

Компьютерные методы в фармакологии



В.Б.Сулимов

НИВЦ МГУ

Лекция № 4

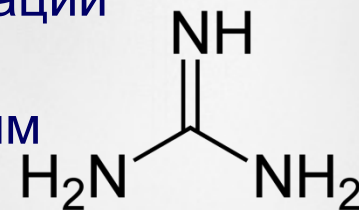
Силовое поле MMFF94, программа докинга SOL

Силловые Поля – Force fields

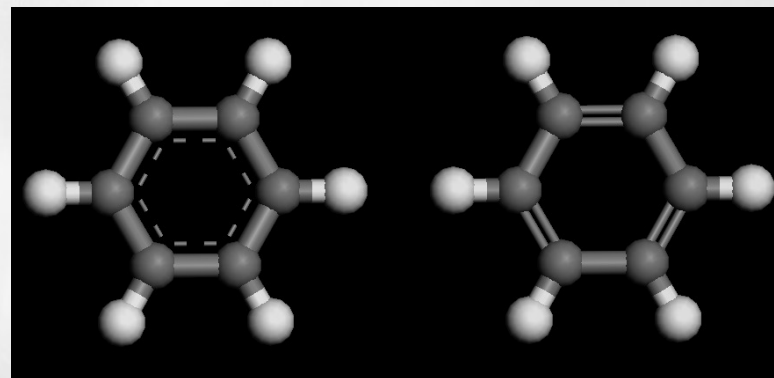
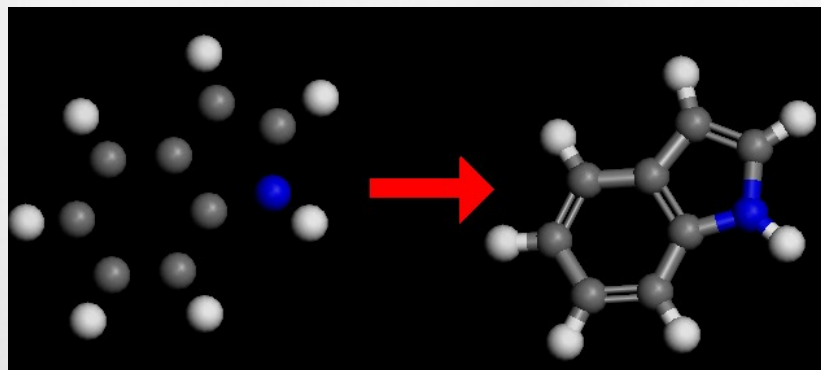
- MMFF94 – *ab initio* force field
 - На основе **квантовой механики**
 - Валентные связи
 - Валентные углы
 - Торсионы – вращения вокруг связей
 - Межмолекулярное взаимодействие
 - Электростатические взаимодействия
 - Ван дер Ваальсовы взаимодействия

MMFF94

- Типизация атомов – ключ к расстановке потенциалов:
- Тип атома задает параметры, определяющие потенциалы:
 - Заряды, радиусы Ван-дер-Ваальса, силовые константы и др.
- Три уровня типизации:
 - 1-ый уровень: положение в Таблице Менделеева (только для водорода воды – вводится уже на этом уровне свой тип); **27 типов**.
 - 2-ой уровень: учет типов и числа связей данного атома (для каждой возможной гибридизации атома C вводится свой тип); различие атомов H, связанных с полярными (O, N, P, S) и неполярными (C) атомами; 1-ый и 2-ой уровни типизации составляют **45 типов**.
 - 3-ий уровень: принадлежность атома к ароматическим подсистемам; более детальная типизация в случае принадлежности атома к особым группам (пептиды, гуанидин, и др.); 1-ый, 2-ой и 3-ий уровни типизации составляют **99 типов**.



Матрица связности



Для каждого атома указываются все атомы, с которыми он образует химические связи, а также порядок связи (одинарная, двойная, тройная)

Матрица связности необходима для визуализации структуры, а также проведения моделирования с использованием силовых полей при расчете энергии деформации связей, валентных и торсионных углов

Форматы записи

HIN-формат

MRK-формат

MOL-формат

SDF-формат

MOPAC-формат

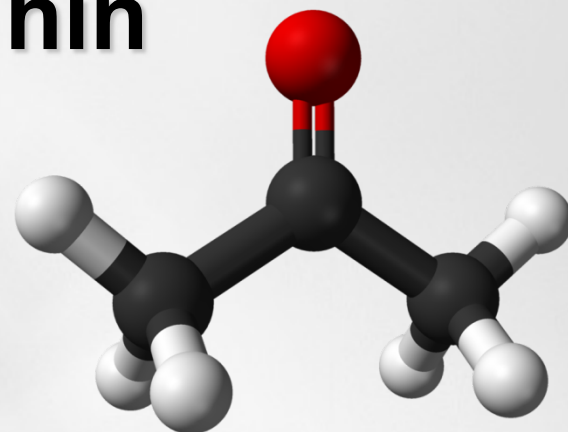
MOL2-формат

PDB-формат

XYZ-формат

Форматы записи: hin

ацетон



transform

mrk2hin

mol 1 "acetone"

atom	1 - H	**	- 0	0.048	-7.886	6.899	1	2 s				
atom	2 - C	**	- 0	0.364	-8.782	6.899	4	3 s	4 s	1 s	5 s	
atom	3 - H	**	- 0	0.047	-9.230	7.675	1	2 s				
atom	4 - H	**	- 0	0.047	-9.230	6.123	1	2 s				
atom	5 - C	**	- 0	1.904	-8.783	6.899	3	6 d	2 s	7 s		
atom	6 - O	**	- 0	2.519	-8.250	5.977	1	5 d				
atom	7 - C	**	- 0	2.674	-9.450	8.054	4	8 s	9 s	10 s	5 s	
atom	8 - H	**	- 0	2.056	-9.811	8.679	1	7 s				
atom	9 - H	**	- 0	3.220	-10.148	7.710	1	7 s				
atom	10 - H	**	- 0	3.220	-8.804	8.486	1	7 s				
endmol 1												

