

КОМПЬЮТЕРНОЕ КОНСТРУИРОВАНИЕ

УДК 577.15.04

©Коллектив авторов

ПРЕДСКАЗАНИЕ СВЯЗЫВАЮЩЕЙ СПОСОБНОСТИ БЕЛОК-ЛИГАНДНЫХ КОМПЛЕКСОВ ПРИ ПОМОЩИ НЕЛИНЕЙНЫХ МОДЕЛЕЙ.

В.В. КРЕПЕЦ, Н.В. БЕЛКИНА, В.С. СКВОРЦОВ, А.С. ИВАНОВ

Институт биомедицинской химии им. В.Н. Ореховича РАМН.
Москва, 119832, Погодинская 10. эл. почта: krepets@ibmh.msk.su

Разработана нейросетевая модель для прогнозирования изменения свободной энергии Гиббса при образовании белок-белковых и белок-лигандных комплексов. В качестве обучающей выборки использованы данные о 150 комплексах различной природы, в качестве независимых переменных - различные расчетные значения их физико-химических параметров. Были построены как классическая модель множественной линейной регрессии, так и несколько нейронных сетей прямого распространения с одним скрытым слоем, из которых выбрана лучшая. Показано значимое улучшение показателей качества нейросетевой модели над классической моделью множественной линейной регрессии (R^2 настройки 0,81 и 0,54; R^2 при процедуре скользящего контроля 0,74 и 0,52 соответственно).

Ключевые слова: белок-лигандный комплекс, константа диссоциации, нейронная сеть.

ВВЕДЕНИЕ. В ходе работ по компьютерному конструированию и поиску в базах данных различных биологических эффекторов, одной из наиболее частых проблем становится предсказание аффинности лиганда к белку мишеня. В настоящей работе создана модель предсказания величины аффинности на основе расчетных значений физико-химических параметров комплексов молекул, полученных с использованием информации о трехмерном строении комплексов, и значений изменения свободной энергии Гиббса (ΔG), взятых из литературных источников. В качестве обучающего множества, использовалась выборка,

включающая в себя 150 белок-лигандных комплексов. В качестве простейшего варианта (для сравнения) использовалась классическая модель множественной линейной регрессии. Основной целью работы было построение нейросетевой модели для возможности прогнозирования ΔG .

МЕТОДИКА. Источник данных. Обучающая выборка была сформирована на основе отобранных данных из созданной в лаборатории базы данных KeyLock [1]. Основными критериями отбора служили: наличие для конкретного комплекса трехмерной структуры в Protein Data Bank (PDB) [2] и отсутствие в составе места связывания металлических ионов, таких как Cu, Fe, Zn (последнее обусловлено несовершенством процедуры расчета физико-химических параметров). Всего было отобрано 150 комплексов, из которых 70 представляли белок-белковые комплексы (молекулярный вес мишеней от 5000 до 60000 и лиганда от 5000 до 14000) и 80 - белок-лигандные (молекулярный вес лиганда от 120 до 1200) соединений. Среди них наиболее широко были представлены комплексы различных мутантов барназы с мутантами барстара (53), протеазы ВИЧ-1 (26), сериновых протеаз (13), эндотиапепсина (9) и тромбина (7). Исключая комплексы мутантов барназы и барстара, для всех остальных в PDB имеются файлы с трехмерными структурами: 121d, 1aaq, 1abe, 1abf, 1aht, 1apt, 1apu, 1bap, 1bit, 1bmm, 1bmn, 1cho, 1cse, 1d30, 1dbb, 1dbj, 1dbm, 1dds, 1dhf, 1dif, 1dwb, 1dwc, 1dwd, 1eas, 1eat, 1eed, 1ela, 1elb, 1elc, 1fss, 1hbv, 1hdt, 1hgt, 1hih, 1hiv, 1hos, 1hps, 1hpv, 1hpx, 1hsg, 1hte, 1htf, 1htg, 1hvi, 1hvj, 1hvk, 1hvl, 1hvr, 1hvs, 1lyb, 1nsd, 1ppc, 1ppk, 1ppl, 1pro, 1sbg, 1sbn, 1tpa, 2dbl, 2er0, 2er6, 2er7, 2er9, 2kai, 2ptc, 2sec, 2sni, 2tpi, 2tpi, 2upj, 3er3, 3hfl, 3sgb, 3sic, 3ssi, 4er1, 4er4, 4est, 4ins, 4phv, 4tpi, 5er2, 5sic, 6abp, 7abp, 7gch, 8abp, 9abp (ряд файлов содержит более одного лиганда).

Для всех этих комплексов средствами программы Sybyl [3] была проведена предварительная обработка по единой схеме:

1. Редактирование типов атомов низкомолекулярного лиганда, согласно схеме принятой в Sybyl.
2. Добавление атомов водорода, отсутствующих в файлах PDB в большинстве случаев.
3. Расстановка частичных атомарных зарядов с использованием комбинированной схемы Гастайгера-Хюккеля.
4. Формирование сольватационной оболочки, с использованием бокса содержащего воду, структура которой предварительно уравновешена молекулярной динамикой.
5. Минимизация полученной ячейки в ближайший глубокий локальный минимум (метод - сопряженных градиентов Пауэла, ограничение - не более 500 итераций).

