388  
экстрагентов,  
5-кратный скользящий  
контроль,  
 $R_{\text{det}}^2$  0.941,  $RMSE$  0.095 эВ



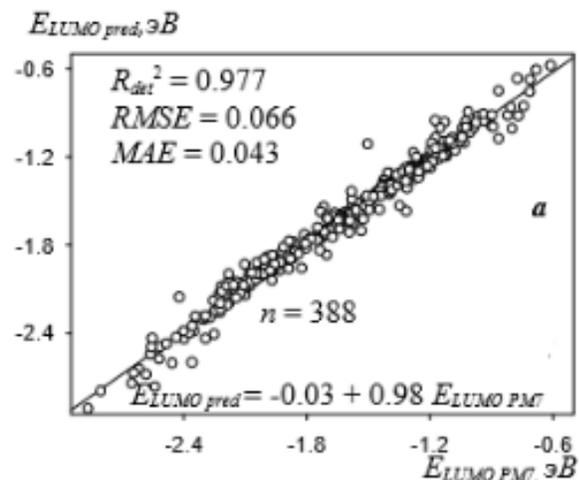
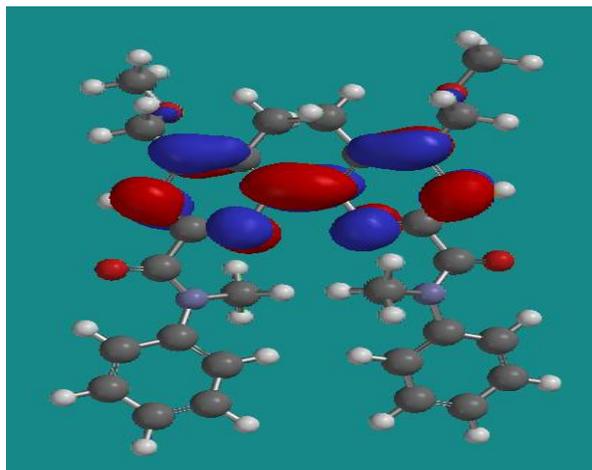
Внешняя независимая  
контрольная выборка 45  
экстрагентов,  
 $R_{\text{det}}^2$  0.806,  $RMSE$  0.26 эВ

# Консенсусные модели “структура-свойство” для прогнозирования энергий НСМО

Построены QSPR модели с высокой прогнозирующей способностью

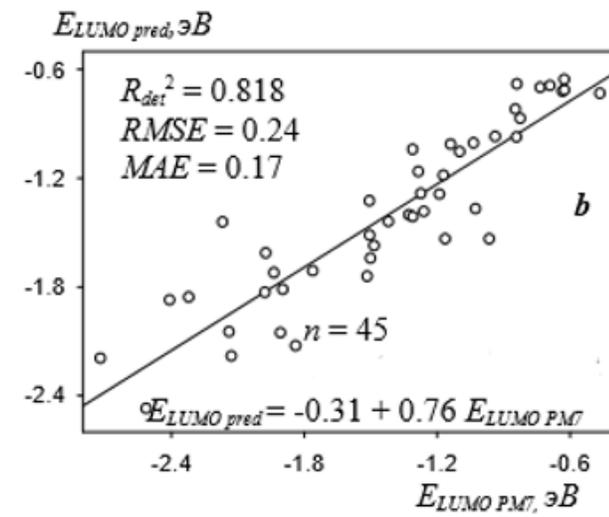
## База данных:

388 экстрагентов,  
результаты расчета  $E(\text{ВЗМО})$   
полуэмпирическим методом  
PM7