Формат hin

«mol», «endmol», «res», «endres», «atom», «formalcharge».

```
;my comment
mol 1 "2 waters"
res 1 H2O 999 - X
atom 1 HW1 H 31 F 0.20523 14.59035 -1.59775 5.2244 1 2 s
atom 2 OW O 70 F -0.41047 14.30144 -0.86623 4.64987 2 1 s 3 s
atom 3 HW2 H 31 - 0.20523 15.07461 -0.81689 4.03482 1 2 s
endres 1
res 2 OOO 999 A Y
atom 4 OW O 70 M -0.41047 13.8672 -0.87086 1.3318 2 5 s 6 s
endres 2
atom 5 HW1 H 31 - 0.20523 13.29637 -1.60993 1.61034 1 4 s
atom 6 HW2 H 31 - 0.20523 13.35378 -0.1264 1.73101 1 4 s
endmol 1
mol 2 "1 water"
atom 1 OW O 70 - -0.41047 16.19888 -0.91569 2.68214 2 2 s 3 s
formalcharge 1 0
atom 2 HW1 H 31 - 0.20523 16.97204 -0.74453 2.11552 1 1 s
atom 3 HW2 H 31 - 0.20523 15.47243 -0.89699 2.01081 1 1 s
endmol 2
```

benzol

XXaplite 12319923593D_1 0.00000 0.00000 1

This file was generated by apLite v0.43

12 12_0 0_0 0_0 0_0 0999 V2000

-6.3533	0.0252	0.1414	C	0_0	0_0	0_0	0_0	0_0	0_0	0_0	0_0	M	-0.1500	37
-6.3298	1.4147	-0.0202	C	0	0	0	0	0	0	0	0	M	-0.1500	37
-5.1675	2.1341	0.2780	C	0_0	0_0	0_0	0_0	0_0	0_0	0_0	0_0	M	-0.1500	37
-4.0288	1.4639	0.7378	C	0	0	0	0	0	0	0	0	M	-0.1500	37
-4.0522	0.0745	0.8995	C	0	0	0	0	0	0	0	0	M	-0.1500	37
-5.2145	-0.6448	0.6013	C	0	0	0	0	0	0	0	0	M	-0.1500	37
-7.2525	-0.5313	-0.0892	H	0	0	0	0	0	0	0	0	M	0.1500	5
-7.2108	1.9331	-0.3759	H	0	0	0	0	0	0	0	0	M	0.1500	5
-3.1712	-0.4439	1.2552	H	0	0	0	0	0	0	0	0	M	0.1500	5
-5.2327	-1.7198	0.7264	H	0	0	0	0	0	0	0	0	M	0.1500	5
-5.1494	3.2090	0.1530	H	0	0	0	0	0	0	0	0	M	0.1500	5
-3.1296	2.0205	0.9685	H	0	0	0	0	0	0	0	0	M	0.1500	5

1 2 4 0_0 0_0
1 6 4_0 0_0 0
1 7 1 0 0 0 0
2 3 4 0 0 0 0
2 8 1 0 0 0 0
3 4 4 0 0 0 0
3 11 1 0 0 0 0
4 5 4 0 0 0 0
4 12 1 0 0 0 0
5 6 4 0 0 0 0
5 9 1 0 0 0 0
6 10 1 0 0 0 0

M END

Формат MOL

Форматы SDF

SDF-формат является расширением MOL-формата

позволяет записывать в один файл несколько молекул.

Формат полностью аналогичен MOL-формату, но

каждая молекула отделяется строчкой

\$\$\$\$

от других молекул.

MMFF94

$$U_{MMFF} = \sum EB_{ij} + \sum EA_{ijk} + \sum ESB_{ijk} + \sum EOP_{ijk;l} + \sum ET_{ijk;l} + \sum EVdW_{ij} + \sum EQ_{ij}$$

$$EQ_{ij} = 332.0716 \frac{q_i q_j}{R_{ij}} \quad - \text{электростатические взаимодействия}$$

$$EVdW_{ij} = \varepsilon_{IJ} \left(\frac{1.07 R_{IJ}^*}{R_{ij} + 0.07 R_{IJ}^*} \right)^7 \left(\frac{1.12 R_{IJ}^{*7}}{R_{ij}^7 + 0.12 R_{IJ}^{*7}} - 2 \right) \quad - \text{Ван дер Ваальс}$$

