

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Институт химии и инженерной экологии

(наименование института полностью)

Кафедра «Химия, химические процессы и технологии»

(наименование кафедры)

04.03.01 «Химия»

(код и наименование направления подготовки, специальности)

«Медицинская и фармацевтическая химия»

(наименование (профиль)/специализации)

БАКАЛАВРСКАЯ РАБОТА

на тему: QSAR-модели для изыскания новых ингибиторов рецептора
эпидермального фактора роста

Студент

Ю.С. Чуркина

(И.О. Фамилия)

(личная подпись)

Руководитель

Е.В.Варакина

(И.О. Фамилия)

(личная подпись)

Консультанты

Е.Ю.Аношина

(И.О. Фамилия)

(личная подпись)

(И.О. Фамилия)

(личная подпись)

Допустить к защите

Заведующий кафедрой д.х.н., профессор Г.И. Остапенко

(ученая степень, звание, И.О. Фамилия)

(личная подпись)

« _____ » _____ 2018 г.

Тольятти 2018

АННОТАЦИЯ

Название выпускной работы: «Построение модели QSAR для поиска ингибиторов рецептора эпидермального фактора роста». Эта работа посвящена молекулярному моделированию ингибиторов EGFR.

Целью этой выпускной работы является разработка валидных QSAR моделей для изыскания новых структурных ингибиторов киназ семейства эпидермального фактора роста.

Для достижения цели мы провели построение классической 2D QSAR модели. Затем построили 3D QSAR модели методом молекулярного поля. Провели анализ применимости данных моделей для предсказания новых

Результаты исследования показали, определенные количественные характеристики представленной модели, с помощью набора результатов для валидации, при тренировочной выборке в 75% достигаются наилучшим значениям корреляции перекрёстного контроля Q^2 достигающие значения выше 0.75. Модели, полученные в этой работе, могут быть использованы для оптимизации комбинаторной библиотеки на основе 5-цианопиримидина, чтобы найти более активные лидирующие соединения.

ABSTRACT

The title of the graduation work is “QSAR models for the search for inhibitors of the epidermal growth factor receptor”. This work is dedicated to the study of the molecular modeling of EGFR molecule inhibitors.

The aim of this graduation work is to develop valid QSAR models for the discovery of new structural kinase inhibitors of the epidermal growth factor family.

To achieve the aim, we carried out the construction of a classical 2D QSAR model. Then, 3D QSAR models were constructed using the molecular-field method. An analysis of the applicability of these models to predict new inhibitors was carried out.

The graduation work consists of an introduction, 4 chapters, a conclusion, list of 365 references, all of which are foreign sources. The text of the work contains 13 figures, 16 tables and 51 formulas.

The first part of the work includes a literature overview on a current topic. It also studies methods of molecular modeling.

The second part represents the practical implementation of this topic by providing experiments and outlines results the work.

The results of the research showed that certain quantitative characteristics of the presented model with a set of results for validation in the training sample of 75% are achieved with better cross-correlation values Q^2 , reaching values above 0.75. The models obtained in this work can be used to optimize a combinatorial library based on 5-cyanopyrimidine to find more active lead compound.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	6
ЛИТЕРАТУРНЫЙ ОБЗОР	7
1.1. История развития QSAR.....	7
1.1.1. Взаимодействие «лекарство-рецептор».....	12
1.2. Deskрипторы	22
1.2.1. Deskрипторы липофильности	30
1.2.2. Deskрипторы поляризуемости.....	37
1.2.3. Электронные deskрипторы.....	41
1.2.4. Стерические deskрипторы.....	48
1.3. Количественные модели.....	51
1.3.1. Экстратермодинамический подход (анализ Ганча).....	51
1.3.2. Модель аддитивности (анализ FreeWilson).....	59
1.3.3. Другие подходы QSAR.....	63
2. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ.....	72
3. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ	93
3.1. Подготовка комбинаторной библиотеки	93
3.2. QSAR.....	93
3.3. 3D-QSAR	93
ЗАКЛЮЧЕНИЕ	95
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ	96

ПРИНЯТЫЕ СОКРАЩЕНИЯ

В настоящей работе применяют следующие термины с соответствующими определениями:

QSAR	Количественные отношения структуры и активности
CoMFA	Сравнительный молекулярный анализ поля
CoMSIA	Анализ показателей сравнительного молекулярного сходства
PCA	Анализ основных компонентов
ТМП	Триметоприм
PGDP	Акцептор водородной связи
ЯМР	Ядерный магнитный резонанс
ИК	Инфракрасная спектроскопия
ВЭЖХ	Высокоэффективная жидкостная хроматография
DDT	1, 1, 1 трихлор-2,2-бис(4-хлорфенил)
cAMP	Специфическая фосфодиэстераза
ДНК	Дезоксирибонуклеиновая кислота
БПК	Биологическая потребность в кислороде
IC ₅₀	Концентрация 50% ингибирования
MSD	Минимальная стерическая разность
MTD	Минимальная топологическая разность
РА	Парашор
MV	Молярный объем
MW	Молекулярная масса
DHFR	Дигидрофолатредуктаза
MSA	Анализ молекулярной формы
ALS	Адаптивный метод наименьших квадратов
PLS	Метод частичных наименьших квадратов
FA	Факторный анализ
CAS	Анализ значимости кластеров
DEF	Файл определения модуля

ВВЕДЕНИЕ

Одна из важнейших задач современного молекулярного моделирования является прогнозирование свойств молекул вычислительными методами с дальнейшей визуализацией результатов. Среди методов исследования новых ингибиторов рецептора эпидермального фактора роста выделяют молекулярное моделирование, которое за последние годы значительно усовершенствовалось. В его основе используют количественные соотношения структуры и активности, получают модели (QSAR), которые описывают структурную зависимость биологической активности либо по физико-химическим параметрам (анализ Ганча), по переменным индикаторам, координирующим различные структурные особенности (свободный анализ Уилсона), либо по трехмерным характеристикам молекулярных свойств соединения (сравнительный анализ поля, CoMFA).

Представленная работа посвящена разработке валидных QSAR моделей для изыскания новых структурных ингибиторов киназ семейства эпидермального фактора роста.

Для реализации поставленной цели были решены следующие задачи:

1. Построение классической 2D QSAR модели.
2. Построение 3D QSAR модели методом молекулярного поля.
3. Анализ применимости данных моделей для предсказания новых ингибиторов

ЛИТЕРАТУРНЫЙ ОБЗОР

1.1. История развития QSAR

В 1868 году Крам-Браун и Фрейзер [1] опубликовали уравнение, которое считается первой общей формулировкой количественного соотношения структура-активность. В своих исследованиях по различным алкалоидам они признали, что алкилирование основных атомов азота приводит к значительному разному биологическому воздействию полученных постоянно заряженных соединений четвертичного аммония по сравнению с основными аминами. Поэтому они предположили, что «физиологическая активность» Φ должна быть функцией химической структуры C (уравнение 1).

$$\Phi=f(C) \quad (1)$$

Рише [2] обнаружил, что токсичность органических соединений обратно пропорциональна их водорастворимости. Такое соотношение соответствует уравнению 2, где $\Delta\Phi$ - различия в значениях биологической активности, вызванные соответствующими изменениями в химических и особенно физико-химических свойствах ΔC .

$$\Delta\Phi = f(\Delta C)(2)$$

Строго говоря, до сих пор нет способа применить уравнение 1 к биологическим данным. Все уравнения QSAR соответствуют уравнению 2, поскольку только различия в биологической активности количественно коррелируют с изменениями липофильности и другими физико-химическими свойствами исследуемых соединений.

На рубеже прошлого века Мейер [3] и Овертон [4] независимо друг от друга наблюдали линейные зависимости между липофильностью, выраженной как коэффициент распределения нефти и воды, и наркотическим действием [5]. Фойнер [6] понял, что в рамках гомологичной серии наркотические действия увеличиваются в геометрической прогрессии, то есть 1: 3: 32: 33. Что дало первые доказательства аддитивности групповых

вкладов в значения биологической активности. Этот результат был подтвержден многими другими исследованиями, в которых использовались различные параметры липофильности для описания различных видов неспецифической биологической активности. Фергюсон дал термодинамическую интерпретацию таких неспецифических соотношений структура-активность, что также объясняло часто наблюдаемое «отсечение» значений биологической активности за пределами определенного диапазона липофильности [7].

Методология QSAR быстро развивалась с середины пятидесятых годов: Бруис, Хараш и Винзер [8] сформулировали групповые вклады в значения биологической активности в серии аналогов гормона щитовидной железы, которые можно рассматривать как первый анализ типа FreeWilson. Захрадник [9 – 11] попытался применить понятие уравнения Гамметта (уравнение 3) [12], которое в то время использовалось в течение трех десятилетий для количественного описания реакционной способности органических соединений, а также биологических данных (уравнение 4).

$$\log k_{R-X} - \log k_{R-H} = \rho \sigma (3)$$

$$\log \tau_i - \log \tau_{Et} = \alpha \beta (4)$$

τ_i , в этом «биологическом уравнении Гамметта» означает значение активности i -го члена ряда, τ_{Et} - значение биологической активности этилового соединения той же серии, β - постоянная заместителя (соответствующая электронному параметру σ в уравнении Гамметта), а α - константа, характеризующая биологическую систему, которая соответствует константе реакции Гамметта ρ . К сожалению, уравнение 4 справедливо только для неспецифических биологических активностей, чаще всего в рамках гомологичных рядов и в пределах определенного диапазона липофильности.

В 1962 году Хансен [13] получил первое реальное отношение типа Гамметта между токсичностью замещенных бензойных кислот и электронными постоянными их заместителей. В том же году появилась

первая публикация QSAR Корвина Ганча «Соотношение биологической активности феноксиуксусных кислот с константами заместителя Гаммета и коэффициенты распределения»[14].

1964 год можно считать годом рождения современной методологии QSAR. Время было готово к более общим формулировкам, как относиться к соотношениям структуры и активности количественным образом. Независимо, были опубликованы две статьи, одна из которых была опубликована Ганчем и Фуджитой по « ρ - σ - π Анализ. Метод корреляции биологической активности и химической структуры» [15], вторая - Фри и Уилсоном «Математический вклад в исследования структурной активности»[16]. Оба вклада начали разработку двух новых методов количественных отношений структуры и активности, которые позднее назывались анализом Ганча (линейный подход, связанный с свободной энергией, экстратермодинамический подход) и анализ FreeWilsonсоответственно. Настоящий прорыв в QSAR был обусловлен сочетанием различных физико-химических параметров в линейной аддитивной форме (уравнение 5, $\log 1/C$ - логарифм обратной молярной дозы, которая производит или предотвращает определенный биологический ответ, $\log P$ является логарифмом коэффициента распределения Рн-октанола/воды), как это было сделано ранее в теоретической органической химии. Дальнейшими вкладами было определение рассчитанного параметра липофильности π (уравнение 6), которое должно использоваться вместо измеримых значений $\log P$ (например, значения Гаммета σ используются вместо равновесных констант органических реакций) и формулировка параболического уравнения для количественного описания зависимостей нелинейной липофильности и активности (уравнение 7) [17 – 19].

$$\log 1/C = a \log P + b \sigma + \dots + \text{const.} \quad (5)$$

$$\pi_X = \log P_{R-X} - \log P_{R-H} \quad (6)$$

$$\log 1/C = a (\log P)^2 + b \log P + c \sigma + \dots + \text{const.} \quad (7)$$

Учитывая значительный вклад Фуджиты и Пана [20], модель FreeWilson определяется формулой 8, где a_{ij} - групповой вклад заместителя X_i в положении j , а μ - (теоретическая) величина биологической активности эталонного соединения в серии; все групповые вклады a_{ij} различных заместителей X_i относятся к соответствующим заместителям (чаще всего к водороду) этого эталонного соединения.

$$\log 1/c = \sum a_{ij} + \mu \quad (8)$$

За последние три десятилетия обе модели остались неизменными. Некоторые улучшения были результатом сочетания уравнений Ганча с индикаторными переменными [21], которые можно рассматривать как смешанную модель Ганча/Free Wilson [22] и из формулировки теоретически полученных нелинейных моделей для транспорта и распределение лекарств в биологической системе, например, билинейная модель [23].

$$\log 1/C = a \log P - b \log (\beta P + 1) + c \quad (9)$$

Были предприняты различные попытки использовать распознавание образов [24, 25] в исследованиях QSAR и успешных заявках. Способы мягкого моделирования, например, метод частичных наименьших квадратов (PLS) [26, 27] теперь предлагает лучшие возможности. С помощью этого основного компонентноподобного метода объясняющая способность многих, даже сотен или тысяч переменных может использоваться для ограниченного числа объектов, причем задача абсолютно невозможна в регрессионном анализе, в котором количество объектов всегда должно быть больше чем число переменных.

Трехмерные количественные соотношения структура-активность (3DQSAR) были разработаны из первых попыток нанести на карту рецепторную поверхность путем анализа уравнения QSAR для нековалентных взаимодействий лигандов в разных положениях замещения (например, [28]). Холтье [29, 30] распространил этот подход. Он постулировал определенные аминокислотные боковые цепи в качестве партнеров-партнеров, вычислял энергии взаимодействия в стандартных

геометриях и коррелировал эти энергии с значениями биологической активности. Было предпринято несколько других попыток сопоставить гипотетические сайты взаимодействия рецептора, например, метод геометрии расстояния Криппена [31, 32]. GRID рассчитывает энергии взаимодействия некоторых зондовых атомов с поверхностью белка, трехмерная структура которого известна из кристаллографического анализа [33].

Если трехмерная структура белка неизвестна, различные поля лигандов можно сравнить в трехмерном пространстве. Молекулы химически родственных рядов накладываются друг на друга, следуя некоторым гипотезам выравнивания (гипотезы фармакофора), сетка укладывается над молекулами, а значения стерического и электростатического полей (и необязательно других полей) вычисляются в каждой точке сетки для каждой серии молекул. Соответствующий многомерный статистический метод сопоставляет тысячи таких значений энергии в разных точках сетки (каждый из которых представляет собой столбец в блоке X) с биологической активностью. В первой версии, называемой DYLOMMS, был использован анализ главных компонент [34]; более поздний анализ PLS оказался более подходящим [35]. Сравнительный анализ молекулярного поля (CoMFA), как он используется в настоящее время, был сформулирован в 1988 году [36, 37]. Метод, который все еще находится в активной разработке, за короткое время нашел много успешных приложений [38].

Отличная недавно опубликованная монография, которая охватывает всю область QSAR, представляет собой книгу «Количественный дизайн лекарств», том IV шеститомной комплексной медицинской химии [39]. Кроме того, опубликованы многочисленные другие монографии [40-44], непосредственно связанные с методологией и приложениями QSAR [45-49], по физико-химическим параметрам [50-54] или по похожим темам [55-59].

Поскольку невозможно привести всю соответствующую работу, и поскольку выбор оригинальных вкладов всегда более или менее

неоднозначен, читатель ссылается на [39, 60-64], так же к нескольким сериям монографий [65 - 69], к материалам симпозиумов, связанных с QSAR [70-74], к журналу «Количественные отношения структуры и активности», особенно к разделу рефератов этого журнала [75-80], который из года в год содержит около 400 ~ 500 превосходно подготовленных тезисов публикаций [81-85] связанных с QSAR, к другим услугам рефератов [86 - 88], а также к некоторым другим журналам [89], включая публикации QSAR как их регулярное содержание.

История QSAR была рассмотрена в книгах (например, [40]) и в специальных статьях [5, 90-93], разработка методов 3DQSAR комментируется в [36, 38].

1.1.1. Взаимодействие «лекарство-рецептор»

«Corporanonaguntnisifixata» (Ehrlich, 1913) [94] была ранней формулировкой того факта, что лекарственные средства должны взаимодействовать с определенными биологическими макромолекулами для осуществления их биологической активности.

Концепция взаимодействия лекарств с определенными «веществами, с которыми они способны образовывать соединения, в соответствии с их химической близостью к Y», восходит к работе Лэнгли в 1873 - 1878 гг. [95]. Стереоспецифичность таких взаимодействий была признана Фишером в 1894 году. В своих исследованиях по ферментативному расщеплению аномерных гликозидов инвертином и эмульсином (α -глюкозидаза и β -глюкозидаза соответственно) он сформулировал: «*umeinBildzugebrauchen, dassEnzyrnundGlukosidwieSchlossundSchliisselzueinanderpassenmiissen, umeinechemischeWirkungaufeinanderausiibenzukonnen*» (чтобы проиллюстрировать, я хотел бы сказать, что фермент и гликозид должны соответствовать друг другу, как замок и ключ, для того, чтобы оказывать химическое воздействие друг на друга) [96]. Термин рецептор был впервые использован Эрлихом в его исследованиях на красителях и их

взаимодействии с биологическими тканями. В следующем «рецептор» иногда используется как синоним в любой биологической мишени, например: любой специфический сайт связывания макромолекулы; строго говоря, этот широкий смысл неверен из нашего сегодняшнего определения рецепторов как растворимые, закрепленные на мембране или внедренные в мембрану белки, которые способны вызывать определенный биологический отклик в серии наиболее неизвестных событий [59, 97, 98].