Обучающая выборка  
388 соединений

5-кратный скользящий  
контроль:  
 $R_{det}^2$  0.977  
 $RMSE$  0.066 эВ



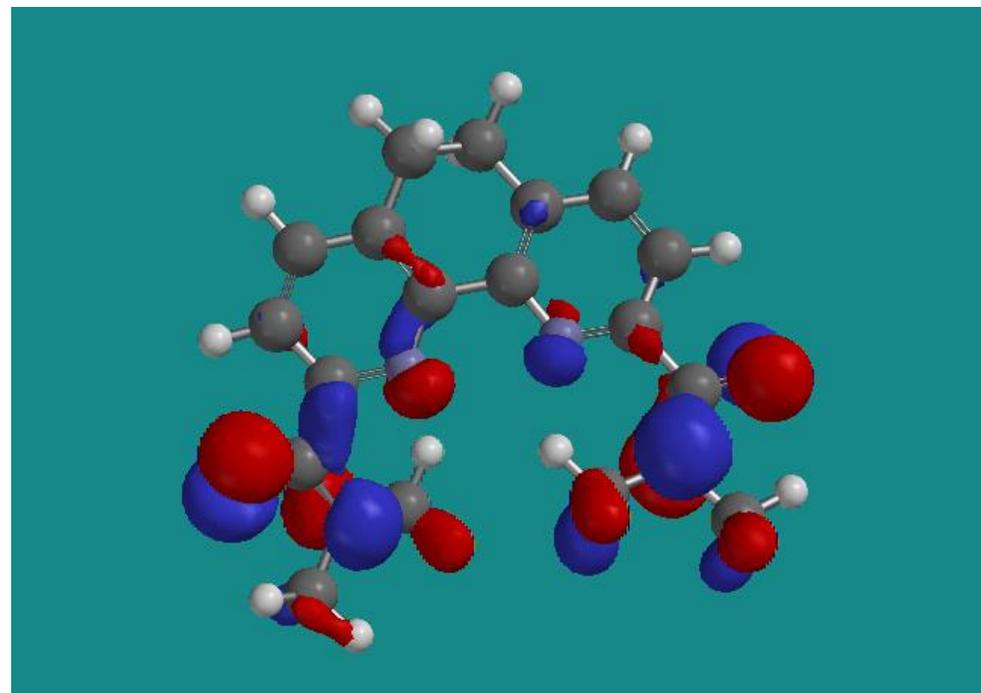
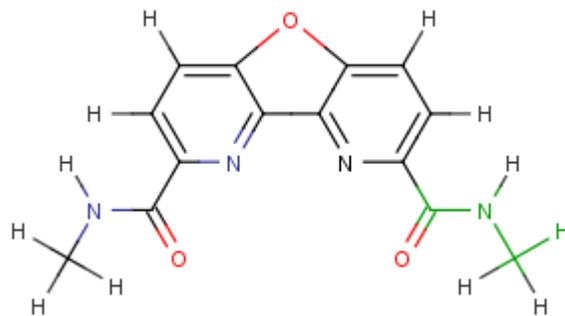
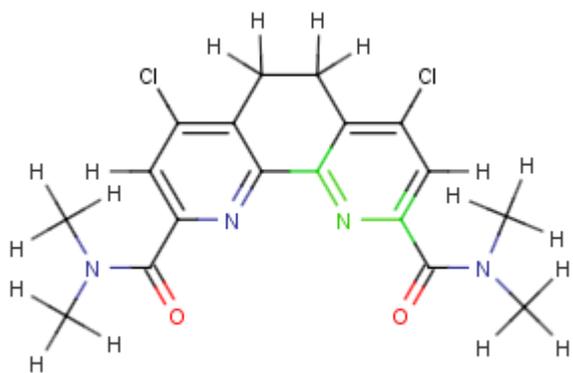
Внешняя независимая  
выборка 45  
соединений

$R_{det}^2$  0,818  
 $RMSE$  0.24 эВ

# Роль структурных фрагментов.

## 1. Вклад в модели для расчета значение энергии ВЗМО

Фрагменты, вносящие наибольший вклад в значение энергии ВЗМО

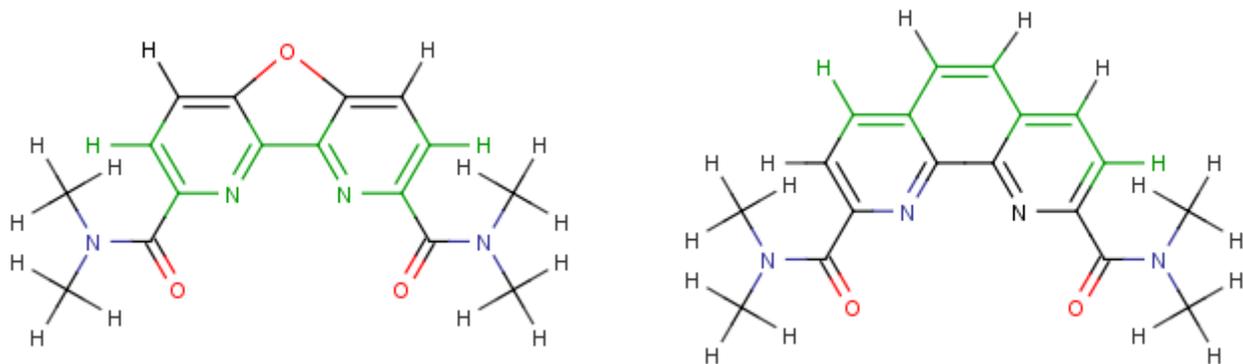


Вклад фрагмента, eV	Диапазон изменения, eV	Число индивидуальных моделей	Число соединений
-3.03	2.0	277	388
-0.81	0.08	85	295

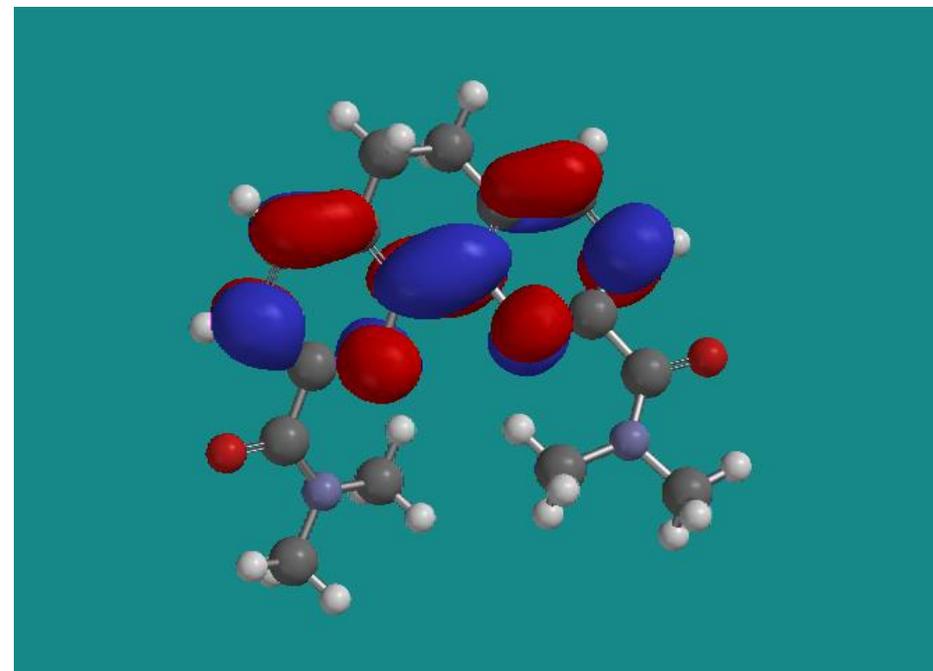
# Роль структурных фрагментов.