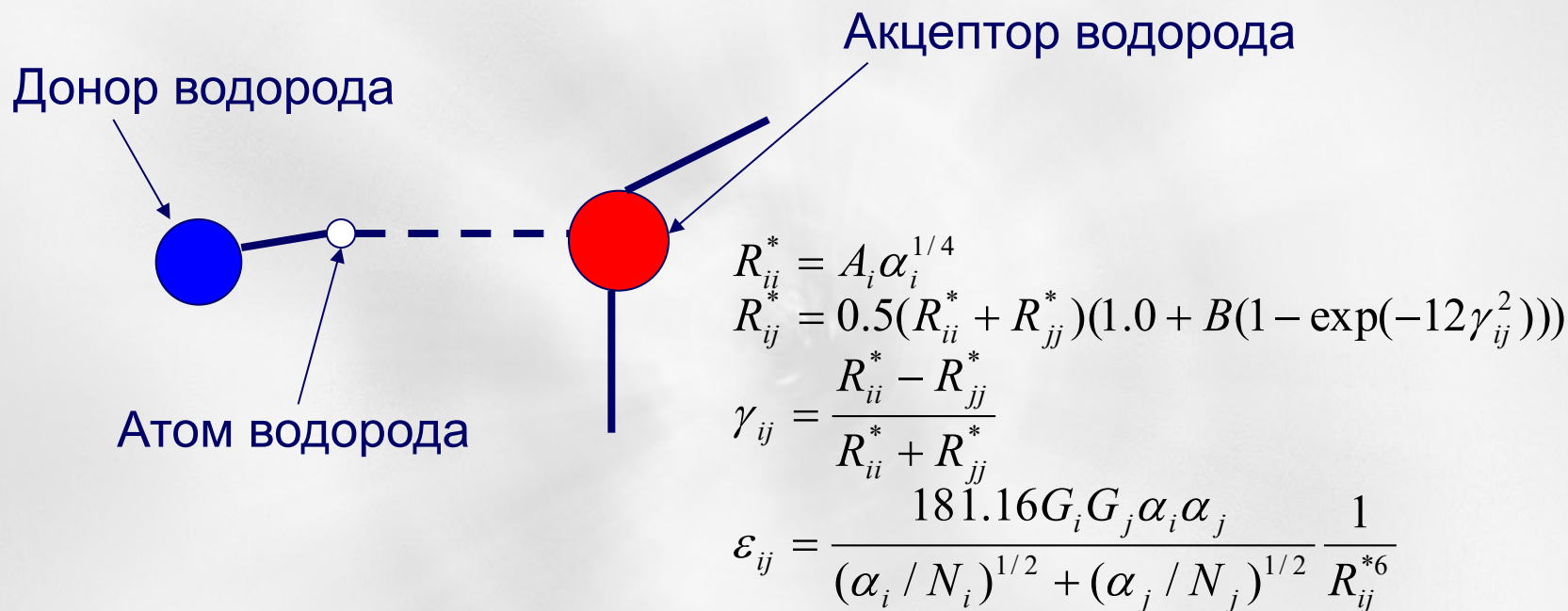
$$R_{ii}^* = A_i \alpha_i^{1/4}$$

$$R_{ij}^* = 0.5(R_{ii}^* + R_{jj}^*)(1.0 + B(1 - \exp(-12\gamma_{ij}^2)))$$

$$\gamma_{ij} = \frac{R_{ii}^* - R_{jj}^*}{R_{ii}^* + R_{jj}^*}$$

$$\varepsilon_{ij} = \frac{181.16 G_i G_j \alpha_i \alpha_j}{(\alpha_i / N_i)^{1/2} + (\alpha_j / N_j)^{1/2}} \frac{1}{R_{ij}^{*6}}$$

Водородные связи



Если атом i и (или) j является донором (акцептором) водородной связи, то $B=0$,
а величины R_{ij}^* и ϵ_{ij} умножаются на 0.8 и 0.5, соответственно.

MMFF94

Энергия деформации валентных связей

$$EB_{ij} = 143.9325 \frac{kb_{ij}}{2} \Delta r_{ij}^2 \left(1 + cs \Delta r_{ij} + 7/12 cs^2 \Delta r_{ij}^2 \right)$$

$$\Delta r_{ij} = r_{ij} - r_{ij}^{(0)} \quad cs = -2 \text{ \AA}^{-1}$$

Энергия деформации валентных углов

$$EA_{ijk} = 0.043844 \frac{ka_{ijk}}{2} \Delta \vartheta_{ijk}^2 \left(1 + cb \Delta \vartheta_{ijk} \right)$$

$$\Delta \vartheta = \vartheta_{ijk} - \vartheta_{ijk}^{(0)} \quad cb = -0.4 \text{ rad}^{-1}$$

MMFF94

Изменение энергии, связанное с одновременной деформацией и длин связей и валентных углов

$$ESB_{ijk} = 2.51210(kba_{ijk}\Delta r_{ij} + kba_{kji}\Delta r_{kj})\Delta\vartheta_{ijk}$$

Изменение энергии, связанное с деформацией торсионного угла φ

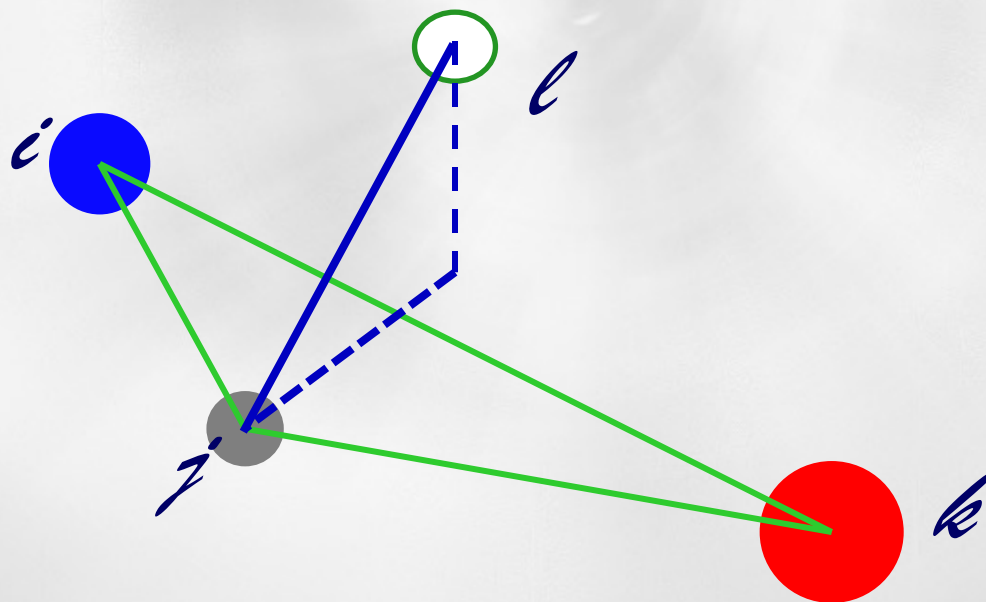
$$ET_{ijkl} = 0.5(V_1(1 + \cos\varphi) + V_2(1 + \cos 2\varphi) + V_3(1 + \cos 3\varphi))$$

MMFF94

Изменение энергии из-за выхода атома ℓ из плоскости
трех атомов: i , j и k

$$E_{OOP_{ijk;l}} = 0.043844 \frac{koop_{IJK;L}}{2} \chi_{ijk;l}^2$$

$\chi_{ijk;l}$ - угол между прямой, проходящей через атомы j и ℓ , и
плоскостью трех атомов i, j, k



MMFF94

Есть еще некоторые нюансы параметризации этого силового поля. Например:

Правило 1-3: внутри одной молекулы несвязевые взаимодействия, т.е. электростатические и Ван-дер-Ваальсовы взаимодействия, между данным атомом и теми, которые с ним образуют химическую связь, или валентный угол или торсионный угол, **не учитываются**.