Следует отметить, что работа Эрлиха также содержит видный пример случайного успеха, основанного на «рациональном» дизайне лекарств, который позже оказался основан на неправильной гипотезе. Prontosil rubrum, р-[(2,4-диаминофенил)-азо] бензолсульфонамид, предназначен для окрашивания и уничтожения инфекционных микроорганизмов. Однако метаболит сульфаниламид является активным агентом, а не самим красителем [97]. Один из самых известных интуитивных примеров (термин, придуманный Горацией Уолполем в 1754 году от Серендипа, прежнее название Цейлона, в старой персидской сказке под названием «Три князя Серендипа», в которой князья совершают случайно или неожиданно интересные открытия) было обнаружение Флеминга антибактериальной активности некоторых грибов. Счастлиное обстоятельство, что он не убрал свои блюда сразу же после неудачного эксперимента, привел к грибковой инфекции культуры бактериальных клеток. Это наблюдение непосредственно привело к открытию пенициллина. Менее известно, что тот же самый Флеминг, который однажды простудился, просто для удовольствия «проверил» свою носовую слизь для антибактериальной активности. Этот незапланированный эксперимент привел к открытию фермента лизоцима, который также встречается в яичном белке, молоке, сыворотке крови, слезах, слюне, некоторых других выделениях и тканях животных и в растительных латексах [99].

Многие лекарства были обнаружены везеньем, интуиция всегда играла важную роль в исследовании наркотиков [100] и, несмотря на все усилия

это будет продолжаться и в будущем. Последствия краткосрочного планирования для вероятности успеха и последствия слишком бюрократического управления исследованиями в области наркотиков были критически прокомментированы [100]. С другой стороны, для проектирования и разработки нового препарата необходимы совместные усилия большой группы специалистов, которые могут работать только в той или иной форме. Кроме того, разработка лекарств является дорогостоящим и трудоемким многоступенчатым процессом, который необходимо планировать и контролировать надлежащим образом.

В течение последних десятилетий первоначально статическая блокировка и ключевая модель взаимодействия лиганд-рецептор были модифицированы до более реалистичной картины с гибкими молекулами лекарств и динамическими рецепторами [101, 102]. Всякий раз, когда лиганд приближается к месту связывания, оба партнера могут изменять свою форму (индуцированная подгонка, гибкая подгонка). Трехмерные структуры только нескольких мембранно-связанных белков и рецепторных белковых комплексов были разрешены при атомном разрешении; среди них - фотосинтетический реакционный центр [103], световодный протонный насос бактериородопсин [104] и бактериальный мембранно-канальный порин [105]. При атомном разрешении нет трехмерных структур рецепторов млекопитающих. Большинство наших знаний о геометрии взаимодействий с лигандсвязывающим сайтом объяснялось трехмерными структурами растворимых белков, особенно ферментов и их ингибиторных комплексов [106 – 110]. Некоторые общие задачи против трехмерных структур белка из кристаллографического анализа могут быть легко рассеяны. Контакты между отдельными молекулами белка в кристалле относительно слабы (что затрудняет кристаллизацию белков). Таким образом, маловероятно, что они будут влиять на нативную конформацию белка, за исключением, возможно, некоторых внешних петель. В принципе, кристаллы белка представляют собой упорядоченные водные растворы, поскольку они могут содержать до

70% воды. Они по-прежнему демонстрируют некоторую гибкость отдельных боковых цепей аминокислот и даже более крупных доменов. Некоторые белки (например, гемоглобин, а также многие ферменты) сохраняют свои функциональные свойства в кристалле, хотя константы скорости могут сильно отличаться от констант в водном растворе из-за менее благоприятных условий диффузии в кристалле; кофакторы и ингибиторы могут быть кристаллизованы или пропитаны кристаллами белка.

Были предприняты попытки моделировать рецепторы, связанные с белком G [111-113], из-за их сходства в количестве транс-мембранных доменов с бактериородопсином. Такие модели оказываются полезными для дальнейшего понимания структуры и функции рецепторов. Однако их ценность для дизайна лигандов ограничена, при атомном разрешении такие модели далеки от реальности.

Важный вклад в концепцию рецептора явилось недавнее исследование Гербета [114, 115] о разделении и распределении лекарств в биологических мембранах. Считается, что правильное пространственное расположение лекарственного средства и его правильная ориентация в мембране относительно места связывания на поверхности, внедренного в мембрану рецептора имеют первостепенное значение для взаимодействия лекарство-рецептор (рисунок 1). Кроме того, модель препарата, переносимого из водной фазы в мембрану, находящегося внутри липидного двойного слоя, уменьшает проблему лиганда, приближающегося к его связывающему участку, от трех измерений до двух измерений.

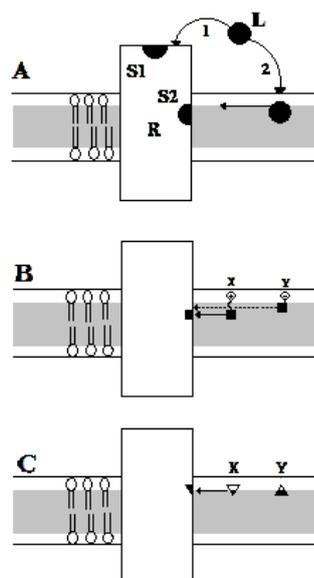


Рисунок 1- Взаимодействие лекарственных препаратов

А) Лиганд L может достигать своего сайта связывания S1 или S2 в рецепторе R путем прямой диффузии в водной среде или (в случае сайта S2) путем разделения на мембрану, а затем диффузии на сайт связывания.

В) Высокоупорядоченная структура липидного бислоя может ограничивать липофильные и особенно амфифильные препараты определенной глубиной проникновения; лекарственное средство X будет соответствовать месту связывания, поскольку оно расположено на надлежащей глубине для оптимального взаимодействия с сайтом связывания, тогда как лекарственное средство Y будет менее активным или неактивным.

С) Ориентация лиганда относительно сайта связывания также может быть оптимизирована мембраной путем ограничения вращательных степеней свободы лекарственного средства; лекарственное средство X будет активно, а лекарственное средство Y не активно. Кроме того, мембрана может стабилизировать конформации лекарственного средства, которые отличаются от тех, которые присутствуют в жидкой фазе, что позволяет или отключает взаимодействие с сайтом рецептора в мембране [114, 115].

Сродство лекарственного средства D к его сайту связывания в рецепторе R определяется разницей свободной энергии ΔG между свободными состояниями обоих партнеров и рецептор-рецепторным комплексом [DR], который составлен из изменения ΔH энтальпии и

изменение энтропии ΔS (уравнение 10). Свободная энергия ΔG связана с константой равновесия K для реакции $D + R = [DR]$ (уравнение 11)

$$\Delta G = \Delta H - T\Delta S \quad (10)$$

$$\Delta G = -2.303 RT \log K \quad (11)$$

Ниже приводится краткий обзор межмолекулярных взаимодействий между лекарственными средствами и их сайтами связывания. Ковалентные связи имеют значения энергии в диапазоне около 170-600 $\text{kJ} \times \text{mol}^{-1}$. Поскольку они необратимы, они не важны для большинства терапевтически значимых лекарств. Только алкилирующие агенты (например, противоопухолевые препараты, такие как циклофосфамид), а также активные ингибиторы фермента, основанные на активном сайте и на основе механизмов (например, пенициллины и цефалоспорины в качестве ингибиторов синтеза бактериальных клеточных стенок, хлорметилкетоны в качестве ингибиторов серина и цистеина протеазы, или паргилин, депренил и родственные аналоги в качестве ингибиторов моноаминоксидазы) образуют ковалентные связи.

Электростатические взаимодействия считаются важными силами притяжения из-за их относительной силы [59, 116-119]. Молекулярное электростатическое поле, которое окружает сайт связывания, направляет правильную ориентацию препарата [59] и отвечает за первый контакт. Однако роль электростатических взаимодействий, в основном ответственных за высокое сродство, подвергалась сомнению из-за часто неблагоприятного баланса энергии сольватации-десольватации. Трудно выразить свой вклад количественно, по ряду причин:

- Диэлектрическая постоянная в кармане связывания может существенно отличаться от ее значения в воде.
- Сила некоторых взаимодействий, например, водородных связей, зависит от геометрии взаимодействия [59, 117, 120].
- В случае дополнительных дисперсионных взаимодействий полученные значения энергии сильно зависят от небольших

различий в расстояниях между атомами, участвующими во взаимодействии.

Даже незначительные изменения в режиме связывания, точнее в геометрии рецепторного комплекса лекарства, при переходе от одного аналога к другому могут значительно увеличивать или уменьшать энтальпию связывания ΔH , что слишком часто игнорируется в количественной структурной активности анализа.

Большинство взаимодействий между заряженными группами включают водородные связи, например, между положительной группой аммония и отрицательной карбоксилатной, фенолятной, фосфатной, фосфонатной или сульфатной группой. Различные значения энергии приведены в литературе (энергии, рассчитанные в вакууме, не следует сравнивать с этими значениями, потому что они не учитывают другие межмолекулярные взаимодействия, например, с окружающими и конкурирующими молекулами воды). Оценки заряженных водородных связей были оценены, главным образом, из исследования мутеинов, в диапазоне $15-19 \text{ kJ} \times \text{mol}^{-1}$, а нейтральные водородные связи оценивались в $2-6 \text{ kJ} \times \text{mol}^{-1}$ [121-123]. Соответственно, введение нейтральной водородной связи увеличивает аффинность связывания примерно в 2-20 раз, а введение заряженной водородной связи увеличивает ее в 400-2000 раз. Различия в значениях свободной энергии, полученные из скоростей реакций лигандов, содержащих гидроксильную группу и лиганды, имеющие вместо этого атом водорода, были скомпилированы для разных ферментов [124]. Из недавнего сравнения энергий связи амидных водородных связей в водном растворе и неполярных растворителях было сделано заключение, что более ранние значения энергий нейтральной водородной связи могут быть слишком малы [125]. Дисперсионные силы являются притягательными силами между атомами на близких расстояниях. Даже молекулы, не имеющие постоянного дипольного момента, из-за движения их электронов локальные дипольные моменты, которые вызывают диполи в

противоположной молекуле, приводят к колебаниям электростатических аттракционов. На более близком расстоянии отталкивающие силы развиваются из-за неблагоприятного перекрытия сфер Ван-дер-Ваальса обеих молекул. Эти отношения обычно описываются потенциалом Леннарда-Джонса, с привлекательным членом r^{-6} и r^{-12} отталкивающим членом (рисунком 2) [59, 116]. Диполь-дипольные взаимодействия и дисперсионные силы значительно слабее других электростатических взаимодействий. Тем не менее, если есть тесный контакт между обеими молекулами на относительно большой площади поверхности, они могут суммироваться до больших значений общих энергий взаимодействия.

Гидрофобные взаимодействия являются наиболее важным единственным фактором, обеспечивающим движущую силу нековалентных взаимодействий в водном растворе, особенно в случае больших гидрофобных областей. Это всего лишь энтропийные взаимодействия. Молекулы, окружающие гидрофобные поверхности, слабо связаны, они имеют определенную степень порядка и поэтому находятся в неблагоприятном энтропийном состоянии. Связь гидрофобных областей лиганда и его связывающего участка вытесняет и освобождает упорядоченные молекулы воды в раствор, что приводит к усилению энтропии. Соответствующий вклад метиленовой группы (которая не находится в окрестности полярной группы, то есть экранирована связанной молекулой воды), оценивается примерно в 2 кДж.моль^{-1} , вклад фенильного кольца составляет около 8 кДж.моль^{-1} [116]. Тем не менее, по-прежнему наблюдается значительное обсуждение фактической силы гидрофобных взаимодействий [118, 126].

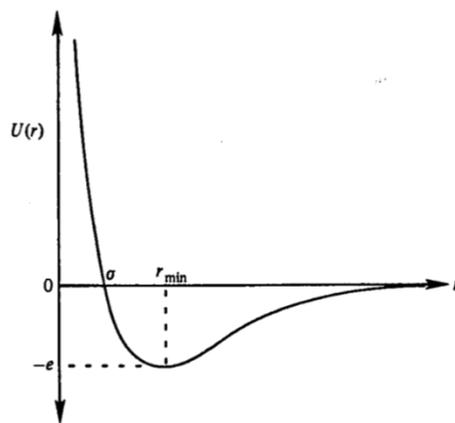


Рисунок 2 - Зависимость потенциальной энергии U двух атомов от их расстояния r (потенциал Леннарда Джонса)

Исходя из бесконечного расстояния r , энергия уменьшается (притяжение из-за электростатических взаимодействий) до достижения минимального расстояния r_{\min} ; отсюда происходит отталкивание из-за увеличения Ван-дер-Ваальсового перекрывания атомов; σ - расстояние, на которое энергия взаимодействия равна нулю [59].

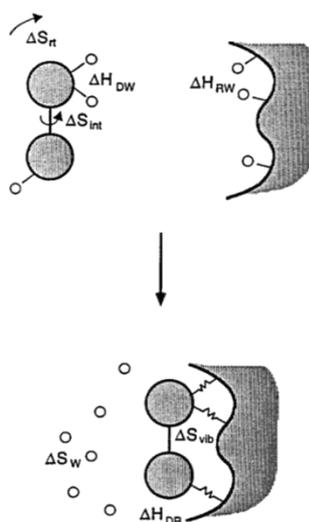


Рисунок 3 - Взаимодействие лекарственных препаратов ΔH_{DW} и ΔH_{RW} представляют собой энтальпии гидратации лекарственного средства и рецептора, соответственно, ΔH_{DR} представляет собой энтальпийный вклад взаимодействия лекарство-рецептор, ΔS_{rot} , представляет собой общую вращательную и поступательную энтропию в растворе, а ΔS_{int} - внутреннюю вращательную энтропию свободного белка ΔS_w - это увеличение энтропии за счет высвобождения молекул воды (небольшие круги) и ΔS_{vib} энтропийное усиление из-за низкочастотных колебательных мод, связанных с нековалентным взаимодействием рецептора лекарственного средства [128].

Отрицательные вклады в связывание рецептора лекарственного средства проистекают из-за потери трансляционных и вращательных энергий лиганда при переходе от свободного состояния в связанное состояние, потери внутренних вращательных степеней свободы (конформационной энтропии) в случае гибких молекул и энтальпии, которая необходима для удаления молекул воды, связанных с полярными группами обоих партнеров, то есть от десольватации. После образования комплекса лекарство-рецептор положительный вклад объясняется увеличением энтропии из-за низкочастотной вибрации, связанной с нековалентными связями лекарства рецептора.

Чистый баланс благоприятных (энтальпийных и энтропийных) и неблагоприятных (энтропийных) вкладов показывает влияние гибкости молекулы лекарственного средства, а также важность качества подгонки. В первом приближении потеря трансляционной и вращательной энтропии не увеличивается пропорционально размеру молекулы, а потеря внутренних конформационных степеней свободы зависит от числа вращающихся связей. Это объясняет, почему жесткие аналоги (если они содержат правильную конформацию фармакофора) часто гораздо более активны и демонстрируют более высокую степень избирательности, чем более гибкие.

Вклад полярных и электростатических взаимодействий часто чрезмерно подчеркивается, поскольку пренебрегают переносом лиганда из водной среды в сайт связывания и особенно отрицательным влиянием десольватации. Но гидрофобные взаимодействия также имеют негативные последствия; ограниченная растворимость делает транспортировку и распределение в биологической системе более сложной или даже невозможной, если молекула лекарственного средства становится слишком липофильной. Наиболее важные благоприятные и неблагоприятные энтальпийные и энтропийные вклады в взаимодействия между рецепторами и рецепторами суммированы на рисунке 3. Независимо от того, является ли специфическое взаимодействие лиганд-рецептор энтальпией или энтропией,

можно определить с помощью термодинамического анализа. Согласно Хитцерманну [127], нельзя сделать общих выводов о том, стабилизируется ли связывание агонистов и антагонистов энтальпией или энтропией. Даже изменение конформации рецептора, вызванное связыванием агониста и продуцирующей рецепторный ответ, не обязательно должно быть результатом энтальпийного вклада. Кроме того, ионная сила среды, например, наличие или отсутствие NaCl в случае опиатных рецепторов может привести к противоположным результатам.

Были рассмотрены различные вклады в общую свободную энергию связывания лиганда с биологической макромолекулой [38, 116, 118, 119, 128, 129] и важную роль окружающих молекул воды [59, 130].

1.2. Дескрипторы

Дескрипторы, которые кодируют определенные структурные особенности и свойства, необходимы для количественной корреляции биологической активности с химическими обеспечить рациональный выбор заместителей при разработке новых аналогов в соответствии с их расстояниями в многомерном пространстве структурами. Особое значение имеют физико-химические свойства, которые непосредственно связаны с межмолекулярными силами, участвующими в взаимодействии лекарство-рецептор, а также с транспортными и распределительными свойствами лекарств. В этом отношении наиболее важны гидрофобные, полярные, электронные и стерические свойства; чаще всего параметры поляризуемости считаются тесно связанными со стерическими параметрами и обсуждаются вместе с этими параметрами, хотя они, безусловно, различны (если структурные изменения подходят).

В этой главе физико-химические и другие параметры будут дифференцироваться в соответствии с:

- Дескрипторами липофильности, например, коэффициенты распределения и хроматографические параметры.

- Deskрипторами поляризуемости, например, молярная рефрактерность, молярный объем и парашор.
- Электронными дескрипторами, например, Константы Гамметта σ , полевые и резонансные параметры, полученные из спектроскопических данных, константы переноса заряда, дипольные моменты и квантово-химические параметры.
- Пространственными дескрипторами, полученными из линейных соотношений свободной энергии или из геометрических соображений.
- Deskрипторы, такие как молекулярная масса, геометрические параметры, конформационные энтропии, индексы связности и другие топологические параметры.
- Индикаторами (свободные значения типа Wilson, фиктивные).