## 2. Вклад в модели для расчета энергии НСМО

Фрагменты, вносящие наибольший вклад в значение энергии НСМО



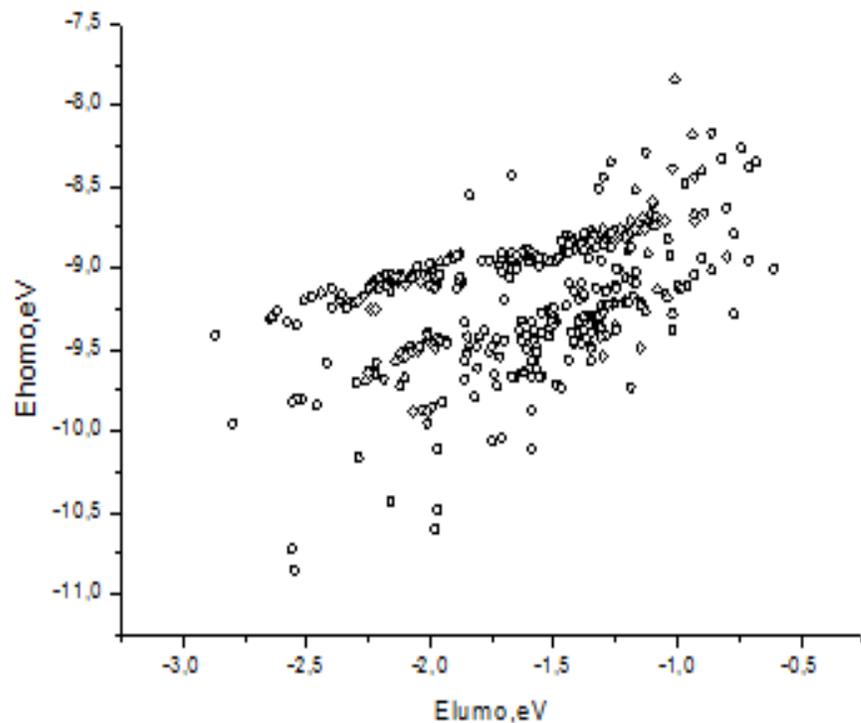
Вклад фрагмента, eV	Диапазон изменений, eV	Число индивидуальных моделей	Число соединений
-0.42	0.09	40	145
-0.44	0.22	333	58



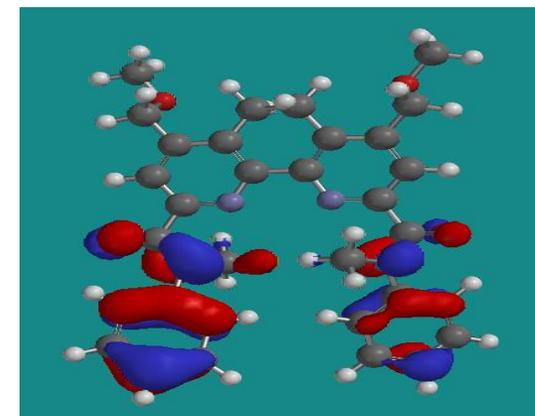
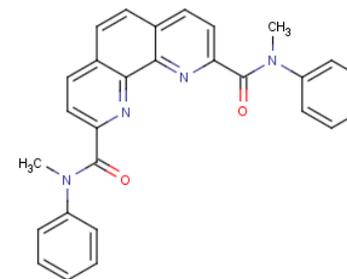
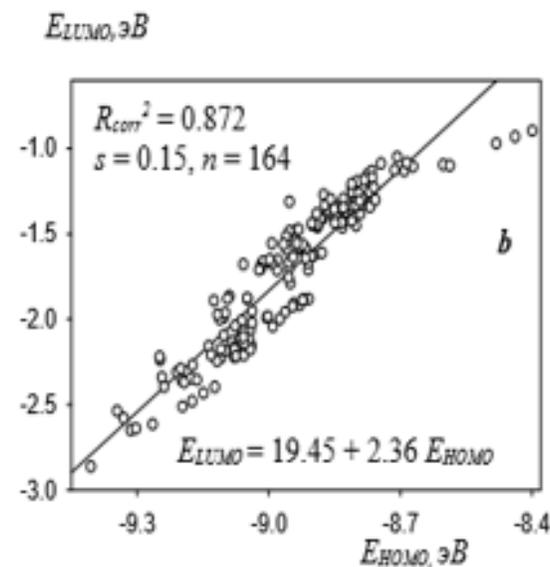
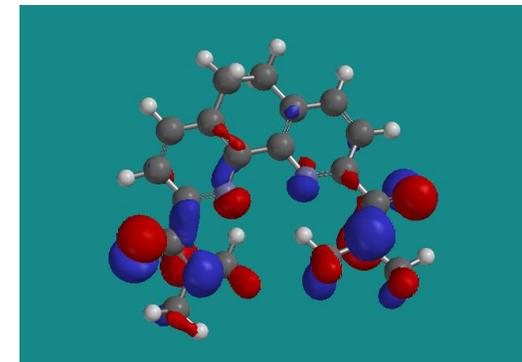
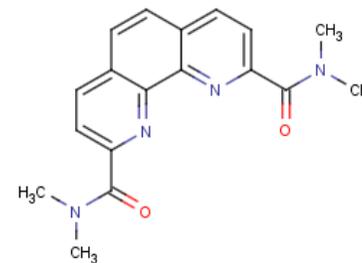
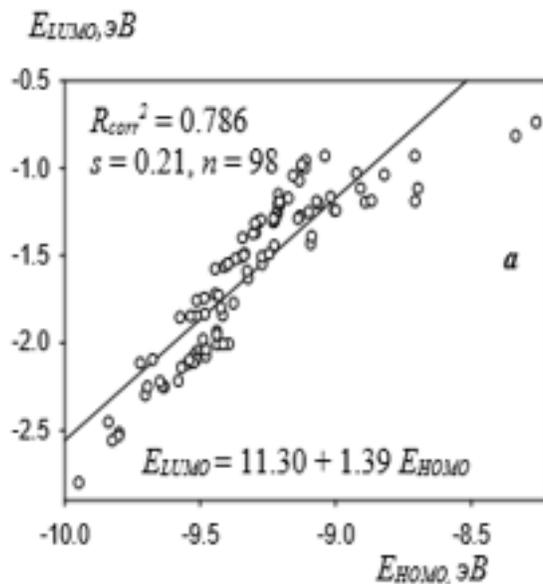
# Роль структурных фрагментов.