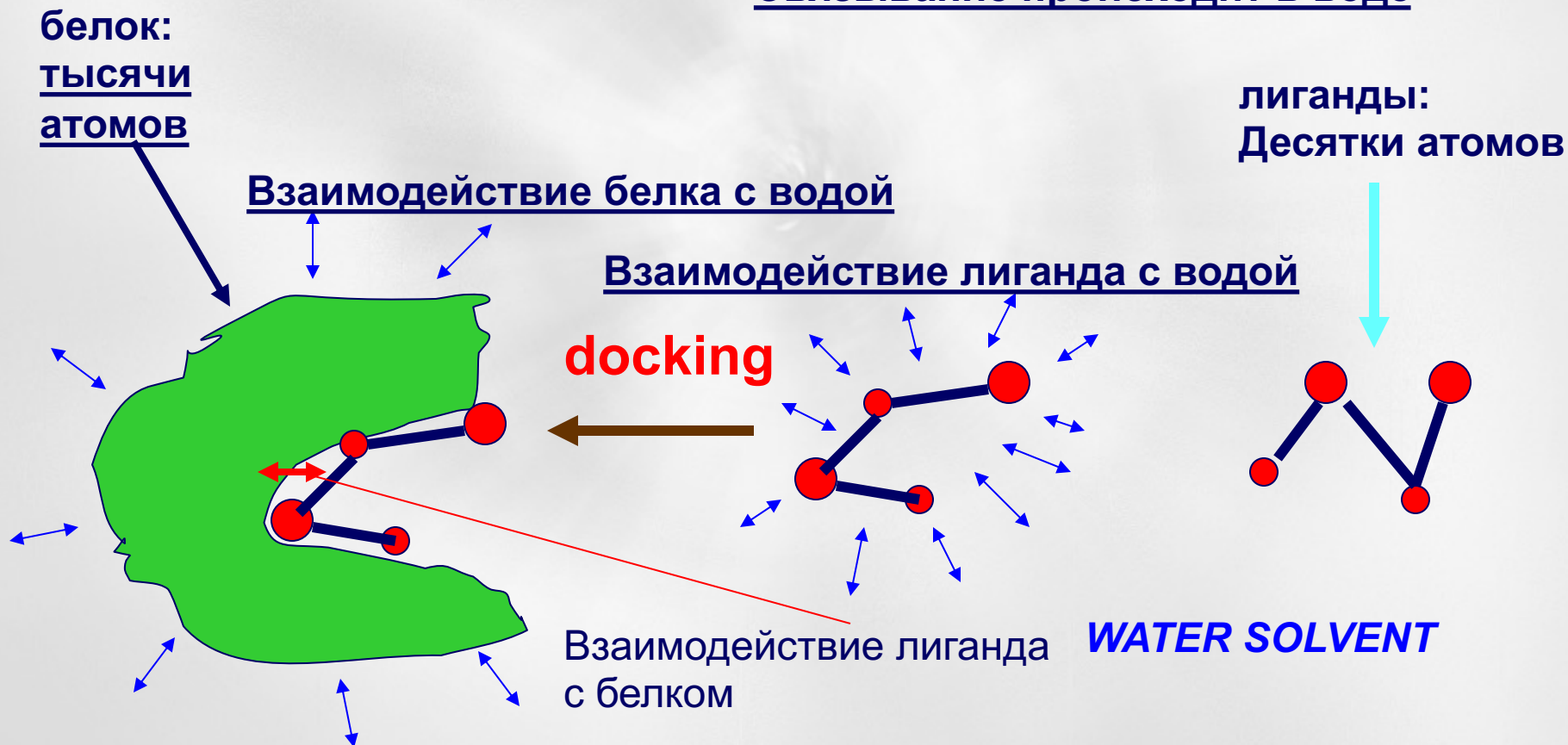
Правило 1-4: при расчете электростатического взаимодействия вклад от атомов, отстоящих от данного на четыре химические связи, **умножается на 0.75**

Scoring – правильная оценка энергии связывания белок-лиганд

Слабые
межмолекулярные
взаимодействия

Высокая точность: ~ 1 kcal/mol ~ 0.05 eV

Связывание происходит в воде



Подготовка белка для докинга

1. Начальная стадия подготовки белка
2. Восстановление структуры белка – «штопка дыр»
3. Упрощение белка
4. Подготовка белка к типизации
5. Расстановка атомов водорода
6. Типизация атомов
7. Определение области докинга

Программа APLITE

Скор: оценка энергии связывания белок-лиганд

компоненты энергии:

- Электростатическое взаимодействие белок-лиганд
- взаимодействие Ван-дер-Ваальса белок-лиганд
- Энергия десольватации:
молекула-растворитель (ES & VdW) & кавитация
- Энергия внутренних напряжений лиганда
- Торсионные степени свободы

$$\Delta G_{\text{bind}} = \alpha(\Delta G_{\text{ES}} + \Delta G_{\text{RF}}) + \beta \Delta G_{\text{VdW}} + \lambda \Delta G_{\text{NP}} + \mu \Delta G_{\text{LIS}} + \Delta G_{\text{TORS}} + C$$

Физический скор

H.-J. Böhm, 1994

- ΔG_{ES} – прямое электростатическое взаимодействие лиганд-белок
- ΔG_{VdW} – прямое VdW взаимодействие лиганд-белок
- $\Delta G_{\text{RF}} = G_{\text{RF}}(\text{PL}) - [G_{\text{RF}}(\text{P}) + G_{\text{RF}}(\text{L})]$ – электростатический вклад в энергию десольватации:
- $G_{\text{RF}}(\text{P})$ – электростатический вклад во взаимодействие белка с водой
- $\Delta G_{\text{NP}} = G_{\text{NP}}(\text{PL}) - [G_{\text{NP}}(\text{P}) + G_{\text{NP}}(\text{L})]$ – VdW вклад в энергию десольватации
- $G_{\text{NP}}(\text{P})$ – VdW вклад во взаимодействие белка с водой
- $G_{\text{NP}}(\text{P}) = B \Sigma_{\text{SAS}}(\text{P}) + A$, $\Sigma_{\text{SAS}}(\text{P})$ – Поверхность белка доступная растворителю
- ΔG_{LIS} – энергия внутренних напряжений лиганда
- ΔG_{TORS} - свободная энергия связанная с потерей степеней свободы лигандом при связывании в активном центре белка