Сотни различных дескрипторов были использованы в исследованиях QSAR за последние 30 лет. Некоторые ученые действительно изобретательны, постоянно придумывая новые дескрипторы, утверждая, что все проблемы в QSAR исчезнут с их использованием. Хотя некоторые дескрипторы действительно могут оказаться полезными, большинство из них не имеют общего значения. Определенный эффект насыщения наблюдается в этой области исследований QSAR. С другой стороны, по-прежнему отсутствуют адекватные дескрипторы для описания некоторых важных взаимодействий, таких как мембранное разделение лекарств, сила водородных связей, влияние энергии десольватации на сродство к рецептору и стерические взаимодействия с (чаще всего неизвестный) сайтом связывания. Ранние коллекции дескрипторов, опубликованные Ганчем и коллегами [131, 132] включают значения π , σ , σ_r , F, R и MR 284 ароматических заместителей. Другая публикация содержала π , π^- , F, R (разделенные по орто-, мета- и пара-замещению) и MR-значения 34 ароматических заместителей [133]. В книге «Константы заместителей для корреляционного анализа в химии и биологии», опубликованной Ганчем и

Лео в 1979 году, π , MR, E8 и все известные значения σ были собраны для около 2000 различных заместителей [50]. Две таблицы включали π , MR, F, R, $\sigma\pi$, $\sigma\pi$ и индикаторные переменные для акцептора водородной связи и донорные свойства 166 ароматических заместителей, а также гидрофобные фрагментарные константы, акцептор водородной связи и свойства донора, MR и F значения 103 алифатических заместителей. По кластерному анализу заместители были разделены соответственно на 20, 10 и 5 различных групп, чтобы обеспечить рациональный выбор заместителей при разработке новых аналогов в соответствии с их расстояниями в многомерном пространстве параметров [50, 134].

Физико-химические дескрипторы, основные теории и их использование в исследованиях QSAR были рассмотрены в нескольких других монографиях [39-43, 48, 49, 51, 52, 54-56] и в специальных статьях (например, [135-138]).

Наиболее полный сборник физико-химических параметров включает более 220 различных параметров и содержит таблицу из 58 параметров 59 различных заместителей, групп и фрагментов вместе с корреляционной матрицей всех 58 параметров [138]. Позднее были добавлены еще 16 параметров того же набора заместителей; было оценено около 18% недостающих значений [139, 140]. Все значения параметров включены в коммерческую базу данных DESBASE [141]. Собственный сборник Эли Лилли, упомянутый в литературе, но недоступный для научного сообщества, содержит более 17 000 физико-химических параметров 3000 различных заместителей [142], организованных в реляционной базе данных для быстрого и эффективного извлечения.

Анализ основных компонентов (PCA) уменьшает матрицы данных физико-химических свойств до меньших ортогональных размеров [139, 140, 143 – 147]. Шесть свойств (энергия водной сольватации, коэффициент распределения, точка кипения, молярная рефракция, объем и энтальпия испарения) 114 жидких соединений почти количественно коррелируют с первыми двумя основными компонентами, называемыми В (объемным) и С

(когезионностью), которые были полученных из большого количества физико-химических свойств этих соединений [144 - 146]. Некоторые другие свойства могут быть объяснены В и С и дополнительными второстепенными компонентами D, E и F; только несколько свойств, например, дипольный момент, температура плавления и молекулярный вес, не коррелируют с этими параметрами. Параметры ВС (DEF) использовались для «прогнозирования» некоторых известных свойств 139 дополнительных соединений с относительно высокой степенью надежности [145]. Анализ основных компонентов нескольких электронных параметров показал, что более 90% информации содержится в их первом главном компоненте [147]. PCA из семи различных химических описателей (липофильные, электронные и стерические свойства 28 ароматических заместителей) приводило к четкой группировке заместителей в графике нагрузок первых двух компонентов в соответствии с акцепторами водородной связи, донорами водородной связи, алкильными группами, и галогенов [148]. Девять переменных дескриптора, такие как π , MR, σ_m , σ_p , и параметры STERIMOLL, V_i , V_{ii} , V_{iii} и V_{iv} набора из 100 ароматических заместителей (таблица 1) были исследованы с помощью PCA [149]. Первые четыре компонента содержали 39%, 21%, 9% и 7% (вместе объясняя 76%) от общей дисперсии; из весов и нагрузок компонентов можно сделать вывод, что первый компонент в основном связан со стерической массой и гидрофобностью, второй компонент с электронными параметрами, а третий - с гидрофобностью и формой.

Таблиц 1- Параметры ароматических заместителей

Заместители	π	MR ^{a)}	σ_m	σ_p	L	V_i ^{b)}	V_{ii} ^{b)}	V_{iii} ^{b)}	V_{iv} ^{b)}
Br	0.86	0.888	0.39	0.23	3.83	1.95	1.95	1.95	1.95
Cl	0.71	0.603	0.37	0.23	3.52	1.80	1.80	1.80	1.80
F	0.14	0.092	0.34	0.06	2.65	1.35	1.35	1.35	1.35
SO ₂ F	0.05	0.865	0.80	0.91	3.50	2.03	2.70	2.45	2.51

Продолжение таблицы

I	1.12	1.394	0.35	0.18	4.23	2.15	2.15	2.15	2.15
NO	-1.20	0.520	0.62	0.91	3.44	1.70	2.44	1.70	1.70
NO ₂	-0.28	0.736	0.71	0.78	3.44	1.70	1.70	2.44	2.44
N ₃	0.46	1.020	0.27	0.15	4.62	1.50	4.18	2.34	2.57
H	0.00	0.103	0.00	0.00	2.06	1.00	1.00	1.00	1.00
OH	-0.67	0.285	0.12	-0.37	2.74	1.35	1.93	1.35	1.35
SH	0.39	0.922	0.25	0.15	3.47	1.70	2.33	1.70	1.70
NH ₂	-1.23	0.542	-0.16	-0.66	2.93	1.50	1.50	1.84	1.84
SO ₂ NH ₂	-1.82	1.228	0.46	0.57	3.82	2.11	3.07	2.67	2.67
NHNH ₂	-0.88	0.844	-0.02	-0.55	3.40	1.50	2.82	1.84	1.84
N=CCl ₂	0.41	1.835	0.21	0.13	5.65	1.70	1.80	1.84	4.54
CF ₃	0.88	0.502	0.43	0.54	3.30	1.98	2.61	2.44	2.44
OCF ₃	1.04	0.786	0.38	0.35	4.57	1.35	3.33	2.44	2.44
SCF ₃	1.44	1.381	0.40	0.50	4.89	1.70	3.69	2.44	2.44
CN	-0.57	0.633	0.56	0.66	4.23	1.60	1.60	1.60	1.60
NCS	1.15	1.724	0.48	0.38	4.29	1.50	2.24	1.64	1.76
SCN	0.41	1.340	0.41	0.52	4.08	1.70	4.45	1.70	1.70
SO ₂ CF ₃	0.55	1.286	0.79	0.93	4.11	2.11	3.64	2.67	2.67
NHCN	-0.26	1.014	0.21	0.06	3.53	1.50	3.08	1.90	1.90
CHO	-0.65	0.688	0.35	0.42	3.53	1.60	1.60	2.00	2.36
CO ₂ H	-0.32	0.693	0.37	0.45	3.91	1.60	1.60	2.36	2.66
OCH ₂ CH ₃	0.38	1.247	0.10	-0.24	4.92	1.35	3.36	1.35	1.90
CH ₂ OCH ₃	-0.78	1.207	0.02	0.03	4.91	1.52	2.88	1.90	1.90
SC ₂ H ₅	1.07	1.842	0.18	0.03	5.24	1.70	3.97	1.90	1.90
NHC ₂ H ₅	0.08	1.498	-0.24	-0.61	4.96	1.50	3.42	1.90	1.90
SO ₂ C ₂ H ₅	-1.09	1.814	0.60	0.72	5.31	2.11	3.67	2.67	2.67
NMe ₂	0.18	1.555	-0.15	-0.83	3.53	1.50	2.56	2.80	2.80
PMe ₂	0.44	2.119	0.03	0.31	3.88	2.00	2.97	2.84	3.29

Продолжение таблицы

cyclo-PROPYL	1.14	1.353	-0.07	-0.21	4.14	1.98	2.24	2.29	2.88
CO ₂ C ₂ H ₅	0.51	1.747	0.37	0.45	5.96	1.90	1.90	2.36	4.29
(CH ₂) ₂ CO ₂ H	-0.29	1.652	-0.03	-0.07	5.96	1.52	3.05	2.35	2.67
NHCO ₂ C ₂ H ₅	0.17	2.118	0.07	-0.15	4.45	1.50	4.97	1.90	5.57
CHMe ₂	1.53	1.496	-0.07	-0.15	4.11	2.04	2.76	3.16	3.16
n-C ₃ H ₇	1.55	1.496	-0.07	-0.13	5.05	1.52	3.49	1.90	1.90
OCHMe ₂	0.85	1.706	0.10	-0.45	4.59	1.35	3.61	1.90	3.16
OC ₃ H ₇	0.85	1.706	0.10	-0.25	6.05	1.35	4.30	1.90	1.90
SC ₃ H ₇	1.61	2.307	0.15	0.00	6.21	1.70	4.90	1.90	1.90
NHC ₃ H ₇	0.62	1.963	-0.24	-0.61	6.07	1.50	4.36	1.90	1.90
2-Thienyl	1.61	2.404	0.09	0.05	5.97	1.65	1.77	3.13	3.16
CH=CHCOCH ₃	-0.06	2.110	0.21	-0.01	5.80	1.60	3.24	1.83	3.73
COC ₃ H ₇	0.53	2.048	0.38	0.50	4.67	2.36	3.69	3.16	3.16
CO ₂ C ₃ H ₇	1.07	2.217	0.37	0.45	6.90	1.90	1.90	2.36	4.83
sec-C ₄ H ₉	2.04	1.959	-	-0.12	5.02	1.90	3.16	2.76	3.49
iso-C ₄ H ₉	-	1.959	-	-0.12	5.05	1.52	4.21	1.90	3.16
n-C ₄ H ₉	2.13	1.969	-0.08	-0.16	6.17	1.52	4.42	1.90	1.90
tert-C ₄ H ₉	1.98	1.962	-0.10	-0.20	4.11	2.59	2.97	2.86	2.86
OC ₄ H ₉	1.55	2.166	0.10	-0.32	6.99	1.35	4.79	1.90	1.90
NHC ₄ H ₉	1.16	2.426	-0.34	-0.51	7.01	1.50	4.97	1.90	1.90
C ₅ H ₁₁	2.67	2.426	-0.08	-0.16	7.11	1.52	4.94	1.90	1.90
C ₆ H ₅	1.96	2.426	0.06	-0.01	6.28	1.70	1.70	3.11	3.11
N=NC ₆ H ₁₅	1.69	3.131	0.32	0.39	8.43	1.70	1.70	1.92	4.31
CH ₂ Br	0.79	1.339	0.12	0.14	4.09	1.52	3.75	1.95	1.95
CH ₂ Cl	0.17	1.049	0.11	0.12	3.89	1.52	3.46	1.90	1.90
CH ₂ I	1.50	1.886	0.10	0.11	4.36	1.52	4.15	2.15	2.15
NHCHO	-0.98	1.031	0.19	0.00	4.22	1.50	1.50	1.94	3.61
CONH ₂	-1.49	0.981	0.28	0.36	4.06	1.60	1.60	2.42	3.07

Продолжение таблицы

CH=NOH	-0.38	1.028	0.22	0.10	4.88	1.60	1.60	1.92	3.11
CH ₃	0.56	0.565	-0.07	-0.17	3.00	1.52	2.04	1.90	1.90
NHCONH ₂	-1.30	1.372	-0.03	-0.24	5.09	1.84	1.84	1.94	3.61
NHCSNH ₂	-1.40	2.219	0.22	0.16	4.62	1.50	4.18	2.34	2.57
OCH ₃	-0.02	0.787	0.12	-0.27	3.98	1.35	2.87	1.90	1.90
CH ₂ OH	-1.03	0.719	0.00	0.00	3.97	1.52	2.70	1.90	1.90
SOCH ₃	-1.58	1.370	0.52	0.49	4.03	1.60	2.93	2.49	3.36
OSO ₂ CH ₃	-0.88	1.699	0.39	0.36	4.03	1.35	3.86	1.90	3.57
SO ₂ CH ₃	-1.63	1.349	0.60	0.72	4.37	2.11	3.15	2.67	2.67
SCH ₃	0.61	1.382	0.15	0.00	4.30	1.70	3.26	1.90	1.90
NHCH ₃	-0.47	1.033	-0.30	-0.84	3.53	1.50	3.08	1.90	1.90
NHSO ₂ CH ₃	-1.18	1.817	0.20	0.03	4.06	1.50	1.90	3.59	3.88
CF ₂ CF ₃	1.68	0.923	0.47	0.52	4.11	1.98	3.64	2.44	2.44
C≡CH	0.40	0.955	0.21	0.23	4.66	1.60	1.60	1.60	1.60
NHCOCF ₃	0.08	1.430	0.30	0.12	-	-	-	-	-
CH ₂ CN	-0.57	1.011	0.16	0.01	3.99	1.52	4.12	1.90	1.90
CH=CHNO ₂	0.11	1.642	0.32	0.26	4.29	1.60	3.24	.183	4.21
CH=CH ₂	0.82	1.099	0.05	-0.02	4.29	1.60	1.60	2.00	3.09
COCH ₃	-0.55	1.118	0.38	0.50	4.06	1.90	1.90	2.36	2.93
SCOCH ₃	0.10	1.842	0.39	0.44	5.19	1.70	4.01	1.90	1.90
OCOCH ₃	-0.64	1.247	0.39	0.31	4.87	1.35	3.68	1.90	1.90
CO ₂ CH ₃	-0.01	1.289	0.37	0.45	4.85	1.90	1.90	2.36	3.36
NHCOCH ₃	-0.97	1.493	0.21	0.00	5.15	1.50	3.61	1.90	1.94
CONHCH ₃	-1.27	1.457	0.35	0.36	5.00	1.60	2.23	2.42	3.07
CH ₂ CH ₃	1.02	1.030	-0.15	-0.15	4.11	1.52	2.97	1.90	1.90
OC ₆ H ₅	2.08	2.768	0.25	-0.03	4.51	1.35	5.89	3.11	3.11
SO ₂ C ₆ H ₅	0.27	3.320	0.61	0.70	5.82	2.11	6.01	2.67	2.67
OSO ₂ C ₆ H ₅	0.93	3.670	0.36	0.33	8.20	1.61	3.64	1.80	3.57

Продолжение таблицы

NHC_6H_5	1.37	3.004	-0.12	-0.40	4.53	1.50	5.95	3.11	3.11
$\text{NHSO}_2\text{C}_6\text{H}_5$	0.45	3.788	0.16	0.01	-	-	-	-	-
cyclo-HEXYL	2.51	2.669	-0.15	-0.22	6.17	2.04	3.49	3.16	3.16
COC_6H_5	1.05	3.033	0.34	0.43	4.57	2.36	5.98	3.11	3.11
OCOC_6H_5	1.46	3.233	0.21	0.13	8.15	1.70	4.40	1.70	1.84
$\text{N=CHC}_6\text{H}_5$	-0.29	3.301	-0.08	-0.55	8.40	1.70	3.55	1.80	3.66
$\text{CH=NC}_6\text{H}_5$	-0.29	3.301	0.35	0.42	8.50	1.70	1.70	2.36	4.07
$\text{CH}_2\text{C}_6\text{H}_5$	2.01	3.001	-0.08	-0.09	3.63	1.52	6.02	3.11	3.11
$\text{CH}_2\text{OC}_6\text{H}_5$	1.66	3.219	0.03	0.04	8.19	1.52	3.09	3.11	3.11

^{a)}MR масштабируется в 0,1 раза, как обычно

^{b)} $V_i - V_{iv}$, определены как в [149]; V_i - наименьшее значение, ортогональное L , V_{ii} противоположно V_i ; V_{iii} и V_{iv} , ортогональны V_i и V_{ii} и расположены таким образом, что $V_{iii} < V_{iv}$.

Анализ основных компонентов был также использован для исследования 74 параметров 59 различных заместителей [139, 140]. Пять значимых компонентов содержали около 84% от общей дисперсии; липофильные, стерические и электронные параметры были хорошо разделены на графике первых двух компонентов (содержащих 33% и 25% информации), в то время как несколько менее четко определенных свойств, а также случайная величина были близки к началу этой диаграммы. После отбрасывания 12 нерелевантных дескрипторов три значимых компонента объяснили 77% общей дисперсии (38%, 28% и 11%). Липофильность интерпретировалась как связанная с объемом (т. е. размером заместителей) и полярностью, что также подтверждалось соответствующими комбинациями таких параметров в уравнениях регрессии [140].

Аминокислоты характеризовались основным компонентным анализом свойств их боковой цепи [150, 151]; во-первых, для 20 закодированных аминокислот были получены три основных компонента Z_1 , Z_2 и Z_3 (интерпретированные как связанные с гидрофильностью, объемными цепями

боковой цепи и электронными свойствами) [150], а затем новые значения z были получены [151] с помощью частичных наименьших квадратов (PLS) анализа свойств боковой цепи этих аминокислот и дополнительных 35 некодированных (неестественных) аминокислот. Использование этих шкал (вместо исходных переменных) было рекомендовано для анализа структурной активности.

Оценки основных компонентов имеют несколько преимуществ по сравнению с исходными переменными:

Индивидуальная ошибка одного значения параметра существенно не искажает значения показателей основного компонента.