## 3. Влияние на $E_{\text{ВЗМО}} / E_{\text{НСМО}}$

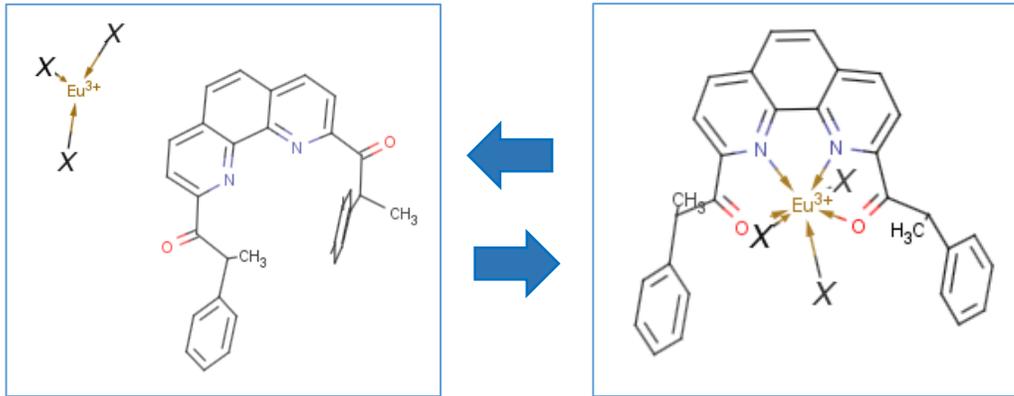
Зависимость  $E_{\text{ВЗМО}} / E_{\text{НСМО}}$



Общее  $N$  структур -388



# Расчет энергии связывания при образовании комплексов $\text{Ln}^{3+}$ с органическими экстрагентами



$$K = \frac{[\text{LLnX}_3]}{[\text{L}][\text{LnX}_3]}$$

$$\Delta G_1^0 = -RT \ln K_1$$

$$\Delta G_2^0 = -RT \ln K_2$$

$$SF_{\text{Ln1}/\text{Ln2}} = \frac{K_1}{K_2} = e^{-\frac{\Delta\Delta G}{RT}}$$



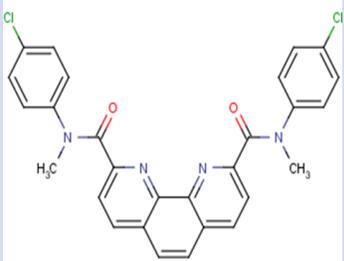
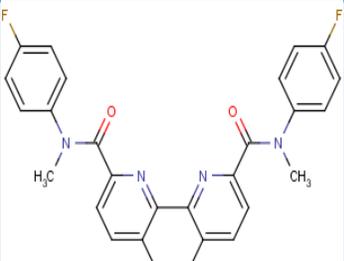
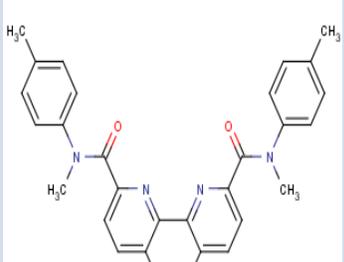
$$\Delta G^0 = \Delta H - T \Delta S$$