Предварительную обработку комплексов мутантов барназы и барстара проводили по несколько иной схеме. Для них модели комплексов были созданы на основе данных о трехмерной структуре комплекса барназы и барстара дикого типа. Соответственно, была изменена схема подготовки комплексов:

1. Замена соответствующих аминокислотных остатков в последовательности дикого типа.
2. Добавление атомов водорода и расстановка частичных атомарных зарядов с использованием комбинированной схемы Гастайгера-Хюккеля.
3. Формирование сольватационной оболочки, с использованием ячейки

содержащей воду, структура которой предварительно уравновешена молекулярной динамикой.

4. Минимизация полученной ячейки в ближайший локальный минимум.

5. Молекулярная динамика в периодически-граничных условиях с сохранением полной энергии ячейки (используемые параметры: метод - Верле, шаг - 1 фс, время проведения - 5 пс, начальное состояние – замороженное: при таких условиях процедура молекулярной динамики вырождается в своеобразный метод глобальной минимизации, т.к. приращение кинетической энергии возможно только за счет увеличения потенциальной, естественно, по абсолютному значению).

6. Повторная минимизация, аналогичной п. 4.

Для всех полученных комплексов был проведен расчет ΔG из значений констант диссоциации (или ингибирования) по широко известной формуле:

$$\Delta G_{bind} = -RT \ln K_d$$

Независимые переменные. Для построения регрессионных уравнений по всем 150 комплексам был рассчитан следующий набор физико-химических параметров, который в дальнейшем мы будем обозначать как "независимые переменные":

1. ΔE_Q - величина энергии электростатического взаимодействия молекул.

2. ΔE_{VDW} - величина энергии вандерваальсовых взаимодействий.

3. ΔE_{HVDW} - величина энергии вандерваальсовых взаимодействий в отсутствие поправочного коэффициента. Эта компонента энергии требует дополнительного разъяснения; её вычисление становится возможным, благодаря особенностям поля сил Tripos, использующегося в Sybyl. В Sybyl, как и во многих других программах, величина энергии вандерваальсовых взаимодействий рассчитывается согласно потенциалу Ленарда-Джонса:

$$E_{VDW} = \sum_{i=1}^{N_{atoms}} \sum_{j>1} E_{ij} \left[\frac{1.0}{a_{ij}^{12}} - \frac{2.0}{a_{ij}^6} \right]$$

где - E_{ij} - вандерваальсова константа для пары атомов i-j, a_{ij} принимает значение $r_{ij}/(R_i+R_j)$, r_{ij} - расстояние между i-тым и j-тым атомом, R_i - вандерваальсов радиус i-того атома. Однако в поле сил Tripos проводят косвенный учет водородного связывания, вводя соответствующий коэффициент у R_i для пар атомов, способных образовать водородную связь (по умолчанию в Sybyl 0.7). Таким образом, вычисляя величину ΔE_{VDW+} без использования коэффициента водородного связывания, можно вычислить:

$$\Delta E_{HVDW} = \Delta E_{VDW+} - \Delta E_{VDW}$$

косвенно характеризующую величину энергии водородного связывания.

4. **MWT** - молекулярный вес молекулы мишени.

5. **MWL** - молекулярный вес молекулы лиганда.

6. **CHT** - суммарный заряд молекулы мишени.

7. **CHL** - суммарный заряд молекулы лиганда.

8. ΔG_A - неполярный компонент сольватационного члена энергии взаимодействия по Хортону и Льюису [4].

9. ΔG_P - полярный компонент сольватационного члена энергии взаимодействия по Хортону и Льюису. ΔG_A и ΔG_P вычисляются по общей формуле:

$$\Delta G_{A(P)} = \sum_{i=1}^{N_{atoms}} \Delta \sigma_i (A_i^0 - A_i^S)$$

где A_i^S - площадь поверхности i-го атома доступной растворителю в несвязанном состоянии, A_i^0 - площадь поверхности i-го атома доступной растворителю в комплексе, $\Delta \sigma_i$ - атомный параметр сольватации для "молекулярной поверхности". Полярный компонент вычисляется для атомов, образующих водородные связи и солевые мостики, неполярный - для остальных.

Предварительный статистический анализ выборки. Понятие независимых переменных, известное из курса математики, подразумевает равные нулю (близкие к нулю для конечного количества наблюдений) попарные корреляции. Рассмотрим корреляционную матрицу, рассчитанную на основании всех 150 наблюдений. Из корреляционной матрицы, приведенной в таблице 1, видно наличие выраженных корреляций между различными "независимыми" переменными (выделено жирным шрифтом), что в сущности делает понятие "независимых переменных" не вполне корректным. Та же корреляционная матрица показывает, что зависимая переменная имеет не очень большие коэффициенты попарной корреляции с каждой из набора независимых переменных.

Таблица 1. Коэффициенты попарной корреляции переменных.