Проблема отсутствия данных значительно сокращается.

Корреляция значений биологической активности с использованием первых нескольких компонентов (которые могут быть истолкованы с точки зрения липофильности, объемных и электронных свойств) дает четкую картину присущих свойств, которые могут отвечать за изменение значений активности.

С другой стороны, есть серьезные недостатки. Значения основных компонентов заместителей явно зависят от правильного выбора исходных переменных. В одном из приведенных выше случаев пять из семи переменных были электронными параметрами [148]. В другом исследовании пять из девяти переменных были связаны с размером заместителей [149]. Еще одним недостатком является то, что более высокие компоненты больше не могут интерпретироваться в физико-химических терминах. В настоящее время использование метода PLS [26, 27], включая все физико-химические параметры в блоке переменных X , представляется лучшей альтернативой.

1.2.1. Дескрипторы липофильности

Ни одно физико-химическое свойство не имеет большого интереса к исследованиям QSAR как липофильность [152, 153] из-за её прямой зависимости от растворимости в фазах индукции, проникновения мембраны

и ее (краткого энтропийного) вклада в связывание лиганда на сайте рецептора. Было опубликовано несколько монографий [49, 50, 54, 56] и многочисленные обзоры (например, [138, 154-158]). Липофильность определяется разделением соединения между водной и неводной фазой. В то время как, ранние определения коэффициента распределения P , относящиеся к «легкой» и «тяжелой» фазам (что приводит к осложнениям в случае органических растворителей, имеющих более высокую плотность, чем вода), в настоящее время P определяется как отношение концентраций вещества в органической и водной фазах двухкамерной системы в равновесных условиях (уравнение 12). Из-за возможной ассоциации растворенного вещества в органической фазе (например, димеры карбоновых кислот) коэффициенты распределения должны измеряться при низких концентрациях или значения P должны быть экстраполированы до бесконечного разбавления растворенного вещества в системе.

$$P = C_{\text{org}}/C_{\text{aq}}(12)$$

В то время как коэффициенты распределения из многих органических систем с органическим растворителем использовались в отношениях структуры и активности. *n*-Октанол стал органическим растворителем, выбранным после новаторской работы Ганча по коэффициентам распределения в замещенные феноксиуксусные кислоты *n*-октаноле/воде и параметра липофильности π из этих коэффициентов распределения [14, 15, 17, 18, 153, 159, 160].

n-Октанол / вода имеет много важных преимуществ по сравнению с другими системами [153, 160]:

Это подходящая модель липидных составляющих биологических мембран из-за ее длинной алкильной цепи и полярной гидроксильной группы.

Его гидроксильная группа является донором и акцептором водородной связи, взаимодействующая с большим разнообразием полярных групп различных растворенных веществ.

Несмотря на липофильный характер, он растворяет гораздо больше органических соединений, чем алканы, циклоалканы или ароматические растворители.

В то время как водная фаза системы н-октанол/вода практически не содержит октанола в равновесии, фаза октанола растворяет значительное количество воды (2,3 моль * л⁻¹, что соответствует молярному отношению н-октанола/воды ≈ 4/1) ; поэтому полярные группы не нужно обезвоживать при их переносе из водной фазы в органическую фазу.

н-Октанол имеет низкое давление пара, что позволяет воспроизводить измерения; с другой стороны, давление его паров достаточно велико, чтобы обеспечить его удаление в мягких условиях.

н-Октанол является ультрафиолетовым прозрачным в широком диапазоне, что делает количественное определение многих соединений относительно легким.

На сегодняшний день большинство коэффициентов распределения были измерены в н-октанол/воде [50, 161]; также рассчитанные коэффициенты распределения относятся к этой системе.

Наконец, многие отношения липофильности и активности, используют коэффициенты распределения н-октанол / вода или параметры липофильности, полученные из них, подтверждают значимость этой системы (например, [18, 19, 162]).

Использование единой стандартной системы для разделения лекарственных средств оправдано уравнением Колландера (уравнение 13) [163], которое связывает коэффициенты распределения из разных систем растворителей.

$$\log P_2 = a \log P_1 + c \quad (13)$$

Например, наклон $\alpha = 1,04$ до $1,06$ и коэффициенты корреляции $r = 1,00$ были получены для разбиения гомологичной серии 4-алкилпиридинов в разных системах органических растворителей. Значения $c = - 1,20, - 1,14, - 0,78, -0,50$ и $0,46$ приведены для органических фаз гексадекана, октана,

дибутилового эфира, четыреххлористого углерода и хлороформа (на основе н-октанола в качестве эталона), что соответствует количественному масштабу полярность растворителя [164]. Многочисленные другие примеры (например, рисунок 4) [154,165,166] подтверждают обоснованность отношений типа Колландера. Для некоторых растворителей были получены различные уравнения для доноров и акцепторных растворов водородной связи [154]. Собственные ограничения уравнения Колландера обсуждались Тейлором [153].

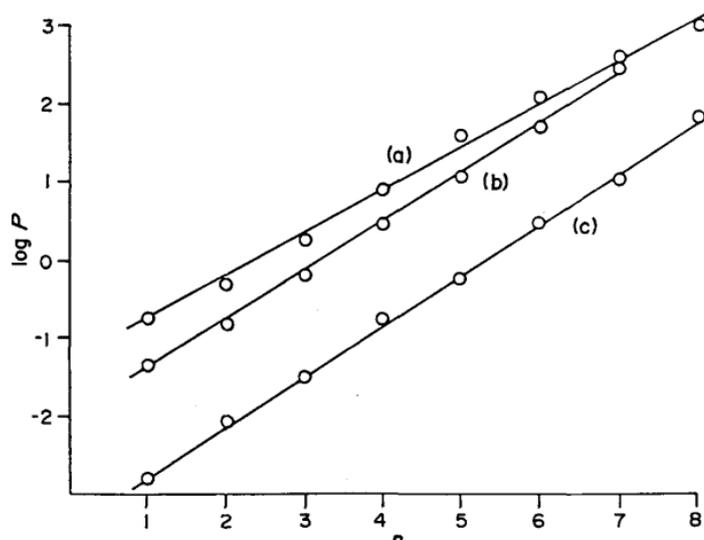


Рисунок 4 $-\log P$ гомологичных спиртов в зависимости от n (число атомов алкильного углерода)

(а) н-октанол / вода, наклон 0,54 (вклад липофильности метиленовой группы); (б) хлороформ / вода, наклон 0,62; (с) циклогексан / вода, наклон 0,65 [153]. Уравнение 14 коррелирует разделение некоторых спиртов между мембранами эритроцитов человека и водным буфером с коэффициентами распределения н-октанол / вода [167].

$$\log P_{\text{membrane}} = 1.003(\pm 0.13) \log P_{\text{oct}} - 0.883(\pm 0.39) \quad (14)$$

$$(n = 5; r = 0.998; s = 0.082)$$

Опубликованы несколько сотен линейных соотношений между различными видами (в основном неспецифическими) биологическими данными и коэффициентами распределения н-октанол / вода [18, 162]. Однако выбор н-октанола / воды в качестве стандартной системы для разделения лекарственных средств должен быть пересмотрен в свете

некоторых недавних результатов. Основной компонентный анализ коэффициентов распределения из разных систем растворителей [168-170] показывает, что липофильность зависит от эффектов объемного, полярного и водородного связывания растворенных веществ [169]; изотропные поверхностные области, то есть области, где молекулы воды не связываются и не гидратируются поверхностными областями, были сопоставлены с первой и второй основными компонентами такого анализа [170]. Потенциал водородной связи I_H различных функциональных групп была впервые определена Сейлером (таблица 2) [171] как разница между коэффициентами распределения циклогексана / воды и н-октанола / воды (уравнение 15).

$$\Delta \log P = \log P_{\text{oct}} - \log P_{\text{cyclohexane}} = \Sigma I_H - 0.16 \quad (15)$$

$$(n = 195; r = 0.967; s = 0.333; F = 107)$$

Таблица 2- Значения потенциала водородной связи I_H различных заместителей и подструктур (уравнение 22) [171]

Заместитель/основание	I_H
Ароматический -COOH	2,87
Ароматический -OH	2,6
-CONH	2,56
-SO ₂ NH	1,93
Алифатический -OH	1,82
Алифатический -NH ₂	1,33
Ароматический -NH ₂	1,18
-NR ₁ R ₂ (R ₁ ,R ₂ <> H)	0,55
-NO ₂	0,45
>C=O	0,31
-C=N	0,23
-O-	0,11
ortho- Substitution to -OH, -COOH, -NR ₁ R ₂	-0,62

Соответствующие шкалы могут быть получены, например, от значений $\log P$, измеренных в *n*-октаноле / воде, гептан / воде и других системах [172]. Так же от «эффекта увлечения водой» (способность растворенного вещества переносить молекулы воды из водной фазы в полярный органический растворитель, такой как, например, дибутиловый эфир) [173, 174], а также константы равновесия воды и газовой фазы различных растворов [175]. Было обсуждено значение параметров водородной связи в исследованиях QSAR [175-178] и приведены примеры их применения [179-183], наиболее интересными из которых являются отношения между проникновением гематоэнцефалического барьера H , -антигистаминные препараты и $\Delta \log P$, разница между коэффициентами раздела *n*-октанол / вода и циклогексан / вода (уравнения 16 и 17).

$$\log (C_{\text{brain}}/C_{\text{blood}}) = -0.604(\pm 0.17) \Delta \log P + 1.23(\pm 0.56) \quad (16)$$

$$(n = 6; r = 0.980; s = 0.249; F = 98.0)$$

$$\log (C_{\text{brain}}/C_{\text{blood}}) = -0.485(\pm 0.16) \Delta \log P + 0.889(\pm 0.50) \quad (17)$$

$$(n = 20; r = 0.831; s = 0.439; F = 40.2)$$

Для моделирования различных мембран и тканей [184, 185] было предложено использование различных растворителей, алканов (инертных), *n*-октанола (амфипрота), хлороформа (донора водородной связи) и дипеларгоната пропиленгликоля (PGDP, акцептор водородной связи); было измерено большое количество коэффициентов распределения PGDP / воды. Хотя *n*-октанол, по-видимому, является подходящей моделью амфипротических областей мембраны, PGDP напоминает больше его липидную часть. Поэтому PGDP следует рассматривать как дополнительный растворитель, а не как замену *n*-октанола [153].

Таблица 3- Мембранный и *n*-октанол / буферный (pH = 7,2) коэффициенты распределения [114]

Лекарственное средство	Биологические мембраны	н-Октанол/Буффер
Вау Р 8857	125,000	42
Нисолдипин	6,000	38
Нимодипин	5,000	730
Пропранодол	200	18
Ацетилхолин	32	0,003
Тимолол	16	0,7
Этанол	3	0,6

Систематические исследования Гербета [114, 115, 186-190], основанные на измерениях коэффициента перераспределения мембраны/воды и на малоугловых нейтронографических экспериментах. Исследования показывают, что изотропная двухфазная система растворителей не может быть подходящей моделью для липидного бислоя мембраны с их внутренней гидрофобной частью и внешней, полярной и отрицательно заряженной поверхностью фосфолипидов [191]. Действительно, коэффициенты биологической мембраны/буферного распределения лекарств значительно отличаются от показателей, измеренных в системах н-октанол/вода (таблица 3). Необходимо рассматривать не только разделение лекарств на мембрану, но и их ориентацию относительно сайта связывания на мембранном рецепторе или канальном белке (рисунок 1) [115]

Важная роль мембран в действии лекарств также была отмечена Сейделем [192 - 194]. При взаимодействии лекарств / мембран, мембрана воздействует на молекулы лекарственного средства:

- Диффузия через мембрану может стать лимитирующей стадией.
- Мембрана может препятствовать диффузии в активный сайт.
- Мембрана может связывать или накапливать лекарства.
- Сольватация лекарственного средства в мембране может привести к конформационному изменению его структуры.

- И наоборот, препарат действует на свойства мембраны:
- Препарат может изменять конформацию ацильных групп.
- Препарат может увеличить поверхность мембраны.
- Препарат может изменять толщину мембраны.
- Препарат может изменять текучесть мембраны.
- Препарат может изменять мембранный потенциал и гидратацию основных полярных групп.

Все эти эффекты имеют важное, до сих пор в основном игнорируемое влияние на активность наркотиков [194]. Различия в ингибирующей активности аналогов триметоприма (ТМП) в бесклеточных и цельноклеточных системах штаммов *Escherichiacoli* являются чувствительными и устойчивыми к ТМП, взаимодействию амфифильных бензиламинов с искусственными бислоями, взаимодействию нейролептиков с двухслойными мембранами (измерено по ЯМР методам), и репликация множественной лекарственной устойчивости амфифильными агентами может быть количественно описана в отношении этих эффектов [194].

Роль биомембран в опосредовании селективности подтипов рецепторов пептидов была исследована Швизером [195, 196], уравнение 18 определяет различные вклады в общее взаимодействие лекарственных мембран.

$$\Delta G_{\text{total}} = \Delta G_{\text{electrostatic}} - \Delta G_{\text{transfer}} + \Delta G_{\text{immobilization}} + \Delta G_{\text{amphiphilic}} + \Delta G_{\text{dipole}} \quad (18)$$

1.2.2. Дескрипторы поляризуемости.

Молярный объем MV , молярная рефрактерность MR и парашор PA являются теоретически и практически тесно взаимосвязанными параметрами (уравнения 19-21 MW = молекулярная масса, ρ = плотность, n = показатель преломления, γ = поверхностное натяжение) [197]. Другим связанным параметром является молярная поляризация P_M (диэлектрическая постоянная ϵ вместо n^2 в уравнении 20) [198].

$$MV = MW/\rho \quad (19)$$

$$MR = MV * \frac{n^2-1}{n^2+2} \quad (20)$$

$$PA = MV * \gamma^{1/4} \quad (21)$$

Молярный объем сам по себе не является строго аддитивным, но скорректированные параметры объема MR и PA являются аддитивными конститутивными молекулярными свойствами, такими как logP и параметр σ Гаммета. В то время как молярная рефракторность привлекла большое внимание [50, 197-199], молярный объем [55, 200] и парашор [50, 201] редко использовались в исследованиях QSAR. MR все еще является хамелеоном [91] среди физико-химических параметров, несмотря на его широкое применение в исследованиях QSAR. Он коррелирует с липофильностью, молярным объемом и стерическим объемом. Конечно, из-за его компонента MW / ρ это действительно связано с объемом и размером заместителя. Но два последних утверждения о том, что MR основаны только на этих свойствах [91, 202], не могут быть приняты. Поправочный член, связанный с показателем преломления в MP, учитывает поляризуемость и, следовательно, размер и полярность определенной группы [138, 153, 197]. Чем больше полярная часть молекулы, тем больше ее значение MR. Даже гидрофобные заместители обладают слабым энтальпийным взаимодействием, обусловленным дисперсионными силами, в дополнение к их энтропийному вкладу в ΔG , происходящему от смещения упорядоченных молекул воды. Для гидрофобных заместителей следует ожидать тесную взаимосвязь между объемом, поверхностью, липофильностью и MR, которая разрушается, если включены полярные заместители. Никакой корреляции между MR и стерическим параметром E_s^c ($n = 142$, $r = 0,67$, $s = 12,82$) [197]. Поскольку показатель преломления n незначительно изменяется для большинства органических соединений, молярный объем (MV) равен обычно очень тесно связаны с MR. В вышеупомянутом наборе параметров 59 различных заместителей [138] MR коррелирует с:

- Объемом Ван-дер-Ваальса ($r = 0,97$);

- парашором ($r = 0,92$);
- объемным параметром В компонентов ВС(DEF) ($r = 0,84$);
- ароматическими ($r = 0,90$) и алифатическими ($r = 0,91$) фрагментарными объемными константами, а также параметров связности ${}^0X_{ar}^v$ ($r=0.95$) и ${}^1X_{ar}^v$ ($r=0.95$).

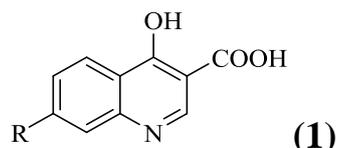
Не коррелирует с:

- Различными параметрами липофильности ($r = 0,53-0,62$);
- стерическими параметрами E_s ($r = 0,56$) и E_s^c ($r = 0,36$);
- параметром Чартона ($r = 0,58$) [158].

Значение молярных коэффициентов рефракции в уравнениях QSAR некоторых взаимодействий лиганд-фермент можно интерпретировать с помощью трехмерных структур. Эти исследования показывают, что заместители, моделируемые MR, связываются в полярных областях, а заместители, смоделированные π , связываются в гидрофобном пространстве [199, 203]. Соответственно, положительный знак MR в уравнении QSAR можно объяснить связыванием заместителей с полярной поверхностью, в то время как отрицательный знак или нелинейное соотношение указывает на ограниченную площадь или стерическое препятствие на этом сайте связывания. Чаще всего значения MR масштабируются в 0,1 раза для достижения разумных значений коэффициентов регрессии полученных QSAR-уравнений.