$$SF_{\text{Ln1}/\text{Ln2}} = e^{-\frac{\Delta\Delta E}{RT}}$$

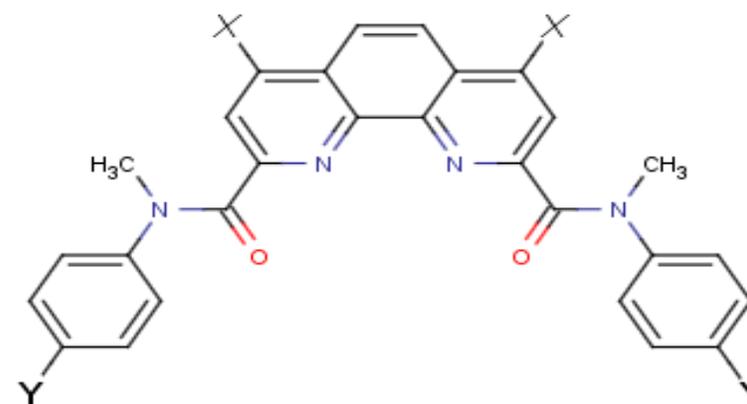
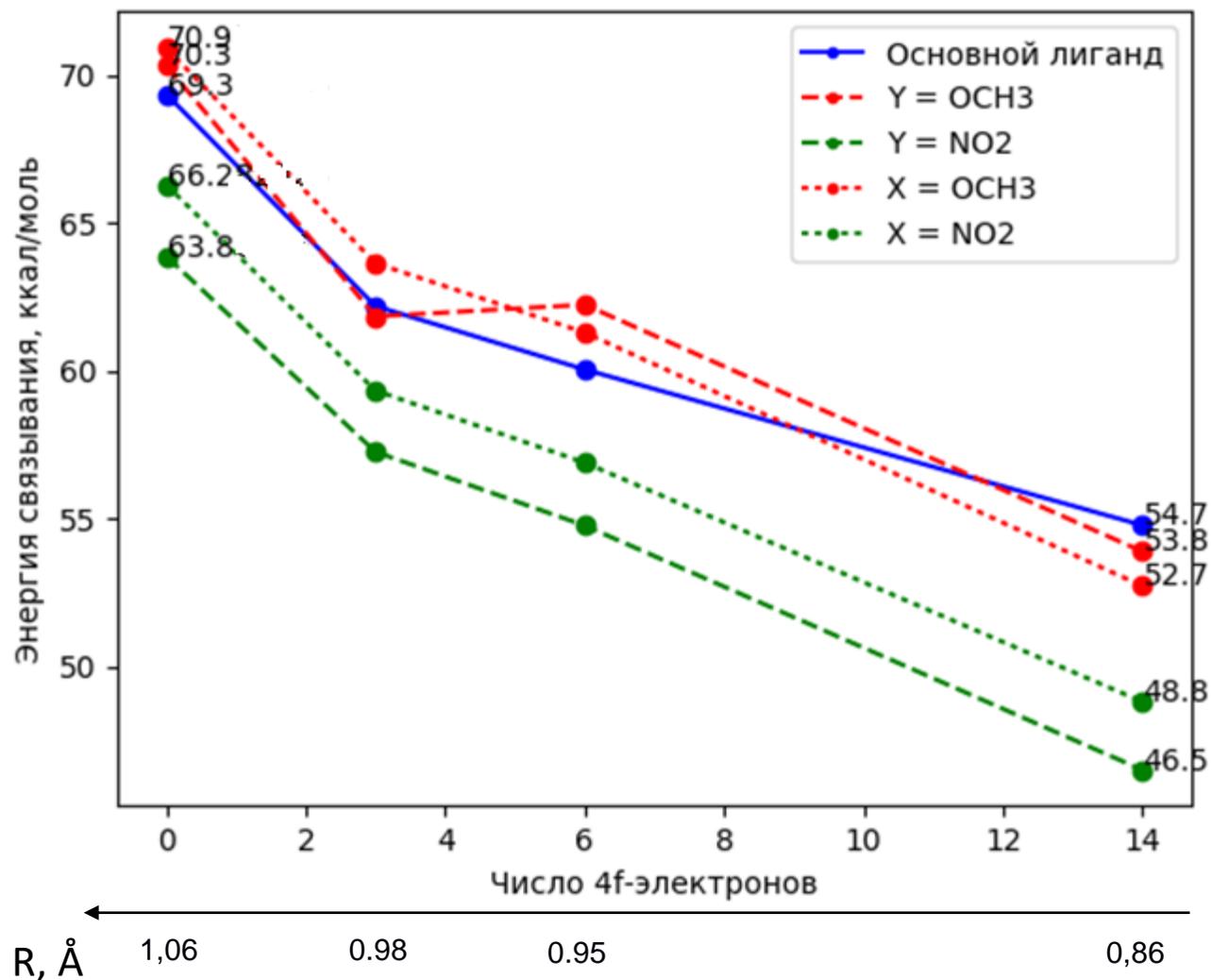
$$E_{\text{свЯЗ}} = E_{\text{LLnX}_3} - E_{\text{L}} - E_{\text{LnX}_3}$$

# Фрагмент комбинаторной библиотеки экстрагентов с результатами *ab initio* расчета энергий связывания и ВЗМО (ВЗЛУР, 6-31G\*)