	ΔE_Q	ΔE_{VDW}	ΔE_{HVDW}	MWT	MWL	CHT	CHL	ΔG_A	ΔG_P
ΔE_Q	1								
ΔE_{VDW}	0,22	1							
ΔE_{HVDW}	-0,97	-0,12	1						
MWT	0,65	0,30	-0,62	1					
MWL	-0,86	-0,45	0,83	-0,64	1				
CHT	-0,13	-0,09	0,16	-0,28	0,18	1			
CHL	0,81	0,17	-0,80	0,69	-0,75	-0,20	1		
ΔG_A	-0,25	-0,96	0,17	-0,28	0,45	0,10	-0,18	1	
ΔG_P	0,56	0,77	-0,50	0,47	-0,75	-0,11	0,38	-0,80	1
ΔG	0,54	0,57	-0,49	0,44	-0,60	-0,26	0,41	-0,56	0,63

Зависимая переменная ΔG хорошо описывается законом нормального распределения (среднее: -12,37, стандартное отклонение: 3,13, асимметрия: 0,02, эксцесс: -0,46). Плотность распределения (см. рис. 1) имеет менее выраженный пик, чем плотность нормального распределения. Однако нельзя сказать про распределение некоторых независимых переменных.

Плотность распределения независимой переменной CHL - "заряд лиганда" приведена на рисунке 2. Анализируя диаграмму, можно выделить два пула. Один пул с зарядами от -6 до -7, и второй, который может быть описан законом нормального распределения с параметрами (среднее: 0,36, стандартное отклонение: 1,78, асимметрия: 1,32, эксцесс: 5,86).

Плотность распределения независимой переменной MWL - "масса лиганда" приведена на рисунке 3. На диаграмме явно просматриваются три класса. Один пул - низкомолекулярные лиганды с массой до 1200, второй пул - лиганды с массой от 6000 до 8000 и третий пул - лиганды с массой более 10000.

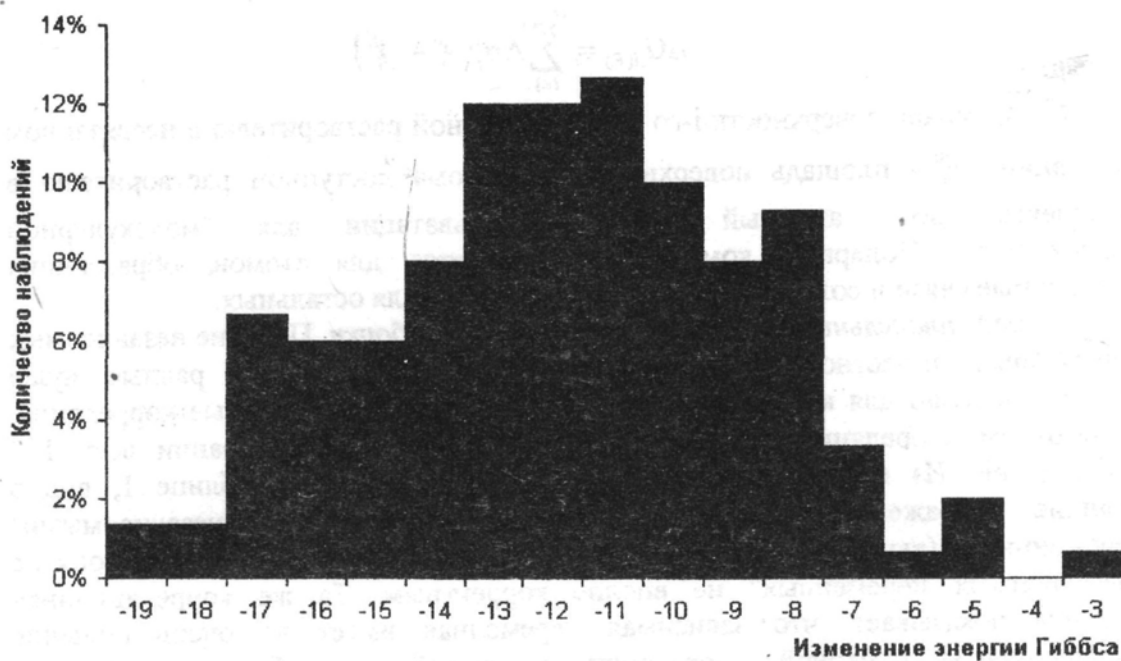


Рисунок 1.

Плотность распределения зависимой переменной (ΔG).