Различная природа MR по сравнению с гидрофобными и стерическими свойствами может быть обнаружена только в тех случаях, когда это позволяет правильный выбор заместителей. Одним из таких примеров является ингибирование малатдегидрогеназы 4-гидроксихинолин-3-карбоновыми кислотами (**1**), где взаимодействие лигандов с ферментом лучше описано MR (уравнение 22), чем на π (pI_{50} против π : $n = 13$, $r = 0,604$, $s = 0,716$) и ингибирование дыхания асцитных опухолевых клеток, где перенос в клетки или накопление в клетках более адекватно описывается π (уравнение

30, тот же набор соединений), что и MR (pI_{50} против MR: $n = 14$, $r = 0,699$, $s = 0,554$) (таблица 4) [204].



$$pI_{50}(\text{MDH}) = 0.70(\pm 0.17) \text{ MR} + 2.29(\pm 0.30) \quad (22)$$

$$(n = 13; r = 0.939; s = 0.315)$$

$$pI_{50}(\text{ascites}) = 0.46(\pm 0.11) \pi + 3.22(\pm 0.16) \quad (23)$$

$$(n = 14; r = 0.933; s = 0.280)$$

Таблица 4 - 7-Замещенные 4-гидроксихинолин-3-карбоновые кислоты (5) в качестве ингибиторов клеточного дыхания [204]

Заместитель R	pI_{50} , Малатдегидрогеназа	pI_{50} , Асцит	π^a	MR ^{b)}
H	-	2.98	0.00	0.103
Cl	2.44	3.84	0.55	0.603
F	1.98	3.30	0.06	0.092
OCH ₃	-	3.28	0.49	0.787
COCH ₃	3.04	3.10	-0.39	1.118
N(CH ₃) ₂	3.32	3.33	1.10	1.555
OCH ₂ C ₆ H ₅	4.49	4.41	1.81	3.174
OCH ₂ C ₆ H ₃ (3,4Cl ₂)	5.32	4.82	3.23	4.174
NO ₂	2.72	3.24	-0.40	0.736
CONH ₂	3.13	2.24	-1.18	0.981
COOH	2.97	2.24	-2.80	0.605
SO ₂ CH ₃	3.18	2.75	-1.39	1.349
OH	3.31	3.04	0.06	0.285
SO ₂ NH ₂	3.02	2.47	-1.36	1.228
SO ₃ ⁻	2.67	2.88 ^{c)}	-4.76	0.971

^{a)} рассчитывается из измеренных значений $\log P$

^{b)} значения MR масштабируются в 0,1 раза

^{c)} не включенные в расчет

Атомные молярные рефракторные вклады были определены. Программное обеспечение MEDCHEM также содержит процедуру вычисления значений MR.

1.2.3. Электронные дескрипторы.

Электронные свойства молекул [12, 40-43, 53, 57, 138, 205-208] могут быть описаны широким спектром различных параметров, например:

- Константами Гаммета σ ;
- полевыми и резонансными параметрами F и R;
- значениями pK_a ;
- параметрами, полученными из молекулярной спектроскопии;
- константами переноса заряда;
- дипольными моментами;
- параметрами водородного связывания;
- параметрами, полученными из квантово-химических расчетов, например, орбитальных энергий и частичных зарядов.

Поскольку все эти дескрипторы описывают влияние определенной группы или заместителя на распределение электронной плотности, все они были использованы в исследованиях QSAR. В отличие от глобальных молекулярных свойств, таких как липофильность и молярная рефракторность, они обычно относятся к определенному атому или группе (единственным исключением являются дипольные моменты). Например: если σ_p используется для описания влияния пара-заместителей на кислотность фенольной группы, этот термин моделирует способность фенольной группы взаимодействовать в качестве донора или акцептора водородной связи с (наиболее часто неизвестным) акцепторным или донорным атомом в сайте связывания.

Использование констант σ_P в исследованиях QSAR началось с раннего обнаружения Хансена [13], что токсичность замещенных бензойных кислот

против личинок комаров коррелирует с значениями σ Гаммета. Уравнение 24 было получено после исключения п-нитро-аналога (все аналоги: $n = 14$; $r = 0,711$; $s = 0,427$).

$$\log 1/C = 1.454(\pm 0.42) \sigma + 1.787(\pm 0.17) \quad (24)$$

$$(n = 13; r = 0.918; s = 0.243; F = 58.91)$$

Действительно, все аналоги можно описать намного лучше, используя вместо него (уравнение 25) [15]. Отношение $\log 1/C$ против σ в уравнении 38 существует только удача. Это связано с близкой связью между π и σ , если исключен нитро-аналог ($n = 13$; π против σ : $r = 0,91$); все уравнения пересчитывались с использованием значений π_{benzene} , σ_m и σ_p , представленных в таблице 5.

Таблица 5 - Токсичность замещенных бензойных кислот против личинок комаров [13,15]

Заместитель R	π^a	σ	$\log 1/C$ obsd.	$\log 1/C$ calc. ^{b)}
3,4,5-Tri-I	3.36	0.88	3.54	3.40
3,5-Di-I	2.24	0.70	2.85	2.80
4-I	1.12	0.18	2.31	2.20
3,4-Di-Cl	1.42	0.60	2.28	2.36
4-Cl	0.71	0.23	2.06	1.98
4-Br	0.86	0.23	2.03	2.06
3-Cl	0.71	0.37	2.00	1.98
3,4-CH=CH-CH=CH-	1.27	0.17	1.92	2.28
4-F	0.15	0.06	1.85	1.68
4-CH ₃	0.56	-0.17	1.66	1.90
H	0.00	0.00	1.64	1.60
4-OCH ₃	-0.02	-0.27	1.60	1.59
4-NO ₂	-0.28	0.78	1.52	1.45
4-OH	-0.67	-0.37	1.29	1.24

Значение π_{benzene}

Уравнение 25

$$\log 1/C = 0.535(\pm 0.09) \pi + 1.602(\pm 0.11) \quad (25)$$

$$(n = 14; r = 0.969; s = 0.151; F = 181.88)$$

При рассмотрении электронных эффектов нужно различать полевые (индуктивные) эффекты и резонансные эффекты. Из-за характерных особенностей бензольного кольца σ_m в основном описывает индуктивный эффект, а σ_p обозначает комбинацию обоих эффектов с преобладанием резонансного эффекта. За последние десятилетия в органической химии было разработано много разных масштабов, помимо σ_m и σ_p также σ^+ (для учета заместителей, которые подают электроны в систему ароматических колец прямым резонансным взаимодействием), σ^- (для соответствующих акцепторных заместителей), σ^o и σ^n (нормальные или неконтролируемые σ постоянные), σ_I и σ_R (индуктивные и резонансные вклады) и т. д.

В 1968 году Суэйн и Луптон [209] попытались остановить распространение σ -шкал. Они определили полевые и резонансные компоненты F и R, предположив, что любой набор значений σ может быть выражен взвешенной комбинацией $aF + bR$, что резонансный вклад в случае 4-замещенного бицикло [2.2.2] октан - карбоновых кислот ($b = 0$) и что резонансный вклад заместителя $N^+(CH_3)_3$ ($R = 0$) отсутствует. Они смогли сопоставить 43 различных электронных параметра с линейными комбинациями этих двух параметров (многие значения r больше 0,98), например, уравнения 26 и 27.

$$\sigma_m = 0.60(\pm 0.00)F + 0.27(\pm 0.00)R + 0.00(\pm 0.00) \quad (26)$$

$$(n = 42; r = 1.00; s = 0.00)$$

$$\sigma_p = 0.56(\pm 0.00)F + 1.00(\pm 0.00)R + 0.00(\pm 0.00) \quad (27)$$

$$(n = 42; r = 1.00; s = 0.00)$$

Ганч переопределил значения F и R более последовательно (Уравнения 28 и 29). Значения F в уравнении 28 теперь находятся в правильном масштабе по сравнению со значениями R [50, 151].

$$F = 1.369(\pm 0.19)\sigma_m - 0.373(\pm 0.14)\sigma_p - 0.009(\pm 0.04) \quad (28)$$

$$(n = 14; r = 0.992; s = 0.042)$$

$$R = \sigma_p - 0.921F(29)$$

Было значительное обсуждение за [210, 211] и против [212- 216] обоснованности основополагающих допущений. Несмотря на этот (более философский) вопрос, разделение значений σ на индуктивные и резонансные эффекты представляется оправданным, по крайней мере, с практической точки зрения.

Недавняя сборка [217] содержит σ_m , σ_p и переопределенные значения F и R 530 заместителей вместе с значениями σ^+ , σ^- , R^+ (уравнения 30, значения R^+ π -электронных донорных заместителей, уравнение 32, значения R^+ заместителей-акцепторов π -электронов) и значения R^- (уравнение 33, значения R^- π -электронных донорных заместителей) 223 заместителей, а также некоторые другие электронные параметры, полученные из спектроскопических данных; по сравнению с нормальными значениями 92, значения R^+ и R^- π -электронных донорных заместителей значительно улучшены (уравнения 30 и 31).

$$R^+ = 1.90(\pm 0.26)R - 0.07(\pm 0.11) \quad (30)$$

$$(n = 29; r = 0.945; s = 0.170)$$

$$R^+ = 1.16(\pm 0.45)R - 0.01(\pm 0.07) \quad (31)$$

$$(n = 16; r = 0.829; s = 0.060)$$

$$R^- = 1.93(\pm 0.59)R + 0.19(\pm 0.11) \quad (32)$$

$$(n = 34; r = 0.758; s = 0.173)$$

Кроме того, в исследованиях QSAR были определены и использованы оптимизированные и нормированные значения S (полевое индуктивное (σ -возмущение связи) и значения P (возмущение резонансной π -связи, ортогональное S) [218]. Факторизация значений σ в другое поле и резонансные вклады рассмотрены в [217].

Орто-заместители представляют собой особые проблемы, поскольку их значения σ включают стерический вклад; кроме того, по сравнению с мета- и пара-заместителями многие орто-заместители вызывают конформационные изменения, иногда благоприятные для связывания, иногда очень

неблагоприятные; хотя проблема параметрирования орто-заместителей подробно обсуждалась [41, 207], она останется сложной проблемой в исследованиях QSAR.

Алифатические σ -константы (значения σ^*) определяются уравнением 33, где $\log k$ и $\log k_0$ - константы скорости кислотного (индекс А) и основного (индекс В) гидролиза RCOOR' и CH₃COOR' соответственно [219].

$$\sigma^* = \frac{\log \frac{k}{k_0} - \log \frac{k}{k_0}}{2.48} \quad (33)$$

Значения pK_a [53, 206] прямо отражают электронные свойства из-за определения значений Гамметта σ относительно констант ионизации бензойных кислот ($e = 1,00$); таким образом, они могут использоваться в качестве заменителей значений σ . Однако их использование неадекватно, если они используются для описания относительных количеств объединенной формы различных лекарств. Одним из редких исключений из этого правила является, например, (уравнение 34), в котором описаны антибактериальные активности сульфонамидов против кишечной палочки [220].

$$\log 1/C = 1.044(\pm 0.13)pK_a - 1.640(\pm 0.18) \log (\beta^* 10^{pK_a} + 1) + 0.275(\pm 0.65) \quad (34)$$

$$\log p = -5.96$$

$$pK_{a \text{ оптимальная}} = 6.22$$

$$(n = 39; r = 0.956; s = 0.275; F = 124.1)$$

Параметры, полученные из молекулярной спектроскопии, например, от ИК-измерений или ЯМР-измерений, относительно рано использовались в исследованиях QSAR (например, [138, 221-223]). Они чрезвычайно ценны при описании электронного влияния заместителей в системах, для которых не известны значения σ , что применимо к большинству гетероциклических соединений.

Константы переноса заряда СТ [224] и К [225, 226] были получены из констант ассоциации разных аналогов с тетрацианоэтиленом и 1,3,5-

тринитробензолом соответственно. Хотя предполагается, что взаимодействия переноса заряда важны для взаимодействия некоторых лекарств со своими сайтами связывания (например, в интеркаляции ДНК) [116], они не привлекли большого внимания в исследованиях QSAR. Значения известны лишь несколькими заместителями, но они определенно взаимосвязаны с другими электронными параметрами.

Использование дипольных моментов [207, 227] в исследованиях QSAR было предложено Лиеном [227-229]; групповые дипольные моменты 311 ароматических [228] и 214 алифатических заместителей [229] приведены в литературе. Определенной проблемой использования дипольных моментов является их направленность. Векторное дополнение должно использоваться для оценки дипольного момента соединений, имеющих более одного места замещения.

В дополнение к значениям I_H от Сейлера [191] было получено несколько других шкал, например, Ганчем и Лео (различающие донорные, акцепторные и нейтральные заместители) [50, 230] и Тафтом (значения РКНВ, полученные из комплексного образования оснований с п-фторфенолом в CCl_4) [231, 232, 207].

В отличие от огромного значения водородных связей в взаимодействиях рецептора лекарства, только несколько анализов QSAR были выполнены с использованием таких параметров. Это может отражать тот факт, что водородсвязывающая способность связана с некоторыми другими электронными параметрами, по крайней мере, внутри структурно связанных рядов. Более пессимистичное, но, тем не менее, правдоподобное объяснение может заключаться в том, что выбор параметров в исследованиях QSAR в основном зависит от их доступности и готовой к использованию таблицы в справочниках. Похоже, что предстоит большая работа в отношении водородных связей в взаимодействиях рецепторов лекарств и их надлежащего, то есть количественного рассмотрения в исследованиях QSAR.

Квантово-химические параметры [208] являются самостоятельными. Их частые злоупотребления в ранние героические времена QSAR дискредитировали их в течение длительного времени. Квантово-химические расчеты дают так много разных значений, например, чистые атомные заряды, плотности зарядов, суперделокализованности, электростатические потенциалы, значения для индуктивных, резонансных и поляризуемых эффектов, а также энергии HOMO и LUMO и т. д., что их некритическая комбинация, как это делали в эти ранние дни, часто заканчивалась случайными корреляциями. В настоящее время некоторые более подходящие приложения демонстрируют свою реальную ценность для исследований QSAR [233- 237], например, уравнение 35 (коррелирующая биоразлагаемость, выраженная как БПК = биологическая потребность в кислороде, с параметром разности атомных зарядов, $\Delta|\delta|_{x-y}$) [235] и уравнением 36 (мутагенность различных ароматических соединений, TA98 = результаты испытаний Эймса, с использованием кишечной сальмонеллы, штамм TA98) [238-239]. Преимущество ϵ_{LUMO} (= самая низкая незанятая молекулярная орбиталь) в уравнении 36, по сравнению с классическими значениями σ , является результатом возможности включения различных гетероароматических соединений, для которых не известны значения σ

$$\% \text{BOD} = 1015, \Delta|\delta|_{x-y} + 1.523. \quad (35)$$

$$(n = 197; r = 0.991; s = 3.822)$$

$$\log \text{TA98} = 0.65(\pm 0.16)\log P - 2.90(\pm 0.59) \log (\beta P + 1) - 1.38(\pm 0.25)\epsilon_{LUMO} + 1.88(\pm 0.39) I_1 - 2.89(\pm 0.81)I_a - 4.154 (\pm 0.58)(36)$$

$$\log \beta = -5.48 \quad \text{оптимальный } \log P = 4.93$$

$$(n = 188; r = 0.900; s = 0.886)$$

Классические и магнитные ароматизации гетероциклов были определены и коррелированы с биологической активностью [240].

Быстрые расчетные процедуры электронных эффектов в органических молекулах были предложены Гастейгером и Марсили [241-243] в качестве

альтернативы abinitio и полуэмпирическим квантово-химическим расчетам, которые являются относительно трудоемкими процедурами.

1.2.4. Стерические дескрипторы.

Стерические дескрипторы трудно описать из-за того, что трехмерные структуры сайтов связывания лекарств чаще всего неизвестны. Несколько обзоров стерических эффектов и пространственных дескрипторов [50, 52, 198, 219, 244-247] способствуют этой проблеме, но общее решение кажется по своей сути невозможным.

Константы E_s , как определено уравнением 37 (кислотно-катализируемый гидролиз RCOOR' против CH₃COOR') [219] были первыми параметрами, которые были использованы для описания стерических эффектов в исследованиях QSAR.

$$E_s = \log(k/k_0)_A \quad (37)$$

Ван-дер-Ваальсовы радиусы симметричных заместителей приведены в уравнении 38 ($r_{v(av)}$ = средний Ван-дер-Ваальсов радиус заместителей) [50]

$$E_s = -1.839r_{v(av)} + 3.484 \quad (38)$$

$$(n = 6; r = 0.996; s = 0.132)$$

Чартон [248-250] определил параметр пространственного заместителя u (r_v = минимальный Ван-дер-Ваальсовский радиус заместителя) (Уравнение 39), который сильно коррелирован с значениями E_s (уравнение 40) [50].

$$u_x = r_{vX} - r_{vH} = r_{vX} - 1.20 \quad (39)$$

$$E_s = -2.062(\pm 0.86)u - 0.194(\pm 0.10) \quad (40)$$

$$(n = 104; r = 0.978; s = 0.250)$$

Хэнкок [251] модифицировал значения E_s на E_s^c , скорректировав их для числа атомов α -водорода, n_H (уравнение 41).

$$E_s^c = E_s + 0.306(n_H - 3) \quad (41)$$

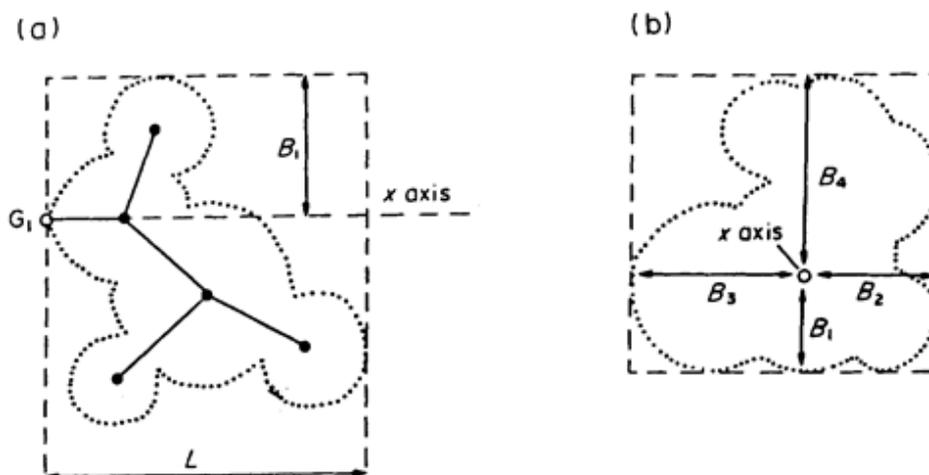


Рисунок 5 - Параметры STERIMOLL и B1-B4

Проекция заместителя (а) вдоль и (б) перпендикулярно оси подстановки (ось x) [335]; в соответствии с определением Скагерберга [169], $B_i = B_1$, $B_{ii} = B_4$, $B_{iii} = B_2$ и $B_{iv} = B_3$ [198].