Структура экстрагента	$E_{\text{связ}La}$ (Хартри)	$E_{\text{связ}Nd}$ (Хартри)	$E_{\text{связ}Eu}$ (Хартри)	$E_{\text{связ}Lu}$ (Хартри)	ВЗМО (эВ)
	-0.10791	-0.09672	-0.09337	-0.08003	-0.21
	-0.11035	-0.09922	-0.09557	-0.09629	-0.21
	-0.11176	-0.10025	-0.09686	-0.0884	-0.20

# Влияние структуры лиганда на энергию связывания “металл - лиганд” в комплексах.

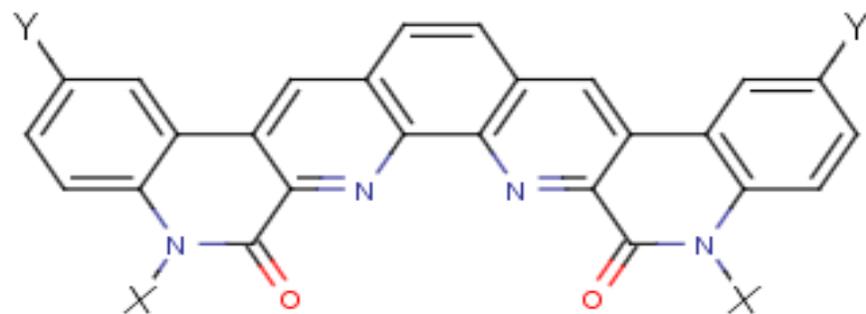
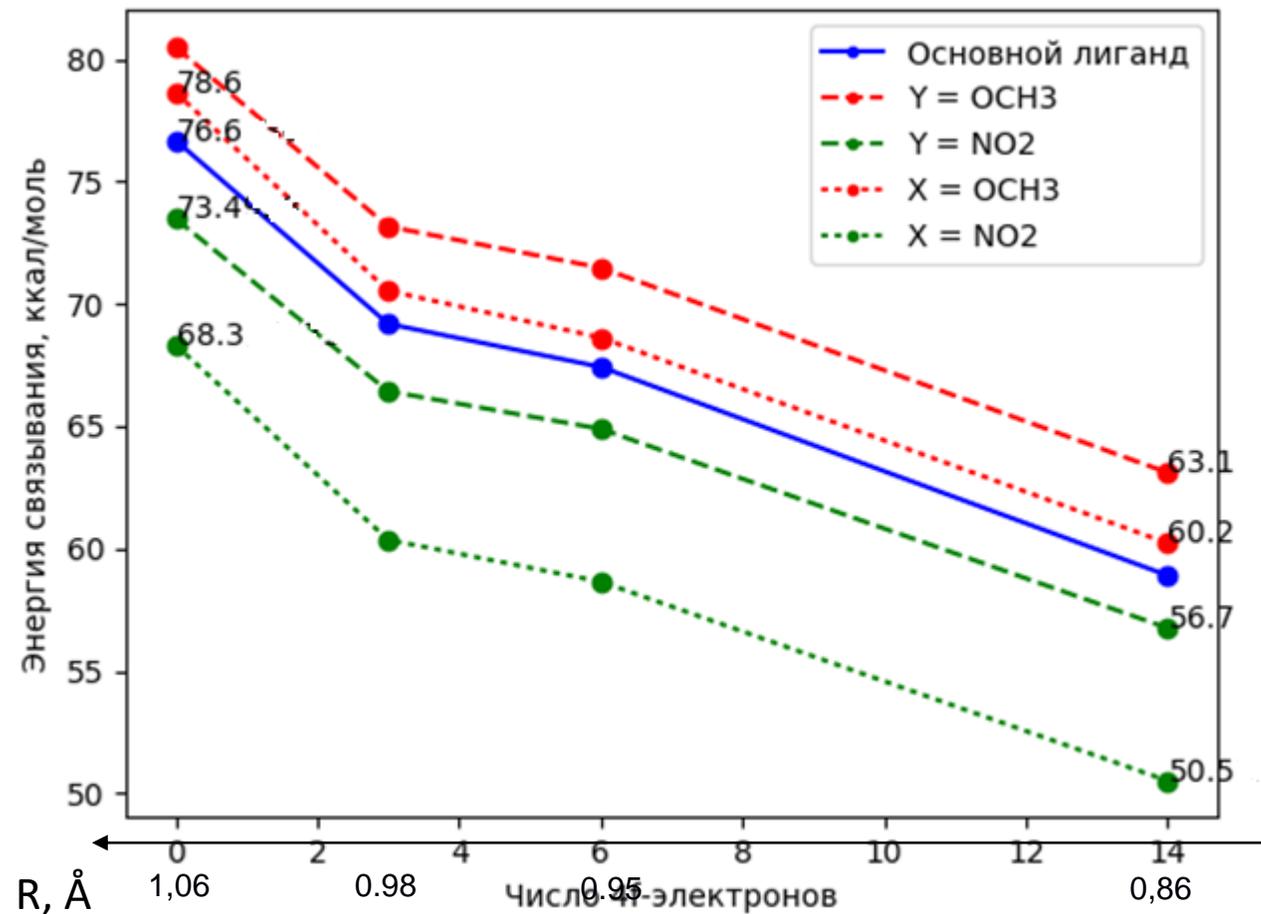
## 1. Зависимость $E_{\text{связ}}$ комплексов производных фенантролинов от природы металла