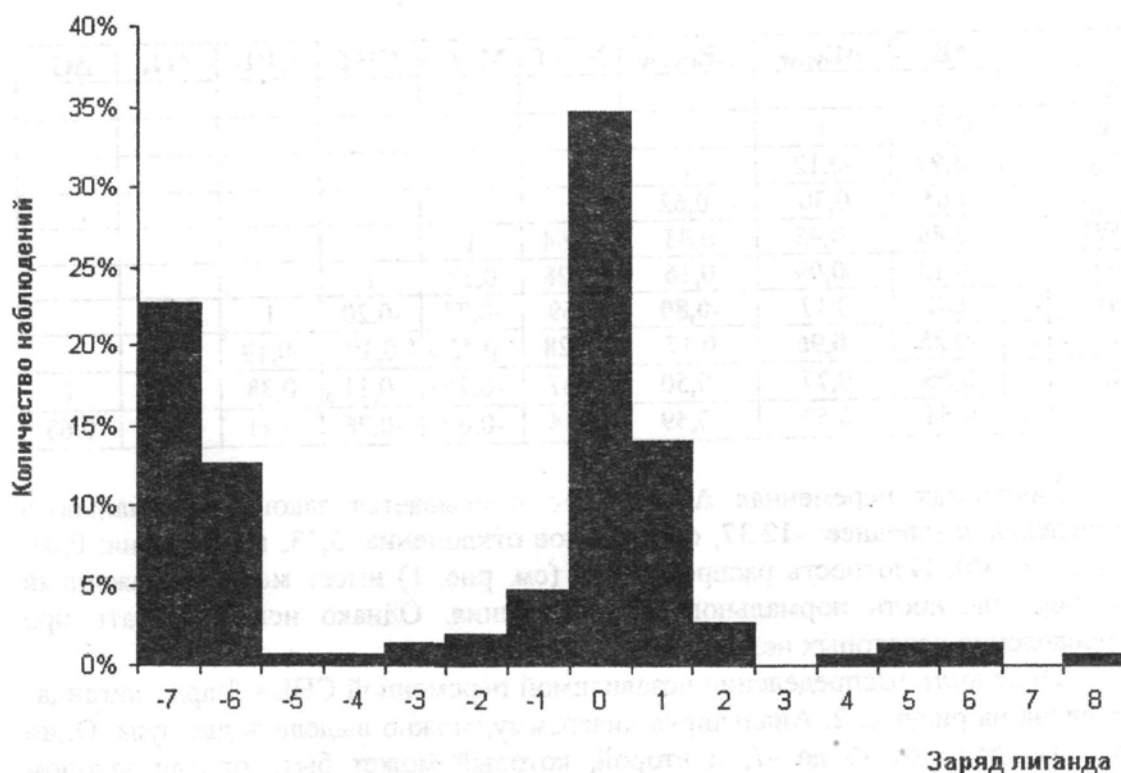


Рисунок 2.

Плотность распределения переменной "Заряд лиганда".

Поскольку плотности распределений независимых переменных существенно отличаются от нормального, имеет смысл рассматривать не только робастное оценивание параметров моделей, но и "стойкое" оценивание

параметров [5]. Такая методика оптимизации оценивания параметров произвольных (в отличие от робастности) распределений по двум признакам (эффективности и устойчивости), с использованием методов вариационного исчисления дает наилучшую возможность прогнозирования случайных процессов и точечных полей, фигурирующих в начальных условиях многих практических задач в естественных и технических науках. Разумеется, она может быть применена и к рассматриваемой проблеме.

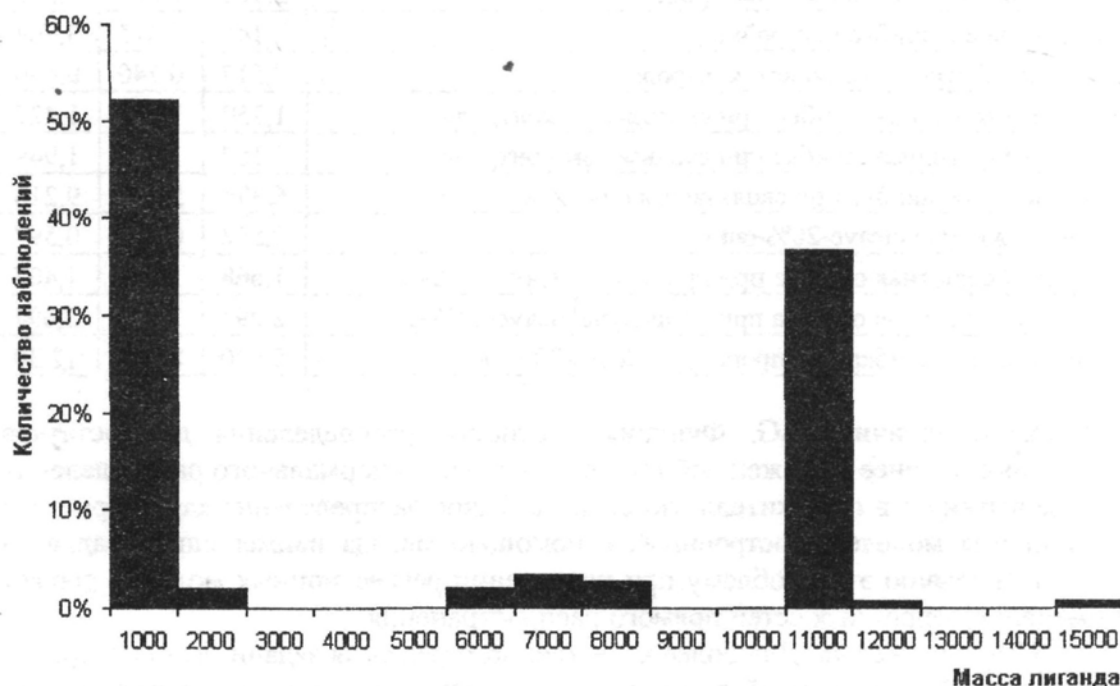


Рисунок 3.
Плотность распределения переменной "Масса лиганда".

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ. Все модели, построенные в рамках настоящего исследования, как и все описанные в настоящей работе модели, были созданы с использованием разработанного оригинального программного обеспечения [6].