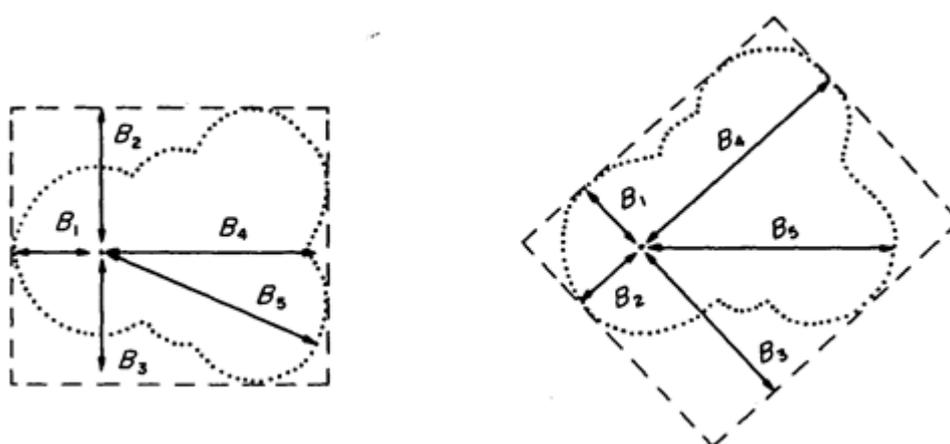


Рисунок 6 - Пересмотренные параметры STERIMOLL, B1 и B5 группы $ОСН_3$

B_1 - минимальная ширина, B_5 - максимальная ширина заместителя [247, 254].

Различные возможности являются результатом определения параметров B_2 - B_4 в зависимости от выбора B_1 [198].

Фуджита выражал E_s^c значения заместителей типа $CR'R_2R_3$ как взвешенную сумму отдельных E_s^c значений R_1 , R_2 и R_3 [252] для преодоления проблем, возникающих в связи с использованием стерических параметров несимметричных заместителей. Простой стерический параметр S_b был составлен на основе разветвления заместителей [253].

Для описания стерических эффектов в уравнениях QSAR были использованы многие другие параметры, относящиеся к размеру, например, объемы Ван-дер-Ваальса, молярные объемы, поверхность, доступная растворителю, молярная преломляющая способность и т. д.

Реальный прогресс был обусловлен определением параметров STERIMOLL, V_1 , V_2 , V_3 и V_4 [246, 247]. L определяется как длина заместителя вдоль оси его замещения родительскому скелету; параметры ширины V все ортогональны L и имеют углы 90 градусов друг к другу. Они расположены таким образом, что V_1 имеет наименьшее значение, а V_4 - наибольшее; в большинстве случаев V_4 близок к максимальной ширине заместителя (рисунок 5). Несколько другое определение было использовано Скагербергом [149]: V_i по-прежнему является наименьшим значением, ортогональным L , но V_{ii} противоположно V_i ; V_{iii} и V_{iv} , ортогональны V_i и V_{ii} и расположены таким образом, что $V_{iii} < V_{iv}$.

Из-за некоторой критики [50] и из-за ряда случайных корреляций в литературе QSAR, возникающих из некритических комбинаций линейных и квадратичных членов (10 возможных значений для одного сайта замещения и гораздо больше для нескольких разных сайтов), Верлуп уменьшил параметры ширины до V_1 , являющиеся наименьшей шириной, и V_5 , которая теперь является наибольшей шириной, ортогональной L , но не зависит от угла между V_1 и V_5 (рисунок 6) [247, 254]. Было рассмотрено использование параметров STERIMOL в исследованиях QSAR [198, 246, 247, 254].

Однако, как говорилось ранее, даже самое изящное описание стерического объёма заместителя не может справиться с проблемой неизвестной геометрии сайта связывания и его стерических ограничений. Количественное описание стерических эффектов является загадочной проблемой даже в случае известных трехмерных структур белка из-за:

Различной конформации различных лигандов,

Небольших (или даже больших) различий в режимах связи разных аналогов и вариации положения боковых цепей и даже опорных атомов белка в различных комплексах лиганд-белок.

Сборники различных стерических параметров приведены в ссылках. [138, 198, 246, 247].

1.3. Количественные модели

1.3.1. Экстратермодинамический подход (анализ Ганча)

Анализ Ганча [14, 15, 17, 18, 40-44, 255] коррелирует значения биологической активности с физико-химическими свойствами путем линейного множественного или нелинейного регрессионного анализа; таким образом, анализ Ганча действительно является моделью отношений собственности и собственности. Поскольку практически все параметры, используемые в анализе Ганча, являются линейными значениями свободной энергии (полученными из констант скорости или равновесия), термины «линейный подход, связанный с свободной энергией» или «экстратермодинамический подход» [255] иногда используются как синонимы для анализа Ганча. Кроме того, значения биологической активности, если они правильно определены, являются значениями линейной свободной энергии (например, константами связывания или ингибирования, константами скорости поглощения и распределения или сложными данными, которые соответствуют взвешенной комбинации нескольких таких единичных процессов).

Ранние попытки сопоставить значения биологической активности с липофильностью, выраженные, например, по коэффициентам растворимости или распределения [3-7], только объясняются неспецифические отношения структуры активности; применение концепции общего биологического уравнения Гамметта (уравнение 4) не удалось. Методологический прорыв произошел из предложения Фуджиты, в то время работавшего в группе Ганча [17], применить подход, использованный ранее Тафтом [219]. Как и он, Ганч

и Фуджита комбинировали различные физико-химические параметры в одном уравнении, например. Уравнение 42 (C = молярная концентрация, вызывающая определенный биологический эффект, k_1 , k_2 , k_3 = коэффициенты, определяемые методом наименьших квадратов, например линейный множественный регрессионный анализ, для соответствия биологическим данным).

$$\log 1/C = k_1 \log P + k_2 \sigma + k_3 \quad (42)$$

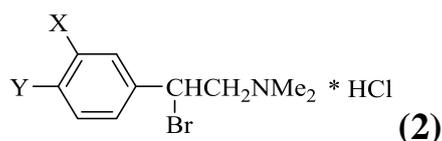
Для данных *in vivo* уравнение 42 было расширено до уравнения 55 путем включения параболической липофильности. Идеей уравнения 55 было то, что молекулы, которые являются слишком гидрофильными или слишком липофильными, не смогут пересекать липофильные или гидрофильные барьеры, соответственно. Таким образом, они будут иметь меньшую вероятность достичь места рецептора, чем молекулы с промежуточной липофильностью, легко растворяются в водных фазах, а также в липидных фазах.

$$\log 1/C = -k_1(\log P)^2 + k_2 \log P + k_3 \sigma + k_4 \quad (43)$$

Уравнения 42 и 43 являются лишь двумя примерами огромного широкого спектра уравнений Ганча. Более поздние стерические параметры были добавлены к этой общей модели и даже более поздние значения молярной рефракции. С помощью таких многопараметрических уравнений можно было описать гораздо более сложные зависимости биологической активности от физико-химических свойств, чем раньше. В течение последних трех десятилетий почти все возможные сочетания липофильных, поляризуемых, электронных и стерических параметров с дополнительными индикаторными переменными и без них использовались и коррелировались с значениями биологической активности в линейных, параболических и билинейных уравнениях. Здесь приведен только один иллюстративный пример, чтобы описать и объяснить правильное применение анализа Ганча. Грэм и Каррар [256] определили антиадренергическую активность ряда альфа-бромфенэтиламинов (2). Ганч и Лиен [257] получили уравнение 44,

который в это время считался наилучшим количественным описанием данных (в таблицу 10 включены только некоторые выбранные физико-химические параметры); несколько разные результаты получают, если различные шкалы, например. Значения π из системы феноксиуксусной кислоты, значения X-бензол или любые другие значения χ .

Таблица 10 - Антидренергическая активность мета- и пара-замещенных N,N-диметила-бромфенэтиламинов [218,256]



meta (X)	para (Y)	π	σ^*	E_s^{meta}	$\log 1/C$ obsd.	$\log 1/C$ calc. ^{a)}	$\log 1/C$ calc. ^{b)}
H	H	0.00	0.00	1.24	7.46	7.82	7.88
H	F	0.15	-0.07	1.24	8.16	8.09	8.17
H	Cl	0.70	0.11	1.24	8.68	8.46	8.60
H	Br	1.02	0.15	1.24	8.89	8.77	8.94
H	I	1.26	0.14	1.24	9.25	9.06	9.26
H	Me	0.52	-0.31	1.24	9.30	8.87	8.98
F	H	0.13	0.35	0.78	7.52	7.45	7.43
Cl	H	0.76	0.40	0.27	8.16	8.11	8.05
Br	H	0.94	0.41	0.08	8.30	8.30	8.22
I	H	1.15	0.36	-0.16	8.40	8.61	8.51
Me	H	0.51	-0.07	0.00	8.46	8.51	8.36
Cl	F	0.91	0.33	0.27	8.19	8.38	8.34
Br	F	1.09	0.34	0.08	8.57	8.57	8.51
Me	F	0.66	-0.14	0.00	8.82	8.78	8.65
Cl	Cl	1.46	0.51	0.27	8.89	8.75	8.77
Br	Cl	1.64	0.52	0.08	8.92	8.94	8.94
Me	Cl	1.21	0.04	0.00	8.96	9.15	9.08
Cl	Br	1.78	0.55	0.27	9.00	9.06	9.11
Br	Br	1.96	0.56	0.08	9.35	9.25	9.29
Me	Br	1.53	0.08	0.00	9.22	9.46	9.43
Me	Me	1.03	-0.38	0.00	9.30	9.56	9.47
Br	Me	1.46	0.10	0.08	9.52	9.35	9.33

^{a)} Уравнение 46.

^{b)} Уравнение 49.

$$\log 1/C = 1.22\pi - 1.59\sigma + 7.89(44)$$

$$(n = 22; r = 0.918; s = 0.238)$$

Позже Каммарата [358] представил уравнение 45, которое описывает метасоставления по их значениям TC и орто- и пара-заместителям стерическим параметром r_v^{para} . Однако стерический параметр имеет неправильный знак; положительное значение его коэффициента регрессии подразумевает, что стерическая масса увеличивает биологическую активность, что не может быть правдой. Кроме того, трудно понять, как электронный эффект может быть получен только для мета-заместителей.

$$\log 1/C = 0.747(\pm 0.26) \pi_m - 0.911(\pm 0.52) \sigma_m + 1.666(\pm 0.26) r_v^{para} + 5.769(\pm 0.45) \quad (45)$$

$$(n = 22; r = 0.962; s = 0.168; F = 74.01)$$

Соответственно, уравнение 45 было подвергнуто критике со стороны Унгера и Ганча [218] в примечательной статье, которая является важной вехой в развитии анализа Ганча. Они сформулировали правила вывода экстратермодинамических уравнений, которые суммируются здесь из-за их общей достоверности. Выбор независимых переменных. Следует попробовать широкий диапазон различных параметров, таких как $\log P$ или π , σ , MR и стерические параметры. Параметры MO и переменные индикатора не следует упускать из виду. Параметры, выбранные для «наилучшего уравнения», должны быть по существу независимыми (коэффициенты межкорреляции r не должны быть больше 0,6-0,7, исключениями являются комбинации линейных и квадратичных членов, таких как $(\log P)^2$ и $\log P$, которые обычно сильно взаимосвязаны, причем значения $r > 0,9$).

Обоснование выбора независимых переменных. Все «разумные» параметры должны быть подтверждены соответствующей статистической процедурой, например, пошаговым регрессионным анализом. «Наилучшее уравнение» обычно является наименьшим стандартным отклонением, причем все термины значительны (обозначены 95% доверительными интервалами

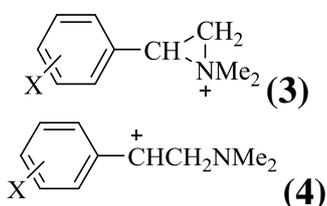
или последовательным F-критерием). В качестве альтернативы, уравнение с наивысшим общим значением F может быть выбрано как «лучшее».

Принцип скупости (Оксеймская бритва, Уильям Оккем, 1285 - 1349/50, английский философ и логик). Если все вещи (приблизительно) равны, нужно принять простейшую модель.

Количество терминов. Для избежания случайных корреляций должно быть не менее пяти-шести точек данных (это правило применяется только к наборам данных промежуточного размера, для небольших наборов данных могут быть разрешены дополнительные параметры, если они основаны на разумной модели, для больших наборов данных, например $n > 30$, эта рекомендация приводит к уравнениям, которые включают слишком много переменных).

Качественная модель. Важно иметь качественную модель, которая согласуется с известной физико-органической и биомедицинской химией рассматриваемого процесса.

Следуя этим рекомендациям, Унгер и Ганч пересмотрели механизм действия α -бромфенэтиламинов и утверждали, что КТ вместо КТ может быть лучшим электронным дескриптором, поскольку предполагается, что соединения взаимодействуют с рецепторным сайтом путем образования иона этилениминия (3), а затем бензиловый катион (4). Они получили уравнение [218], что дало гораздо лучшее описание данных, чем уравнение 44.



$$\log 1/C = 1.151(+0.19) \pi - 1.464(\pm 0.38) \sigma' + 7.817(\pm 0.19) \quad (46)$$

$$(n = 22; r = 0.945; s = 0.196; F = 78.63)$$

Из многих других уравнений, полученных Унгером и Ганчем [287] будут обсуждены только уравнения 47 и 48.

$$\log 1/C = 0.83(\pm 0.27) \pi_m + 1.33(\pm 0.20) \pi_p - 0.92(+0.50) \sigma_m^+ - 1.89(+0.57) \sigma' + 7.80(+0.17) \quad (47)$$

$$(n = 22; r = 0.966; s = 0.164)$$

$$\log 1/C = 0.86(\pm 0.30) \pi + 0.47(\pm 0.26) S_m - 0.36(\pm 0.21) S_p - 0.92(\pm 0.61) P_p + 0.62(\pm 0.49) r_v^{para} + 7.08(\pm 0.62) \quad (48)$$

$$(n = 22; r = 0.967; s = 0.167)$$

Позднее уравнение 61 было получено из предположения, что гидрофобные и электронные влияния идентичны в мета- и пара-положениях, но что может существовать стерическое препятствие для мета-заместителей (шкала E, противоположная знаку к шкале r_v , имеющая меньшие и даже отрицательные значения для больших заместителей, поэтому положительный знак коэффициента в уравнении 61 действительно указывает на отрицательное стерическое влияние больших заместителей).

$$\log 1/C = 1.259(\pm 0.19) \pi - 1.460(\pm 0.34) \sigma' + 0.208(\pm 0.17) E_s^{meta} + 7.619(\pm 0.24) \quad (49)$$

$$(n = 22; r = 0.959; s = 0.173; F = 69.24)$$

Уравнения 44-49 раскрывают типичную дилемму в анализе Ганча: в то время как уравнения 46-49 значительно лучше, чем уравнение 51 и основаны на более разумных предположениях, чем уравнение 45. На основе коэффициентов корреляции r уравнение 48 является предпочтительным; уравнение 47 представляется лучшим, если рассматривать стандартное отклонение s , гораздо лучший критерий. Различия в коэффициентах корреляции r и стандартных отклонений s уравнений 46-48 довольно малы.

Однако, если применить принцип суждения, уравнения 47 и 48 следует исключить, поскольку для такого небольшого набора данных включено слишком много параметров. Доверительный интервал E, член в уравнении 49 достаточно велик, а общее значение F уравнения 48 меньше, чем у уравнения 46, что указывает на то, что может быть слишком мало свободы в пользу выражения 48. С другой стороны, дополнительный E, член в уравнении 49 подтверждается последовательным F-тестом.

Имея имеющиеся данные, невозможно провести различие между обоими уравнениями на рациональной основе и предпочесть либо уравнения 46 или 49. Уравнения получены из менее хорошо продуманной группы соединений; за исключением водорода и метильной группы, все остальные заместители представляют собой галогены. Валидация всех этих уравнений может быть достигнута только после синтеза и испытаний дополнительных соединений с более крупными мета-заместителями, гидрофильными и / или электронными донорскими заместителями.

В дополнение к двусмысленности результатов есть еще одна серьезная проблема в анализе Ганча. Одно соединение исходной серии, парафенильный аналог, не было включено во все анализы, описанные выше, из-за его плохого соответствия. Эта процедура является единственным выбором, если одно соединение представляет собой в значительной степени различный заместитель, который может быть описан только отдельным параметром. Но чаще всего устранение точек данных является произвольной, субъективной и, следовательно, опасной процедурой. Важный эффект может быть упущен или ложная гипотеза может быть неверно оправдана, исходя из неправильного выбора так называемых «выбросов».