Скор – оценка энергии связывания белок-лиганд

$$G_{\text{bind}} = \alpha(\Delta G_{\text{ES}} + \Delta G_{\text{RF}}) + \beta \Delta G_{\text{VdW}} + \lambda \Delta G_{\text{NP}} + \mu \Delta G_{\text{LIS}} + \Delta G_{\text{TORS}} + C$$

$\alpha, \beta, \lambda, \mu, C$ – подгоночные параметры
Обучающий набор:

- подобрать подгоночные параметры так, чтобы хорошо воспроизвести энергию связывания белок-лиганд для набора комплексов.

После завершения подгонки параметры не меняются

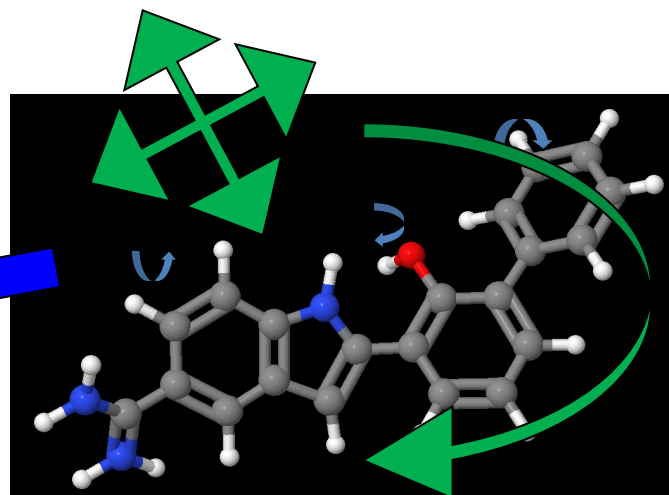
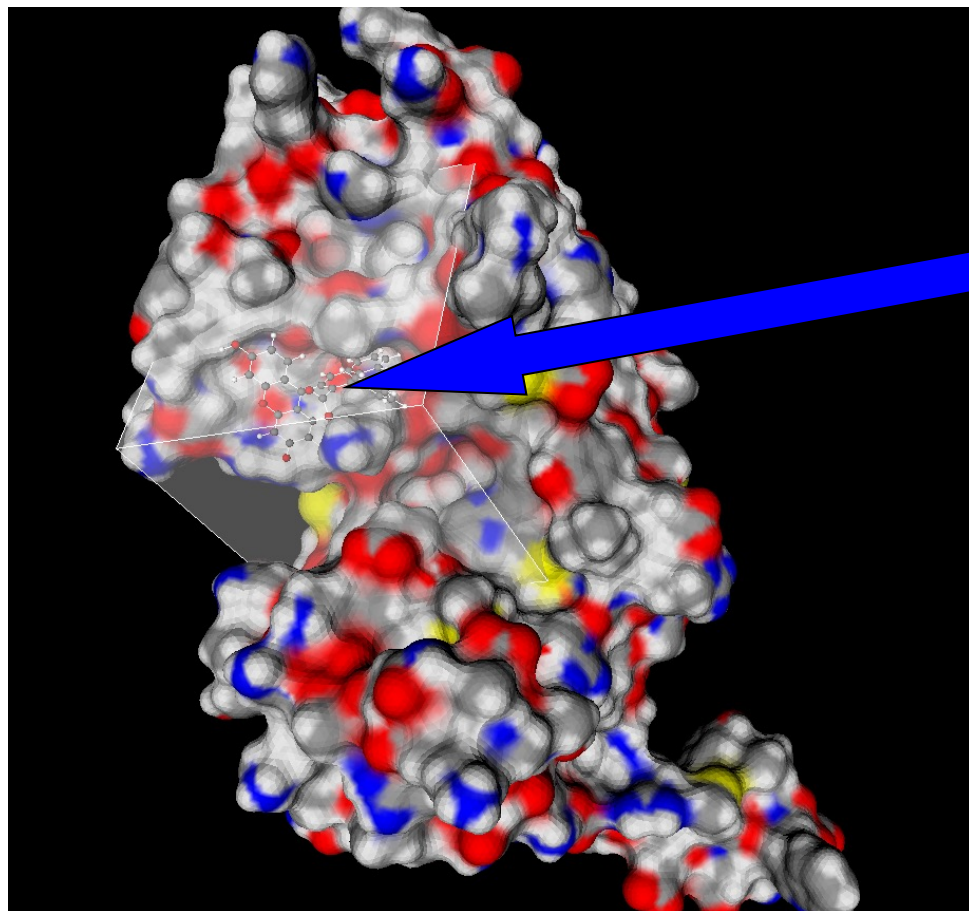
Валидация сора:

- Воспроизвести экспериментально измеренную энергию связывания белок-лиганд из валидационного набора комплексов, которые не входят в обучающий набор

Оригинальная программа докинга SOL

- **Позиционирование лиганда в активном центре заданного белка-мишени**
- **Генетический алгоритм поиска глобального минимума – модель эволюции**
- **Заранее вычисляется набор 3х-мерных сеток потенциалов, представляющих жесткий белок (SOLGRID)**
- **Гибкий лиганд $\lesssim 15$ внутренних вращательных степеней свободы**
- **Универсальное силовое поле MMFF94 – классические потенциалы для описания междатомных взаимодействий**
- **Учет десольватации - подсчет баланса взаимодействий: лиганда с растворителем, белка с растворителем, комплекса белок-лиганд с растворителем**
- **Учет внутренних напряжений лиганда**
- **Точность позиционирования: 0.5-2.0 Å – близкая к AutoDock**
- **Validation: see <http://keenbase.ru>**