Линейная модель. Линейная взаимозависимость изменения энергии Гиббса при образовании комплексов от пространственно-зарядовых показателей компонентов уже исследовалась некоторыми авторами [4,7]. Для исследуемого набора данных классическая модель множественной линейной регрессии (1N), регуляризованная в соответствии с критерием Фишера [8,9,10], приводит к уравнению:

$$\Delta G = (-6,3 \pm 0,6) + (0,0124 \pm 0,0017) \Delta E_Q + (0,073 \pm 0,009) \Delta E_{VDW} + (-0,08 \pm 0,03) \text{СНТ}$$

Диаграмма рассеяния, полученная с помощью этой модели, приведена на рисунке 4. Согласно этой модели, мы получаем $R^2 = 0,54$ и среднеквадратичную ошибку 2,12 (табл.2). Выборочный коэффициент асимметрии остатков модели оказывается равным +0,36, а коэффициент эксцесса остатков модели оказывается равным -0,12 (для нормального распределения коэффициент асимметрии и коэффициент эксцесса равны 0). Это свидетельствует о том, что модель переоценивает малые по абсолютной величине ΔG и недооценивает большие по

Таблица 2. Сравнительные характеристики моделей предсказания ΔG .

Параметры модели	1N	2N	3N
Число связей	3	20	29
R^2 настройки	0,536	0,814	0,829
Средняя абсолютная ошибка настройки	1,714	1,065	1,008
Среднеквадратичная ошибка настройки	2,123	1,343	1,290
Максимальная ошибка настройки	6,146	3,875	3,708
R^2 при процедуре скользящего контроля	0,517	0,740	0,660
Средняя абсолютная ошибка при скользящем контроле	1,750	1,260	1,427
Среднеквадратичная ошибка при скользящем контроле	2,167	1,595	1,949
Максимальная ошибка при скользящем контроле	6,334	5,590	9,216
R^2 при процедуре «leave-20%-out»	0,552	0,721	0,597
Средняя абсолютная ошибка при процедуре «leave-20%-out»	1,668	1,253	1,407
Среднеквадратичная ошибка при процедуре «leave-20%-out»	2,092	1,666	2,094
Максимальная ошибка при процедуре «leave-20%-out»	5,620	8,846	12,30

абсолютной величине ΔG . Функция плотности распределения для остатков модели имеет менее выраженный пик, чем плотность нормального распределения и более вытянута в положительную область. Такое распределение характерно для большинства моделей, построенных с помощью метода наименьших квадратов [11, 12]. Частично эту проблему при построении регрессионных моделей решает применение нейронных сетей прямого распространения.

Нейронная сеть. Для более качественного решения задачи было построено несколько нейронных сетей прямого распространения с одним скрытым слоем. Ниже будут описаны две модели, содержащие два (2N) и три (3N) нейрона в скрытом слое. 2N-модель имела в скрытом слое один нейрон с линейной функцией активации и один нейрон с функцией активации $f(x)=x^3$. В результате регуляризации модели в соответствии с критерием Фишера в сети осталось только 20 информативных связей. По этой модели мы получаем $R^2 = 0,82$ и среднеквадратичную ошибку 1,34 (табл.2). Выборочный коэффициент асимметрии остатков модели оказывается равным +0,05, а коэффициент эксцесса остатков модели оказывается равным +0,07. Эта модель также несколько переоценивает малые по абсолютной величине ΔG и недооценивает большие по абсолютной величине ΔG . Функция плотности распределения для остатков модели имеет более ярко выраженный пик, чем плотность нормального распределения и немного вытянута в положительную область. 3N-модель имела в скрытом слое один нейрон с линейной функцией активации, один нейрон с функцией активации $N(0,1)$ и один нейрон с функцией активации $f(x)=x^3$. В результате регуляризации модели в соответствии с критерием Фишера в сети осталось 29 информативных связей. По этой модели мы получаем $R^2 = 0,829$ и среднеквадратичную ошибку 1,290 (табл.2). Выборочный коэффициент асимметрии остатков модели оказывается равным +0,10, а коэффициент эксцесса остатков модели оказывается равным +0,15. Диаграммы рассеяния для обеих моделей приведены на рисунках 5 и 6 соответственно.