Значения биологической активности других аналогов, имеющих разные заместители в положениях, для которых были получены уравнения, можно предсказать из уравнений Ганча. Хотя это легко с математической точки зрения (значения параметров новых заместителей должны быть вставлены в соответствующее уравнение Ганча), такие предсказания действительно являются довольно сложной задачей. Конечно, предсказанные значения зависят от лежащего в основе уравнения; различные предсказания возникают из разных уравнений, например для больших мета-заместителей из 46 и 49. Если точка, которая должна быть предсказана, находится далеко за пределами включенного диапазона параметров, существует не только риск, но, безусловно, гарантия неудачи. В случае фенэтиламинов может случиться так, что гидроксид-аналоги больше не являются антагонистами, а агонистами.

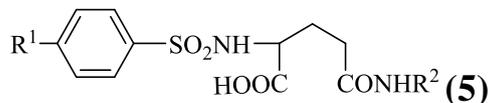
Часто прогностическая способность считается критерием значимости количественных анализов структуры и активности. Хотя это явно разумный спрос, следует понимать, что основная цель анализа Ганча - лучшее понимание, а не прогнозирование. Новые гипотезы могут быть установлены из количественных анализов, которые доказаны или опровергнуты путем синтеза и тестирования новых аналогов. Если предсказанные значения близки к экспериментальным, модель может быть принята.

В противном случае гипотеза была неправильной или ограничивалась определенным диапазоном параметров; должны быть получены новые выводы и новые модели.

Топлисс первым исследовал риск случайных корреляций в анализах Ганча на систематической основе. Несколько симуляций с использованием случайных чисел вместо реальных значений параметров показали, что для данного количества соединений вероятность получения коэффициентов корреляции r , превышающая 0,9, не только значительно возрастает с числом переменных, входящих в уравнение, но и с числом переменных, из которых выбираются различные комбинации [259, 260].

Вероятно, самый ужасный пример чрезмерного предсказания биологических данных, вызванный использованием слишком большого числа переменных, был опубликован в 1986 году (!). Получено несколько сотен уравнений для противоопухолевой активности (типа биологической активности, которая обычно ассоциируется со стандартными отклонениями, намного превышающими 0,3) из восьми производных глутаминовой кислоты ($\mathbf{5}$, $R^1 = \text{H}$, CH_3 , $R^2 = \text{H}$, алкил, циклоалкил) [261]. Только один пугающий пример, уравнение 50 (%ITW = % торможение массы опухоли), цитируется здесь; шесть переменных для восьми соединений оставляют одну степень свободы, достаточно, чтобы не получить сообщение об ошибке компьютерной программой. Кроме того, значения биологической активности определяются ненадлежащим образом, диапазон значений $\log\% \text{ ITW}$ составляет всего 0,24 \log единиц (!), Признаки логарифма P и $(\log P)^2$ являются

неправильными (обратная парабола) , коэффициенты о членов необоснованно велики, не учитываются доверительные интервалы коэффициентов регрессии и дается слишком много десятичных знаков.



$$\log \% ITW = 1.4111(\log P)^2 - 0.5971 \log P - 0.1714\pi_{Al}^2 - 3.2293\sigma_{Al} + 0.9595E_{sAl} - 6.6199\sigma_I + 1.3249 \quad (50)$$

$$(n = 8; r = 0.9912; s = 0.0310)$$

Факторы, важные для значимости и обоснованности отношений QSAR, были рассмотрены в нескольких публикациях [262-265].

1.3.2. Модель аддитивности (анализ FreeWilson)

Подход FreeWilson[16, 20, 266, 267] - это истинная структура отношений структуры и активности. Переменная индикатора генерируется для каждого структурного элемента, который отклоняется от произвольно выбранного ссылочного соединения. Значения 1, указывающие на наличие определенного заместителя или структурного признака, и 0, указывающие на его отсутствие, коррелируют с значениями биологической активности путем линейного множественного регрессионного анализа. Результирующие коэффициенты регрессии переменных индикатора представляют собой вклад биологической активности соответствующих структурных элементов. «Математическая модель», «модель аддитивности» или «подход denovo» - синонимы метода FreeWilson.

Модель FreeWilson была в первоначальной формулировке [16] не так проста. Никакого ссылочного соединения не было выбрано, и были созданы так называемые уравнения симметрии, чтобы избежать проблемы линейных зависимостей между переменными.

Версия, описанная Фуджитой и Баном (уравнение 8) [20, 266, 267], представляет собой прямое применение концепции аддитивности групповых вкладов в значения биологической активности. Поскольку в настоящее время

используется только эта модификация, здесь не рассматриваются детали первоначальной формулировки модели FreeWilson и ее сложных уравнений симметрии.

По сравнению с классической версией анализа Free Wilson вариант Fujita Van предлагает ряд важных преимуществ:

Таблица для регрессионного анализа может быть легко сгенерирована.

Добавление и элиминация соединений является простым и не меняет значения других коэффициентов регрессии.

Любое соединение может быть выбрано в качестве эталонного соединения.

Заместители, которые всегда встречаются вместе в двух разных положениях молекулы, могут быть объединены с псевдопредставителем.

Проблемы сингулярности легко избежать.

Значения групповых вкладов напрямую связаны с групповыми взносами, полученными анализом Ганча.

Индикаторы переменных типа Fujita Van могут быть объединены с анализом Ганча для смешанного подхода [22, 266, 267].

Если метод Free Wilson применяется к α -бромфенэтиламинам (таблица 10) и если в качестве эталонного соединения выбрано незамещенное исходное соединение (**6**, X = Y = H), результаты таблицы 11; регрессионный анализ этих данных приводит к уравнению 51 [266, 267, 268].

$$\begin{aligned} \log 1/C = & -0.301 (\pm 0.50) [m-F] + 0.207 (\pm 0.29) [m-Cl] + \\ & + 0.434 (\pm 0.27) [m-Br] + 0.79 (\pm 0.50) [m-I] + \\ & + 0.454 (\pm 0.27) [m-Me] + 0.340 (\pm 0.30) [p-F] + \\ & + 0.768 (\pm 0.30) [p-Cl] + 1.020 (\pm 0.30) [p-Br] + \\ & + 1.429 (\pm 0.50) [p-I] + 1.256 (\pm 0.33) [p-Me] + \\ & + 7.821 (\pm 0.27) \quad (51) \end{aligned}$$

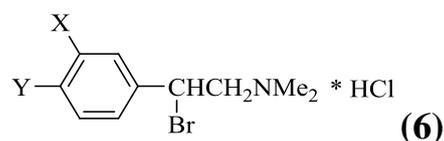
$$(n = 22; r = 0.969; s = 0.194; F = 16.99)$$

Различные коэффициенты регрессии получают, если любое другое соединение выбрано в качестве эталонного соединения или если применяется

классическая модель FreeWilson. Однако эти значения линейно сдвигаются до значений уравнения 51; все статистические параметры идентичны, за исключением только 95% доверительных интервалов [266, 267, 269].

Модель Free Wilson проста в применении. Особенно на ранних этапах анализа структурно-активности это простой способ получения вкладов заместителей и первый взгляд на их возможную зависимость от различных физико-химических свойств.

Таблица 11 - Антиадренергическая активность мета- и пара-замещенных N, N-диметил- α -бромфенэтиламинов (6); таблица для свободного анализа Уилсона [267, 268]



meta (X)	para (Y)	meta-					para-					log1/C obsd.	log1/C calc. ^{a)}
		F	Cl	Br	I	Me	F	Cl	Br	I	Me		
H	H											7.46	7.82
H	F						1					8.16	8.16
H	Cl							1				8.68	8.59
H	Br								1			8.89	8.84
H	I									1		9.25	9.25
H	Me										1	9.30	9.08
F	H	1										7.52	7.52
Cl	H		1									8.16	8.03
Br	H			1								8.30	8.26
I	H				1							8.40	8.40
Me	H					1						8.46	8.28
Cl	F		1				1					8.19	8.37
Br	F			1			1					8.57	8.60
Me	F					1	1					8.82	8.62
Cl	Cl		1					1				8.89	8.80
Br	Cl			1				1				8.92	9.02
Me	Cl							1				8.96	9.04
Cl	Br		1						1			9.00	9.05
Br	Br			1					1			9.35	9.28

Продолжение таблицы

Me	Br							1			9.22	9.30
Me	Me								1		9.30	9.53
Br	Me			1						1	9.52	9.51

Однако модель Free Wilson также имеет некоторые недостатки:

Прежде всего, и самое главное, структурные изменения необходимы, по крайней мере, в двух разных положениях замещения; в противном случае результатом будет бессмысленный групповой вклад, по одному для каждого соединения.

Каждый заместитель, который только один раз встречается в наборе данных, приводит к одноточечному определению; соответствующий групповой вклад содержит всю экспериментальную ошибку этого значения биологической активности.

Только общий вклад активности может быть получен для заместителей, которые всегда встречаются вместе в двух разных положениях молекулы.

В большинстве случаев для описания относительно небольшого количества соединений требуется большое количество параметров, что иногда приводит к статистически статистически значимым (в случае уравнения 51 требуется 10 параметров для описания 22 точек данных, оставляя только 11 степеней свободы) ,

Только небольшое число новых аналогов можно предсказать из анализа FreeWilson (14 других соединений в случае уравнения 51).

Прогнозы для заместителей, которые не включены в анализ, вообще невозможны; оценочные значения могут быть получены из рабочих гипотез о физико-химической интерпретации полученных групповых вкладов.

По его определению анализ FreeWilson ограничивается линейными зависимостями структуры структуры. Подробное обсуждение объема и ограничений модели FreeWilson дано в рефератах [266, 267].

Значения доверительных интервалов имеют разную значимость в анализе Ганча и в модели FreeWilson[266, 267, 269]. Хотя в анализе Ганча они являются мерой значимости соответствующей переменной, в анализе FreeWilson они зависят от выбора ссылочных заместителей; они указывают только, являются ли два групповых вклада (например, групповой вклад ссылочного заместителя, который произвольно установлен равным нулю) и существенно отличаются или нет. Соответственно, «незначительные» термины обычно не устраняются при представлении результатов анализа FreeWilson.

Было предложено использование ступенчатого регрессионного анализа основных компонентов в методе FreeWilson[270]. Другие модификации, которые учитывают и включают только существенные термины [например, 271, 272], как и «уменьшенная модель FreeWilson» [273-275] и метод BELFREE [276]), имеют преимущество большего числа степеней свободы, и поэтому чаще всего получается более высокая статистическая значимость результатов. Однако эти модификации обычно не принимаются. Кроме того, устранение параметров зависит от выбора ссылочных заместителей.

Подход DARC-PELCO [277-281] - простое применение концепции гиперструктуры к методу FreeWilson; в то время как подход может быть подходящим для чрезвычайно больших наборов данных, например, для вывода вкладов липофильности от коэффициентов распределения — это бесполезно для большинства анализов структурной активности из-за слишком большого числа переменных [266, 270]. Результаты Ганча, анализы FreeWilson и DARC-PELCO сравнивались друг с другом [280, 281, 282]; не было видно никаких преимуществ последнего подхода.

1.3.3. Другие подходы QSAR

В течение последних двух десятилетий методы распознавания образов привлекали много внимания [24,25,58,283-287]. В принципе, нет никакой разницы между классическими методами QSAR и распознаванием образов,

только число переменных намного больше в распознавании образов, чем в анализе Ганча [288-292]. В большинстве случаев не проводится предварительный выбор переменных на основе биохимической модели; включается большое количество переменных (физико-химические параметры, квантово-химические показатели, топологические параметры, индикаторные переменные), но только часть данных (обычно 50- 70%, набор тренировок) выбирается для получения количественной модели для прогнозирования остальной части данных [293-297]. Многие проблемы связаны с выбором надлежащего набора тренировок [298-300] и ограничений, возникающих в результате ступенчатого регрессионного анализа (несколько степеней свободы, взаимосвязанных переменных); более последовательные результаты получены с использованием других многомерных методов, таких как анализ основных компонентов или методы мягкого моделирования, например, SIMCA или PLS вместо регрессионного анализа. Обзор методов распознавания образов приведен в двух недавних обзорах [284, 299].

Хотя применение методов распознавания образов, по крайней мере в сочетании с разумным предварительным выбором переменных и использованием метода PLS, представляется оправданным для групп сопутствующих лекарств, которые имеют один и тот же механизм действия, его злоупотребление, чтобы коррелировать и прогнозировать глобальную токсичность, мутагенные, тератогенные, канцерогенные и другие биологические свойства [301-305]. Ни в коем случае не имеет смысла соотносить токсичность лекарств как отличающуюся, например, HCN, дибензодиоксины, стрихнин, тетродотоксин и токсичные пептиды (все они имеют различные механизмы токсического действия) с любой комбинацией физико-химических или других параметров [306-308]. Все полученные результаты должны рассматриваться как случайные корреляции. Общий аргумент о том, что такие исследования полезны или необходимы для целей оценки риска, может быть легко опровергнут: хотя даже сторонники этих

подходов допускают, что существует от 5 до 10% ложноположительных и ложноотрицательных предсказаний (качественным образом, т.е. да или нет), 20-50% неправильных предсказаний, по-видимому, представляют собой более реалистичную оценку для соединений, которые не включены ни в учебный набор, ни в тестовый набор.

То, что эта оценка не слишком пессимистична, подтверждается недавними совместными усилиями нескольких групп для прогнозирования канцерогенности некоторых химических веществ, которые считаются релевантными как опасные для здоровья окружающей среды [309, 310]. В 1990 году Tennant и другие [309] опубликовали предполагаемые прогнозы потенциальной канцерогенности грызунов для 44 химических веществ, которые были протестированы Национальной программой токсикологии США (NTP). Редактор журнала рекомендовал другим группам сделать свои прогнозы, и в результате в 1993 году несколько публикаций и неопубликованных исследований с использованием разных подходов были рассмотрены на семинаре по прогнозированию канцерогенности грызунов [310]. В то время как человеческая экспертиза, сочетающая результаты *in vitro* (тест мутагенеза Сальмонеллы) с другими краткосрочными данными испытаний, химическим и биологическим опытом, привела к корректным предсказаниям (канцерогенная активность да / нет) почти для 80% соединений. Все автоматизированные экспертные системы и другие методы прогнозирования, без использования таких краткосрочных результатов биологических испытаний, являются прогностическими только до 45-65%. При оценке этих цифр следует учитывать, что даже в неуравновешенных группах, то есть независимо от количества активных и неактивных аналогов, включенных в тестовый набор, существует 50% вероятность успеха беспристрастного слепого предположения.

С научной точки зрения такие исследования могут быть оправданы, но для практических целей они бесполезны. Если прогноз отрицательный, отнюдь не является гарантией того, что соединение действительно не

является канцерогенным или токсичным; если «предсказание» дает ложноположительный ответ, тогда будет гораздо труднее опровергнуть отрицательный образ такого соединения, чем без неправильного предсказания.

Несколько более надежным подходом, по-видимому, являются экспертные системы на основе правил [311-315], которые сравнивают и классифицируют структуры новых соединений в отношении информации, которая извлекается из большой базы данных, например, программа CASE (компьютерно-автоматизированная структура) [316-320]. CASE и MULTICASE (модификация CASE, включая иерархический выбор дескрипторов) - это программы искусственного интеллекта, которые автоматически определяют молекулярные особенности, которые способствуют (биофоры) или уменьшают (биофобы) биологическую активность (рисунок 8) [321]. В congeneric серии эти фрагменты коррелированы (что соответствует поэтапному анализу Вилсона) с биологической активностью; в неконденсированном ряду идентифицируются структурные особенности, которые различают активные соединения из неактивных.

Утверждается, что CASE отличается от других методов тем, что полностью автоматизирован и непосредственно изучается из сырых данных, выбирая собственные дескрипторы из практически бесконечного числа возможных структурных сборок и создавая специальный словарь без вмешательства человека [321]. Хотя это утверждение противоречит всем предшествующим опытам с автоматизированными подходами, даже с критической точки зрения нельзя исключать, что подход CASE впервые может быть приближением искусственного интеллекта к интуиции и мастерству фармацевтического химика.

Была разработана и применена программа машинного обучения GOLEM из области индуктивного логического программирования для моделирования отношений структурно-активности ингибиторов

дигидрофолатредуктазы (DHFR) [322]. GOLEM использует информацию о активности, структуре и стереохимии активных и неактивных аналогов для получения индуктивных гипотез, связывающих структуры произвольно выбранных пар аналогов с их действиями. Правила уточняются со всеми возможными парами соединений, и правила, имеющие высокую степень достоверности, добавляются в базу знаний. Процесс генерации правил повторяется до тех пор, пока не будет достигнуто никакого улучшения в прогнозировании. 44 ингибитора DHFR использовались в качестве набора для автоматического определения правил, которые применялись для прогнозирования активности 11 тестируемых соединений. Был сделан вывод о том, что правила, полученные GOLEM, дают более точные данные, чем модель регрессии Ганча, на обучающем наборе, а также на тестовом наборе [322].

Многие различные, но по своей сути связанные методы QSAR начинаются с гиперструктуры, которая представляет собой гипотетическую молекулу, включающую все структурные особенности исследуемых молекул. Наличие и отсутствие определенных атомов или групп гиперструктуры в отдельных молекулах коррелируют с их биологической активностью в свободной форме, подобной Вильсону. Метод DARC-PELCO был использован для анализа и оптимизации большой группы антихолинергических препаратов [323] путем получения общей надстройки.