X	Y	$\Delta E_{\text{связ}}$ , ккал/моль
H	H	-14,6
H	OCH <sub>3</sub>	-16,4
H	NO <sub>2</sub>	-17,4
OCH <sub>3</sub>	H	-18,2
NO <sub>2</sub>	H	-17,5

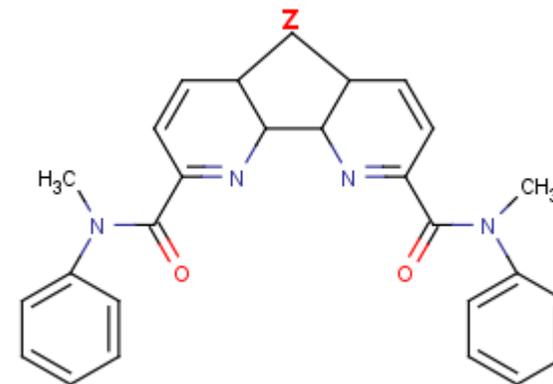
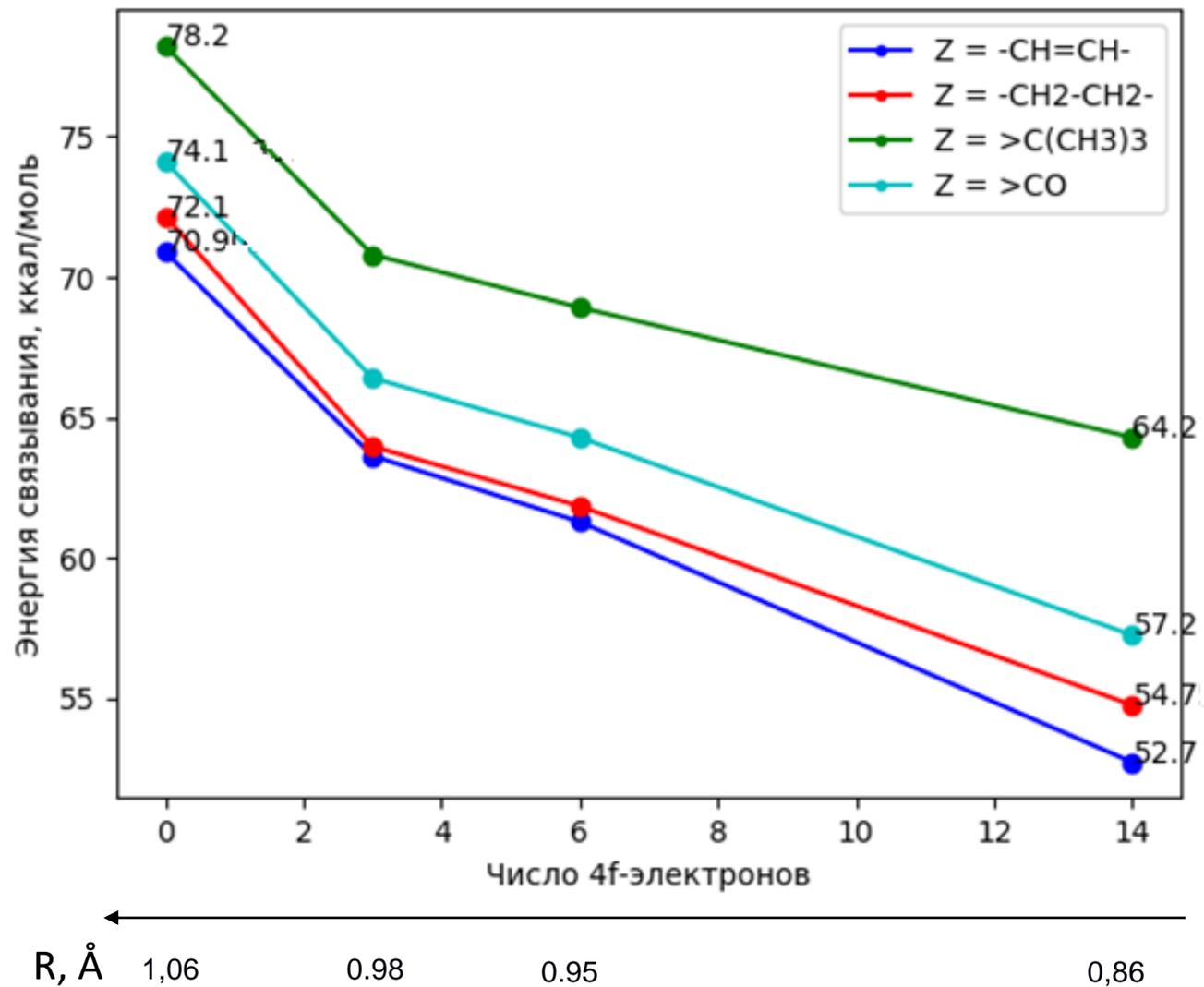
**OCH<sub>3</sub> – донорный заместитель**  
**NO<sub>2</sub> – акцепторный заместитель**

## 2. Зависимость $E_{\text{связ}}$ комплексов производных дилактамов фенантролинов от природы металла (для стабильных лигандов)