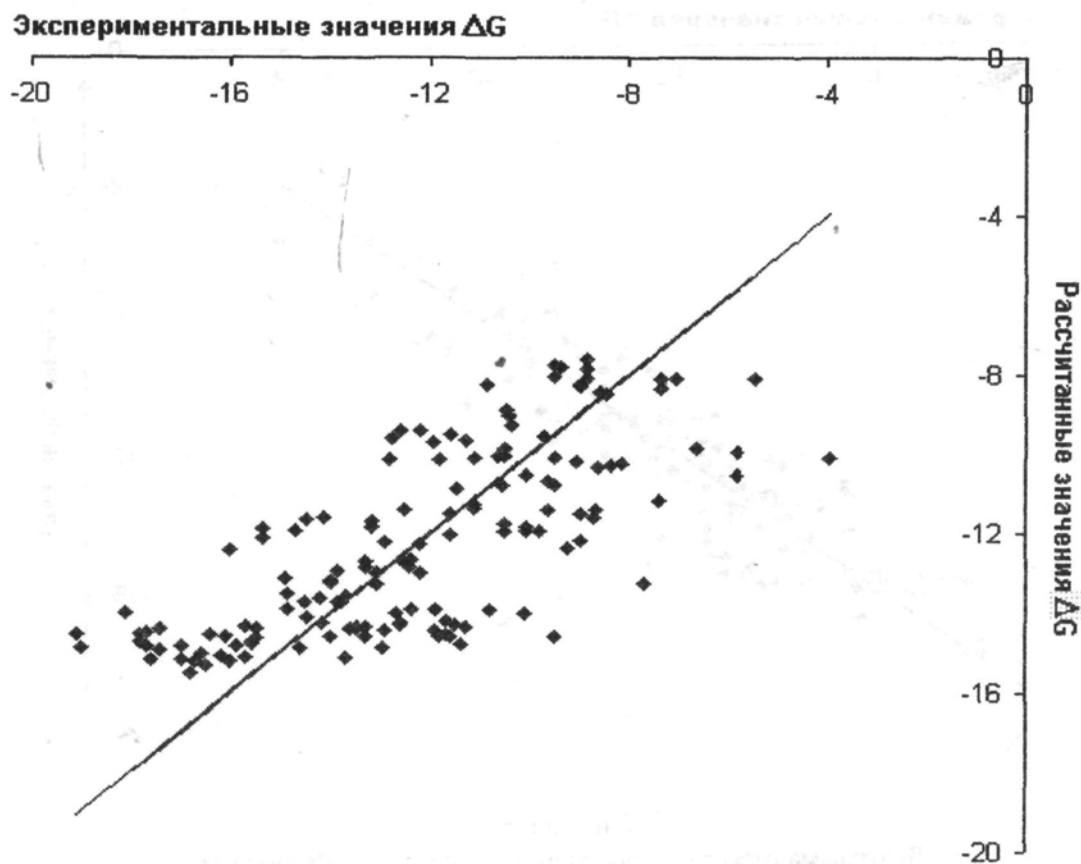


Рисунок 4.

Диаграмма рассеяния, полученная с помощью классической линейной модели.

Результаты проверки. Простейшую проверку каждой модели производили с помощью процедуры скользящего контроля (с удалением по одному). Более качественная проверка адекватности и устойчивости построенных моделей проводилась следующим образом (классическая процедура «leave-20%-out»). Вся выборка из 150 наблюдений была случайным образом разделена на 5 подвыборок по 30 наблюдений в каждой. Каждая (1N, 2N и 3N) модель строилась на основании четырех подвыборок из пяти. После минимизации (настройки) предсказывались значения для оставшейся вне рассмотрения подвыборки. Проведя такие действия 5 раз, были получены контрольные значения для всех 150 наблюдений из полной выборки. Именно это множество значений и сравнивалось с множеством известных ответов. В обоих экспериментах классическая модель множественной линейной регрессии проявила себя как самая устойчивая модель, обладающая, тем не менее, низким качеством предсказания. 2N-модель показала удовлетворительное качество предсказания при достаточной устойчивости. 3N-модель оказалась неустойчивой и, поэтому, несмотря на более хорошие показатели качества при настройке на обучающее множество, показала худшее качество при проверке. Лучше других

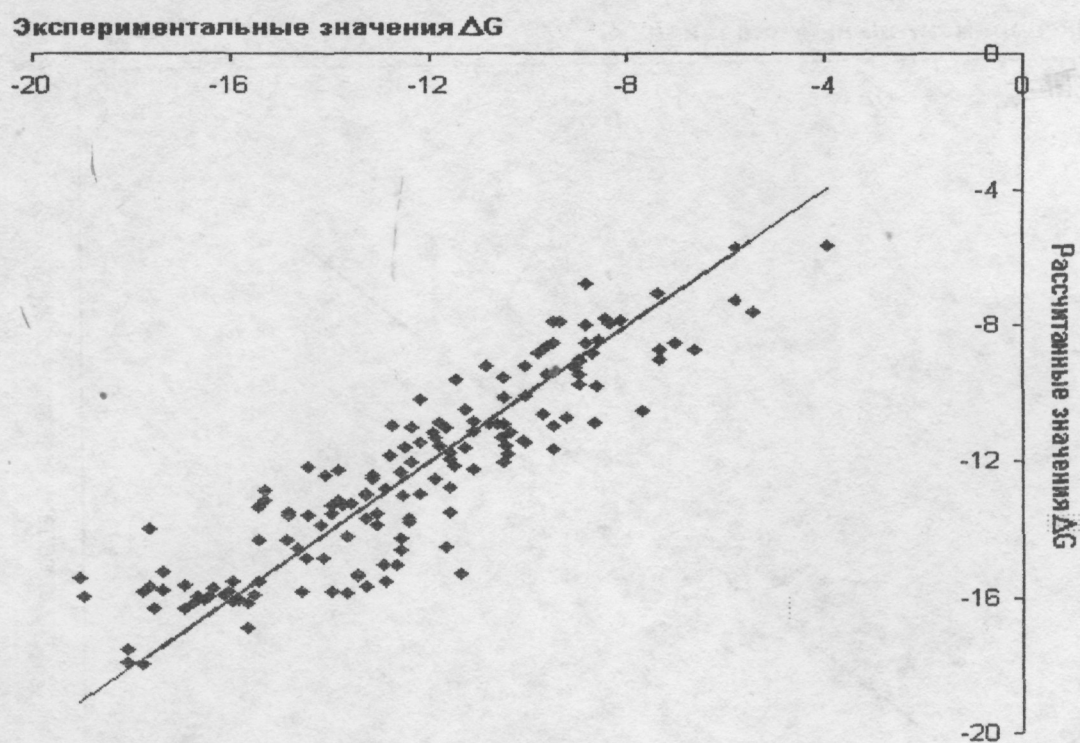


Рисунок 5.
 Диаграмма рассеяния, полученная с помощью 2N-модели.