Программы докинга: позиционирование молекул в активном центре белка-мишени



Десятки тысяч молекул:

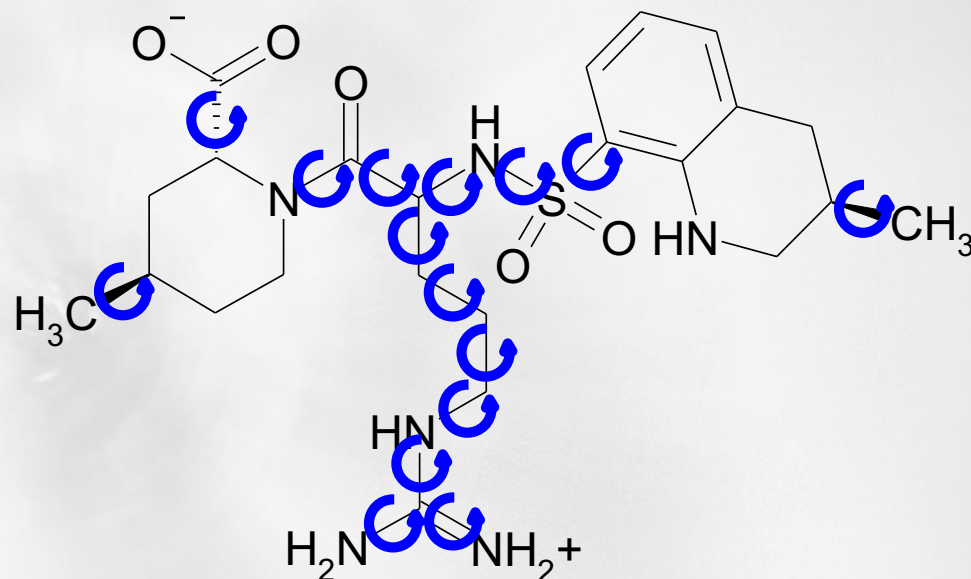
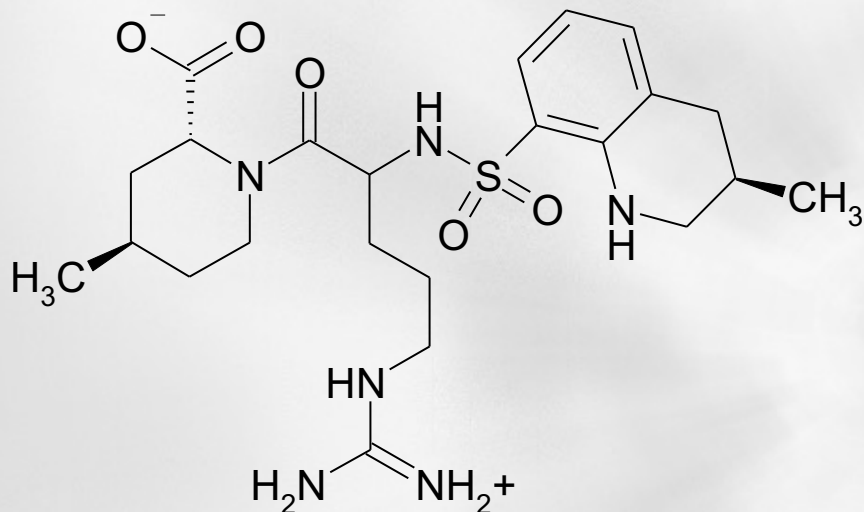
- 3 трансляции как целое тело
- 3 вращения как целое тело
- ≤ 15 внутренних вращений

Белок: жесткий

Открытие ингибиторов: тромбина, урокиназы (uPA), факторов свертываемости крови Xa, XIa, Mpro SARS-CoV-2

Модуль SOL: найти глобальный минимум энергии на сложной многомерной энергетической поверхности

Типичный лиганд – ингибитор тромбина – Аргатробан



Получаем задачу на нахождение глобального минимума в пространстве $15+6=21$ измерений!

Систематический перебор невозможен

Докинг – задача поиска глобального минимума энергии

- на многомерной поверхности >10 измерений
- с большим числом локальных минимумов
- Только полный перебор гарантирует нахождение глобального минимума
- Для докинга это невозможно:
 - Пусть 15 измерений
 - По каждому измерению 10 точек
 - Всего 10^{15} точек – 10^{15} вычислений энергии связывания

Эвристические и термодинамические методы

- Не гарантируют нахождение глобального минимума
- Позволяют решать задачу за меньшее число шагов
- Нет 100% уверенности в том, что найденный минимум - глобальный