X	Y	$\Delta E_{\text{связ}}$ , ккал/моль
H	H	-17,8
H	OCH3	-17,4
H	NO2	-16,8
OCH3	H	-18,4
NO2	H	-17,8

### 3. Зависимость $E_{\text{связ}}$ комплексов производных N –гетероциклов от природы металла (наличие СО сопряжения)

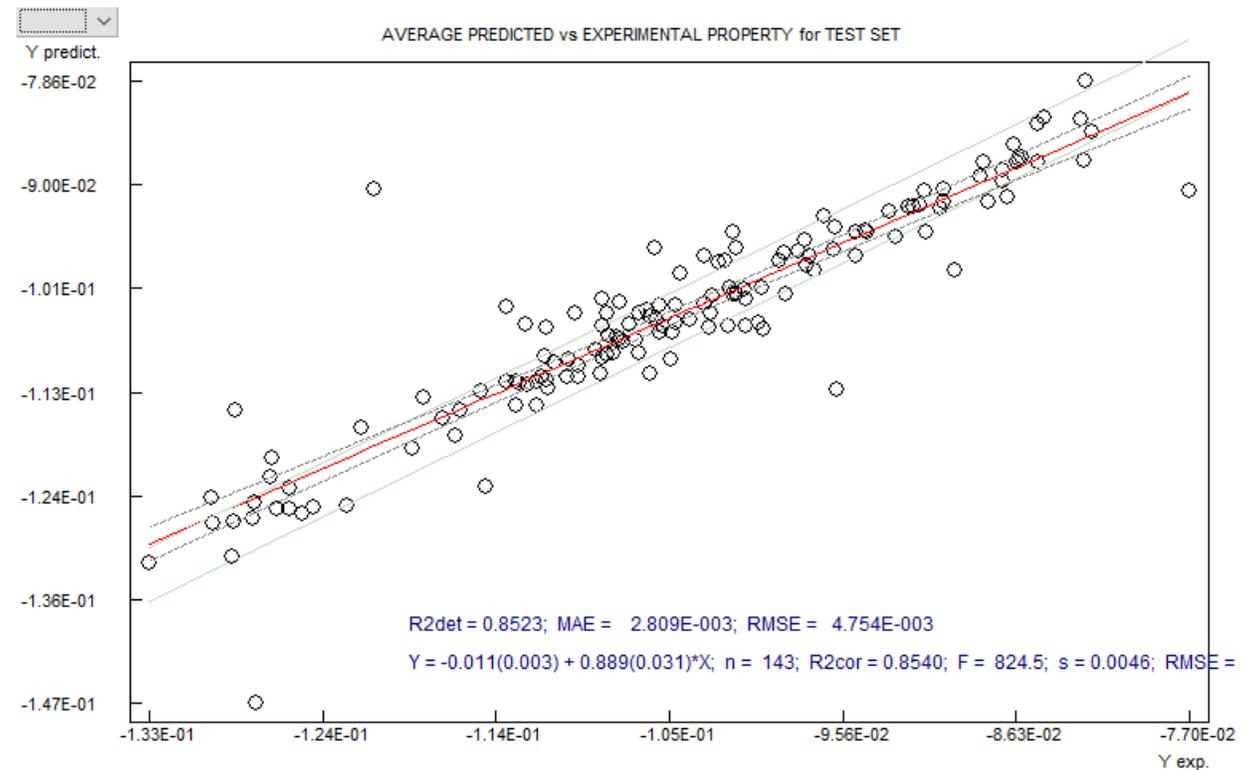


Z	Расстояние между атомами N-N, Å	$\Delta E_{\text{связ}}$ , ккал/моль
-CH=CH-	2.763	-17.3
-CH2-CH2-	2.766	-17.4
>C(CH3)2	3.083	-13.9
>CO	3.070	-16.9

# Прогнозирование $E_{\text{связ}}$ комплексов на основе консенсусных моделей “структура-свойство”

Базы данных – 149 экстрагентов и их комплексов с  $\text{La}^{3+}$ ,  $\text{Eu}^{3+}$ ,  $\text{Lu}^{3+}$ , результаты расчета методом функционала плотности, B3LYP, 6-31G\*

катион	$R^2$	RMSE ккал/моль
$\text{La}^{3+}$	0.852	3.0
$\text{Eu}^{3+}$	0.847	3.2
$\text{Lu}^{3+}$	0.848	3.2



Выводы:

1. С помощью *ab initio* и полуэмпирических квантово-химических методов, а также метода статистического моделирования “структура-свойство” с использованием результатов этих расчетов исследованы закономерности влияния молекулярной структуры лиганда на энергию связывания металл-лиганд в комплексах и построены модели “структура-свойство” для прогнозирования свойств перспективных экстрагенов для разделения РЗЭ на основе полидентатных N-гетероциклических лигандов и их комплексов состава  $LMe(NO_3)_3$  с катионами лантанидов.

2. На основании результатов расчетов предложены рекомендации синтетической группе ИНЭОС РАН по получению новых производных фенантролиновых лигандов. Ряд новых лигандов синтезированы и в настоящий момент проходят экспериментальное тестирование.

С благодарностью профессору Устыньюку Ю.А. (химический факультет МГУ) за неоценимую помощь в работе.