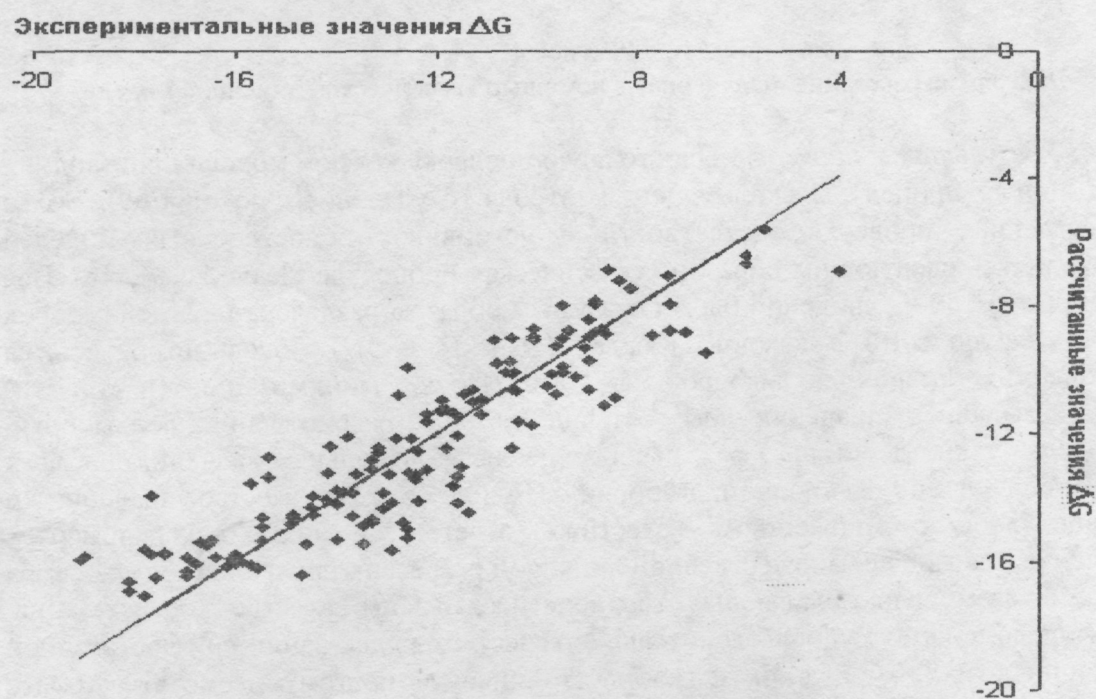


Рисунок 6.
 Диаграмма рассеяния, полученная с помощью 3N-модели.

всеми тремя моделями описываются следующие комплексы: 1htf, 1htg, 1hvi, 1hvj, 1hih, 8abp, 1eas, 1dwc, 3sic. Хуже других описываются: 1pro, 1eed, 4tpi, барназа / барстар, барназа / мутант барстара tyr29phe, барназа / мутант барстара glu76ala, мутант барназы arg83gln / барстар, мутант барназы glu60ala / барстар, мутант барназы arg59ala / мутант барстара asp39ala. В таблице 3 приведен ряд примеров предсказания ΔG для различных комплексов. Полностью таблица помещена в сети Интернет по адресу: http://imgdd.ibmh.msk.su/NNC/dg_example.html.

Таблица 3. Примеры предсказаний ΔG .