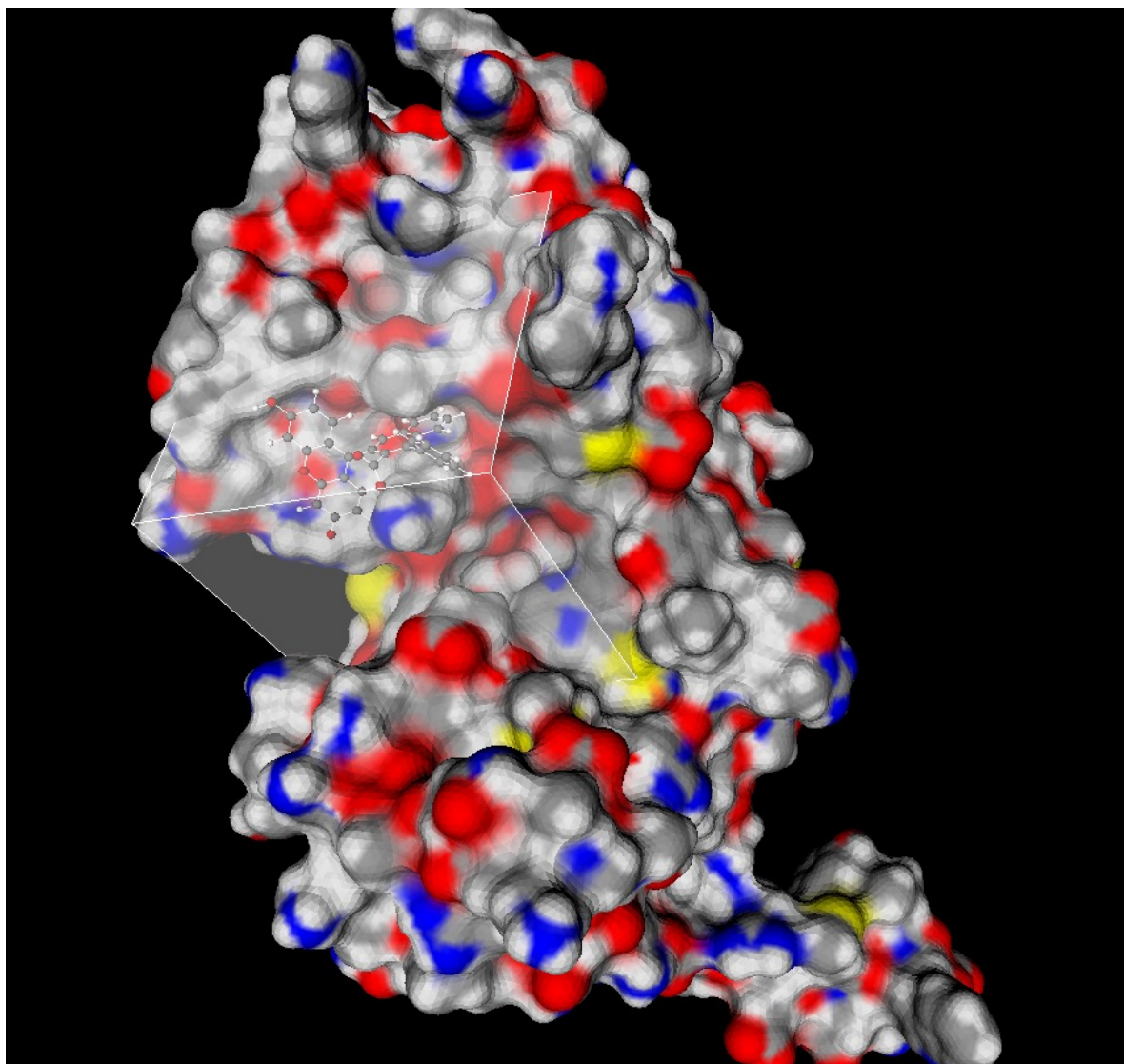
Модель белка – программа SOLGRID

- Белок жесткий и представлен сеткой потенциалов
- Сетка – это набор «сеток» потенциалов для разных типов пробного атома лиганда
- Разные типы взаимодействий пробного атома лиганда и белка:
 - Ван-дер-ваальсовы взаимодействия
 - Кулоновские взаимодействия
 - Энергия десольватации

$$E_{lig-protein} = \sum_{m=1}^N E_m, E_m = E0_m + E1_m + E2_m$$

- Область докинга – куб размера 101 x 101 x 101 точек, длина ребра куба 22 Å – длину ребра куба можно менять

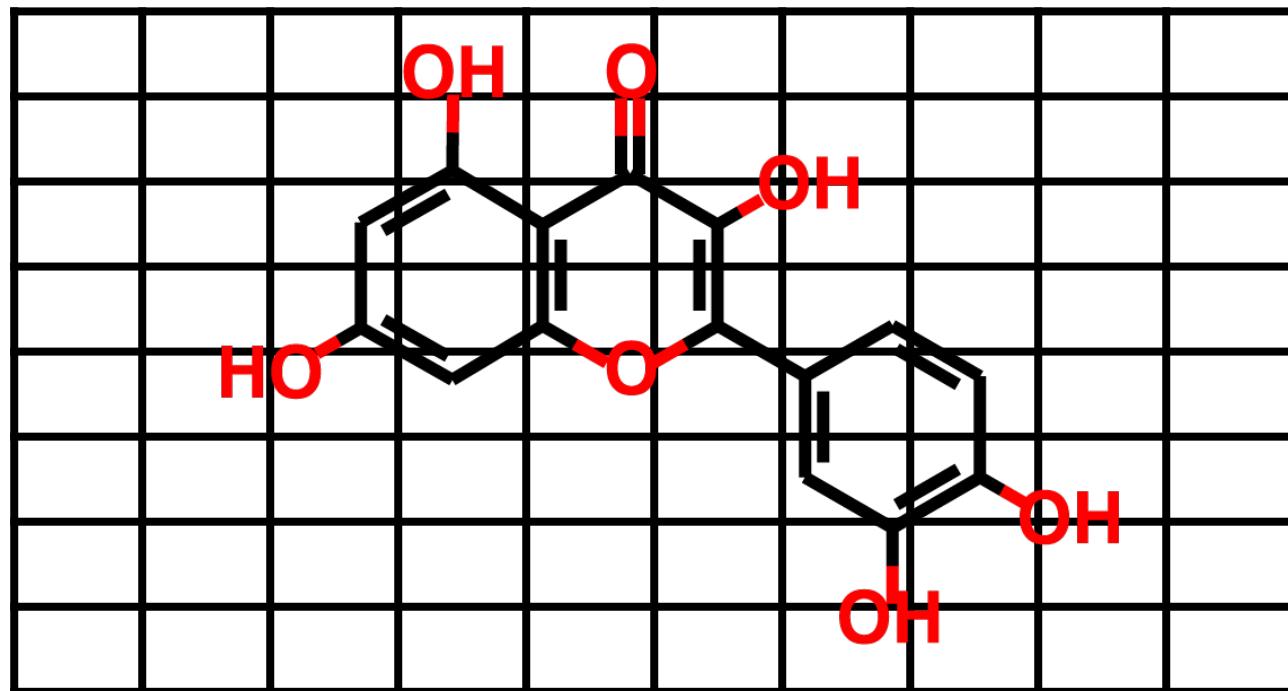
Программа SOLGRID



Вычисление энергии лиганда в поле белка

$$E_{lig-protein} = \sum_{i=1}^N E_i$$

N – Количество атомов в лиганде
 E_i – энергия i -ого атома в поле белка



Если атом попадает между узлов сетки, то его энергия вычисляется в результате интерполяции по значениям энергии в **8 ближайших узлах**

Все потенциалы взаимодействия атомов лиганда с белком для всех типов атомов (C, N, H, S, O ...) во всех узлах 3х-мерной сетки, покрывающей весь активный центр белка-мишени

Решения задачи докинга

- ▶ Приготовление сетки потенциалов, имитирующих белок-мишень – 3-5 часов – однократно.
- ▶ Докинг: на каждый лиганд надо 1-10 часов
- ▶ Поиск наиболее энергетически выгодных поз лиганда: энергия лиганда в поле белка + энергия внутренних напряжений лиганда
- ▶ Решения задачи докинга: позы лиганда в белке
- ▶ 50 независимых запусков Генетического Алгоритма дают 50 решений - 50 лучших поз лиганда: энергия наиболее отрицательная
- ▶ Совпадение решений - кластеризация

Модуль SOLGRID

- ▶ **Запуск модуля SOLGRID:**

solgrid par-file protein-file

par-file – это текстовый файл параметров с расширением .par

protein-file – файл белка в формате PDB с расширением .pdb

- ▶ **Обязательные, важные и дополнительные параметры**

CENTER X: 32.729

CENTER Y: 32.729

CENTER Z: 32.729

RADIUS: 11.0

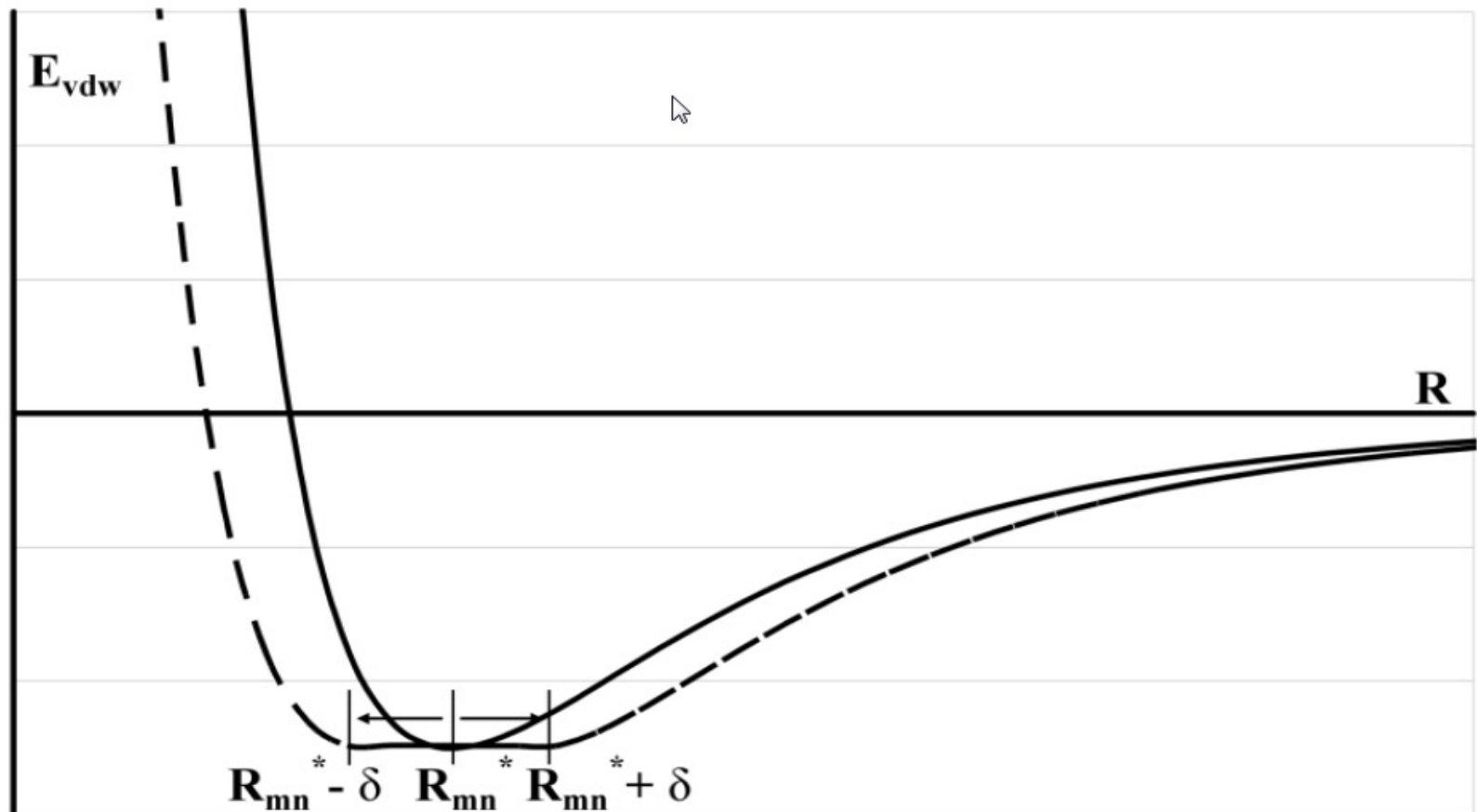
BROADING: 0.4

- ▶ **Модуль работает в однопроцессорном режиме 2-4 часа, в многопроцессорном режиме 1-2 минуты на 128 ядрах**

- ▶ **По завершении работы модуля SOLGRID создается бинарный файл сеток потенциалов – этот файл входной для модуля SOL**

Уширение VDW потенциала: параметр BROADING

Параметр δ называется уширением и задается в файле *par-file* входных параметров модуля SOLGRID.



Спасибо за внимание

- *...Surely every medicine is an innovation; and he that will not apply new remedies, must expect new evils...*
- *...Каждый медицинский метод есть инновация; а кто не хочет применять новые средства, должен ждать новых бед...*

Sir Francis Bacon (1561-1626)



OF INNOVATIONS