Комплекс	Эксп. ΔG	Предск. ΔG (1N)	Предск. ΔG (2N)	Предск. ΔG (3N)
мутант барназы glu73ala / мутант барстара trp38phe	-16.20	-15.12	-15.91	-16.21
8abp	-8.97	-8.24	-9.02	-8.99
1abe	-8.83	-7.84	-8.00	-8.85
5sic	-13.89	-12.97	-12.30	-13.85
3sic	-13.89	-13.72	-13.14	-13.83
2er0	-8.70	-11.61	-8.84	-8.76
2sni	-14.94	-13.17	-14.31	-14.86
1hgt	-14.48	-11.64	-12.20	-14.58
1hvj	-14.20	-13.62	-13.88	-14.09
мутант барназы glu73ser / барстар	-16.00	-15.23	-16.12	-16.14
...
барназа / мутант барстара tyr29ala	-15.60	-14.77	-15.93	-16.26
2dbl	-11.82	-10.14	-10.85	-12.49
1hsg	-12.57	-12.70	-13.03	-13.24
мутант барназы lys49ala / барстар	-15.60	-14.82	-15.92	-16.28
1hps_b	-13.18	-11.87	-12.41	-12.48
1nsd	-8.15	-10.24	-7.87	-7.46
мутант барназы glu73phe / барстар	-16.80	-15.55	-16.15	-16.10
1bmnn	-11.46	-10.86	-9.62	-10.77
4er1	-8.97	-11.50	-9.75	-9.71
2er6	-9.78	-11.94	-8.83	-10.53
...
7gch	-12.60	-9.42	-12.33	-9.94
мутант барназы glu73phe / мутант барстара asp35ala	-12.60	-14.32	-14.59	-15.31
мутант барназы his102ala / барстар	-12.90	-14.48	-15.53	-15.67
1eed	-6.66	-9.86	-8.73	-9.47
мутант барназы glu60ala / барстар	-9.50	-14.63	-12.53	-12.37
4tpi	-3.94	-10.09	-5.62	-6.96
мутант барназы arg83gln / барстар	-11.40	-14.80	-15.28	-14.47
барназа / мутант барстара tyr29phe	-19.10	-14.56	-15.45	-16.01
барназа / барстар (lay7)	-19.00	-14.88	-15.99	-15.90
барназа / мутант барстара glu76ala	-17.70	-14.54	-13.97	-13.99

В заключение следует отметить, что использование методов нелинейной регрессии может привести к существенному улучшению качества построения моделей. К "кажущимся" недостаткам этого метода можно отнести непрозрачность конечного результата, невозможность применения для небольших обучающих выборок, сложность и длительность процесса создания моделей. Интерпретация физического смысла получаемых уравнений действительно затруднена, однако это часто справедливо и для уравнений линейной регрессии. Сложность решения, сейчас, с развитием мощности вычислительной техники, тоже не является препятствием к использованию. Несомненным ограничением остается только размер выборки. Поэтому мы видим развитие данной работы в дальнейшем увеличении объема обучающего множества с привлечением все большего числа разнообразных комплексов, использованием новых расчетных параметров взаимодействия. Тем не менее, во многих случаях использование нелинейных подходов целесообразно. Также, кажется целесообразным применение последних разработок математического аппарата в этой области и использование эффективных и устойчивых оценок для параметров моделей. В случае, когда плотности распределений независимых переменных близки к нормальному распределению, достаточно использовать регуляризацию решений и робастные оценки параметров, поскольку в этом предположении оценки параметров являются состоятельными, эффективными и устойчивыми. Как правило, предположение о нормальности закона распределения не выполняется. В этом случае можно рекомендовать использование "стойких" оценок.

ЛИТЕРАТУРА

1. KeyLock, (1999), <http://imgdd.ibmh.msk.su/KeyLock>
2. *Sussman J.L., Lin D., Jiang J. et al.*, (1998), *Acta Cryst.* **D54**, 1078-1084.
3. Sybyl 6.6, (1999), Tripos Inc., 1699 South Hanley Road, St Louis, Missouri, 63144, USA.
4. *Horton N., Lewis M.*, (1992), *Protein Science* **1**, 169-181.
5. *Шурыгин А.М.*, (2000), Прикладная стохастика. Робастность, оценивание, прогноз. М., Финансы и статистика.
6. Neural Network Constructor, (2000), <http://imgdd.ibmh.msk.su/NNC>
7. *Oprea T.I., Marshall G.R.*, (1998), *Perspectives in Drug Discovery and Design*, Kubinyi H., Folkers G., Martin Y.C. (Eds), **9/10/11, Part I**, 35-61.
8. *Афифи А.А., Эйзен С.П.*, (1982), Статистический анализ. Подход с использованием ЭВМ. - М., Мир.
9. *Krepets V.V.*, (1999), The Second International Conference on Discovery Science, Tokyo, Japan.
10. *Шакин В.В., Крепец В.В.*, (1999), Тезисы докладов IX Всероссийской конференции по математическим методам распознавания образов (ММРО-9), ВЦ РАН, Москва.

11. *Manallack D.T., Livingstone D.J.*, (1992), *Medicinal Chem Res* **2**, 181-190.
12. *Manallack D.T., Livingstone D.J.*, (1993), *J. Medicinal Chem* **36**, 1295-1297.

Поступила 7.06.2000.

**PREDICTION OF BINDING AFFINITIES FOR
PROTEIN-LIGAND COMPLEXES BY USING
NON-LINEAR MODELS.**

V.V. KREPETS, N. V. BELKINA,
V. S. SKVORTSOV, A. S. IVANOV

Orekhovich Institute of Biomedical Chemistry RAMS.
10, Pogodinskaya St., 119832, Moscow, Russia. (krepets@ibmh.msk.su)

A network model for prediction of the free energy changes in protein-ligand complexes has been developed. The 150 complexes of different nature were used as a training set. The computational physics-chemical parameters of these complexes were used as independent variables. Both classical models of multiple linear regression and several network models with one hidden layer were created and the best was chosen. Significant improvement was shown for network model prediction quality in comparison with classical model of multiple linear regression (R^2 on training - 0,81 and 0,54; R^2 on «leave-one-out» procedure - 0,74 and 0,52 respectively).

Key words: protein-ligand complex, dissociation constant, neural network.