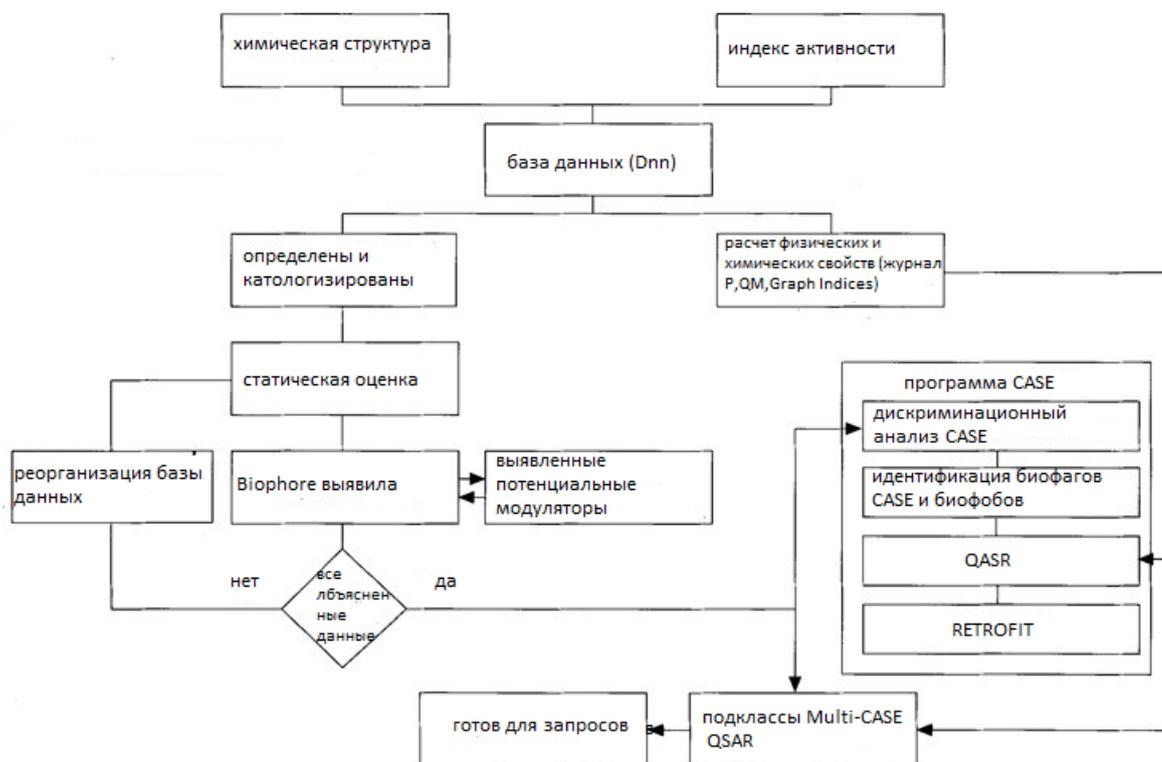


Рисунок 8– Многозадачная программа [321].

В методах топологического фармакофора например, LOCON [324, 325, 327], LOGANA [325 – 327] и EVAL [328]), индикаторы типа Free Wilson находятся в ступенчатой процедуре, связанной логическими операторами, например, отрицание (не), соединение (и), присоединение (включительно или), дизъюнкция (исключение или), импликация (если это так), репликация (так если), эквивалентность (тогда и только тогда), функция Шеффера (несовместимая с) и Функция Никода (ни и не) [328] для корреляции структур и видов деятельности. Методы LOGANA и LOCON применяли к агонистам фенетиламина и антагонистам P-адренергического рецептора [324, 328] и группе из 382 противомаларийных препаратов [326]. Ингибиторы сукцинатдегидрогеназы $R_x - X - Z - Y - R_y$ ($n = 89$, $X = n$ - пропил, фенил, гетероциклы, R_x - заместители у X - фрагмента, $Z = -CONH-COCH_2-$, $Y = H$, фенил, циклогексил, $R_y =$ заместители у Y-фрагмента) были разделены различными логическими операторами на классы активных и

неактивных аналогов. Конъюнкция «-C = C-C = O в X и отсутствие замещенного кольца X» присутствует в 63 из 64 активных соединений, но также в 12 неактивных соединениях. Другое соединение "- C = C - C =O в X и не имеет орто-незамещенного кольца X и никакого кольца в Y и Z \diamond -CONH- и X \diamond -CH = CH-CH₃, а R_x \diamond ортофенил" является присутствующих в 61 активных соединениях и в неактивном соединении [325].

Также минимальная стерическая разность (MSD) [52, 329-331], MCD (версия MSD Монте-Карло) [52, 332-334], минимальная топологическая разность (MTD) [52, 331, 335-339] и Методы SIBIS [340, 38, 52, 198] определяют гипотетические гипермолекулы. Эти гипермолекулы генерируются из приближенной атомарно-атомной суперпозиции всех молекул; атомов водорода и незначительных различий в положениях атомов, длинах связей и углах связи пренебрегают. Определение гипермолекулы является относительно простой и простой процедурой, если в серию входит жесткий и высокоактивный аналог. Во многих других случаях применяется интуитивный подход, и принимаются произвольные предположения о связности гипермолекулы; проблемы, связанные с суперпозицией различных кольцевых систем, обсуждаются в [339].

Что касается сайта связывания, некоторые атомные положения гипермолекулы будут находиться в полости рецептора ($\epsilon = -1$), некоторые на поверхности сайта связывания («полостная стенка») ($\epsilon = +1$), а остаток в водном растворе ($\epsilon = 0$). Минимальная стерическая разность (MSD) определяется как разница в количестве атомных позиций, которые заняты каждым аналогом, относительно аналога наивысшего значения биологической активности внутри серии, который считается оптимальным для привязки сайт. Для минимальной топологической разности (МПД) используется немного другое определение. MTD - это сумма всех значений E, умноженная на их появление в определенной молекуле. Поскольку значения E в разных положениях являются априорными неизвестными, произвольные значения +1, 0 и -1 (стартовая карта) присваиваются

различным атомам (вершинам) гипермолекулы и коррелируются с значениями биологической активности. На следующем шаге все значения E систематически изменяются для достижения максимальной корреляции между оптимизированными картами и биологической активностью. Это неоднозначная процедура, приводящая к огромному количеству различных перестановок и нескольким оптимизированным рецепторным картам; проверочный тест, например, перекрестная валидация, должна быть выполнена для доказательства значимости и прогностической способности различных карт.

Маги [341, 342] объединил концепцию гиперструктуры со стратегиями анализа Ганча и смешанного подхода. Поскольку только несколько атомов или групп молекулы модулируют биологическую активность, каждое положение гипермолекулы можно охарактеризовать как благоприятное, неблагоприятное или равнодушное к биологической активности. Позиционные эффекты для липофильных взаимодействий, поляризуемости, электроотрицательности, стерических взаимодействий и водородных связей могут быть отнесены к определенным позициям.

Хотя все эти методы рассматривают гипермолекулы как жесткие рамки, анализ молекулярной формы (MSA) [38, 198, 343-347] рассматривает конформационную гибкость. Минимальные конформации энергии рассчитываются для всех молекул внутри серии [348-352]; различия в объемах перекрытия с эталонной структурой вычисляются и коррелируются с биологической активностью. Также в этом подходе выгодно начинать с жесткого аналога высокой биологической активности в качестве шаблона [353].

Применение нейронных сетей в исследованиях отношений структуры и активности - это недавнее развитие в QSAR [354-358]. Нейронные сети - это подходы к моделированию и моделированию обработки информации в мозге. Многослойная сеть, как правило, применяется в исследованиях QSAR, состоит из входного слоя, небольшого количества скрытых уровней

соединения, нейронных систем смещения и уровня вывода. Каждый нейрон связан с каждым нейроном соседних слоев и с нейронами смещения. На этапе обучения сеть питается входными значениями (независимыми переменными) и ожидаемыми выходными значениями (зависимая переменная) [359-363]. Весы подключения и смещения сети итеративно модифицируются для уменьшения различий между наблюдаемыми и прогнозируемыми значениями зависимой переменной. Выполняя достаточное количество итераций, сеть «учится» распознавать шаблоны в данных. После успешного обучения сеть может использоваться для прогнозирования значений биологической активности новых аналогов из новых входных значений.

Слишком рано судить о реальной пригодности нейронных сетей для исследований QSAR; дальнейшие исследования, которые сравнивают классический анализ структурной активности и результаты нейронных сетей [540], необходимы для оценки объема и ограничений нейронных сетей. Некоторые проблемы нейронных сетей, например, дизайн сети, отсутствие конвергенции, случайные корреляции и перетренированность сети обсуждались и критически комментировались [359, 364-367].

2. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В данном разделе приведены результаты построения количественных моделей соотношения структура-свойство. Объектами исследования стали синтезированные в нашей лаборатории производные 5-цианопиримидина, проявляющие высокую цитотоксичность к клеткам эпидермоидной карциномы A431, обильно экспрессирующим EGFR. На рисунке 9 приведена общая структура исследуемых соединений и их модель выравнивания.

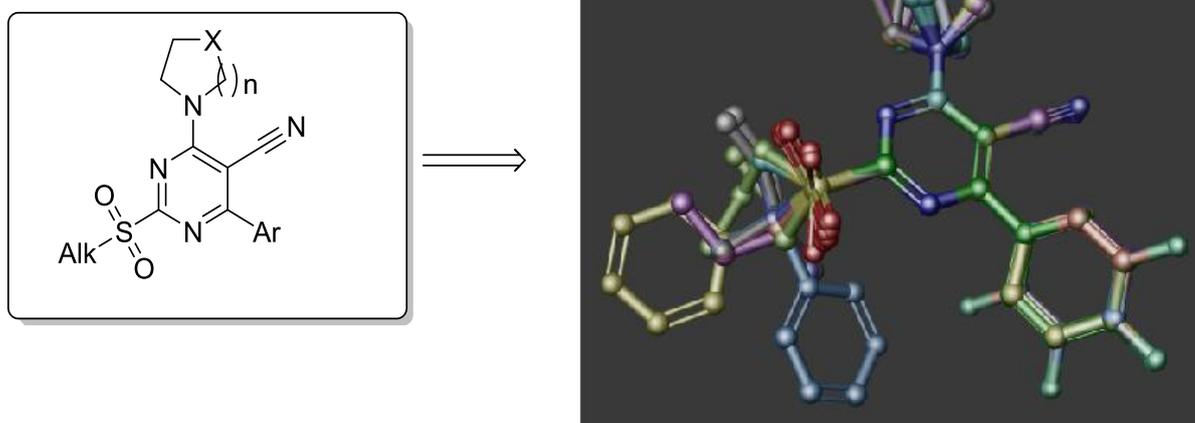
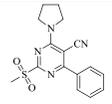
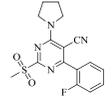
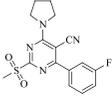
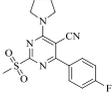
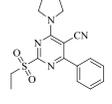
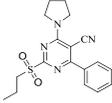
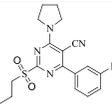
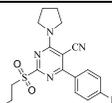
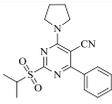
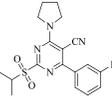


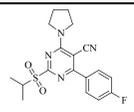
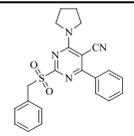
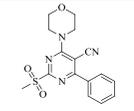
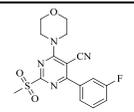
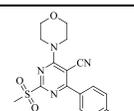
Рисунок 9

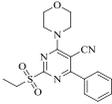
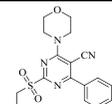
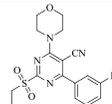
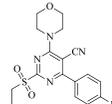
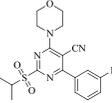
В таблице 12 представлены физико-химические характеристики исследуемых соединений и значения биологической активности (цитотоксичности). Анализ физико-химических характеристик, таких как липофильность, молекулярная масса и объем молекулы позволяет заключить, что все соединения имеют хороший drug-like профиль и в полной мере соответствуют правилам Липински.

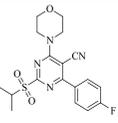
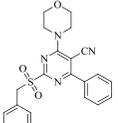
Таблица 12

№	Соединения	EC ₅₀ (A431) μM	pEC ₅₀	MW	SASA	FOSA	FISA	PISA	WPSA	volume	donorHB	accptHB	logPo/w	PSA
1	 SV-260	12.15	4.915	328.39	605.698	255.761	147.586	201.696	0.655	1050.537	0	8	1.470	87.988
2	 TK-387	3.92	5.407	346.38	605.468	262.052	137.413	176.270	29.733	1058.453	0	8	1.673	86.438
3	 TK-388	3.36	5.474	346.38	611.364	263.826	137.044	163.650	46.844	1063.123	0	8	1.757	87.133
4	 TK-403	4.72	5.326	346.78	612.206	265.594	136.326	163.424	46.862	1063.498	0	8	1.764	86.794
5		5.09	5.293	342.42	611.630	285.852	129.357	196.421	0.000	1084.249	0	8	1.807	83.919

	SV-340													
6	 SV-332	6.12	5.213	358.44	634.387	313..257	121.559	199.571	0.000	1134.118	0	8	2.190	80.481
7	 SV-334	6.08	5.216	374.43	633.902	302.340	125.475	159.366	46.720	1146.398	0	8	2.375	82.071
8	 SV-335	4.67	5.333	374.43	643.486	313.151	121.552	161.746	46.837	1150.035	0	8	3.429	80.461
9	 TK-406	3.75	5.426	356.44	647.576	327.524	119.039	201.013	0.000	1149.678	0	8	2.311	82.874
10	 TK-405	3.34	5.476	374.43	656.097	328.108	118.382	162.743	46.863	1165.211	0	8	2.551	80.281

11	 SV-350	1.67	5.777	374.43	657.986	328.250	119.734	163.166	46.836	1166.048	0	8	2.547	82.890
12	 TK-411	2.55	5.593	422.48	721.140	203.945	119.668	350.692	46.834	1288.758	0	8	3.568	84.785
13	 SV-261	19.03	4.721	344.39	578.674	251.900	128.728	198.050	0.031	1035.122	0	8	0.665	94.154
14	 TK-391	6.13	5.213	362.38	599.110	252.777	137.833	161.771	46.728	1061.830	0	8	0.912	95.750
15	 TK-404	24.95	4.603	362.38	588.234	252.907	128.848	159.618	46.861	1050.033	0	8	0.896	94.244

16	 SV-333	18.31	4.737	358.41	609.926	277.292	135.169	197.465	0.000	1093.243	0	8	0.999	94.600
17	 SV-338	13.4	4.873	372.44	633.151	322.809	110.285	200.057	0.000	1141.123	0	8	1.487	90.919
18	 SV-336	11.03	4.957	390.43	616.269	285.082	122.570	160.742	47.876	1132.475	0	8	1.485	89.274
19	 SV-337	12.09	4.918	390.43	623.795	295.967	119.463	161.527	46.837	1136.234	0	8	1.525	88.427
20	 SV-348	6.92	5.160	390.43	623.890	309.075	123.412	152.685	47.719	1149.857	0	8	1.580	91.923

21	 SV-347	7.33	5.135	390.43	629.919	297.051	130.298	154.691	47.878	1143.632	0	8	1.494	91.234
22	 SV-341	21.08	4.676	420.48	721.015	209.884	129.841	380.146	1.144	1300.868	0	8	2.609	92.264

QSAR анализ был проведен с применением оригинального алгоритма AutoQSAR, который с учетом более чем 300 дескрипторов от молекулярных до топологических с помощью линейного регрессионного анализа строит соответствующую зависимость. В таблице 13 представлены лучшие 10 полученных моделей.

Таблица 13

Код Модели	Score	S.D.	R ²	RMSE	Q ²	Q ² MW (Нулевая Гипотеза)
kpls_linear_29	0.8775	0.1364	0.8737	0.0922	0.7852	-0.4379
kpls_radial_24	0.8670	0.1329	0.8626	0.0758	0.9352	-0.5223
kpls_linear_13	0.8654	0.1375	0.8616	0.0634	0.9330	-0.7325
kpls_molprint2D_45	0.8489	0.1495	0.8439	0.0702	0.9123	-0.4628
kpls_linear_24	0.8461	0.1435	0.8397	0.0758	0.8968	-0.5223
kpls_molprint2D_24	0.8440	0.1344	0.8594	0.1275	0.8168	-0.5223
kpls_molprint2D_13	0.8432	0.1495	0.8366	0.0926	0.8571	-0.7325
kpls_linear_34	0.8405	0.1518	0.8337	0.0908	0.8974	-0.1168
kpls_linear_45	0.8339	0.1562	0.8295	0.0591	0.9378	-0.4628
kpls_linear_39	0.8311	0.1546	0.8368	0.1347	0.6167	-0.3369

Анализ представленных данных позволяет заключить, что наилучшее значения коэффициента детерминации и оценочной (скоринг) функции удастся получить с применением линейных корреляционных алгоритмов. В то время, как модели, построенные на топологических дескрипторах исследуемых соединений практически не представлены в топ-10.

В таблице 14 представлены расширенные характеристики полученных моделей. В целом полученные модели можно считать рабочими. Значение среднеквадратичного отклонения во всех моделях не превышает значения 0.15, а значение коэффициента RMSE не превышает 0.1 и во многих случаях меньше представленного значения. Особенно стоит отметить

высокие значения коэффициента корреляции перекрёстного контроля Q^2 , который для большинства моделей составил более 0.7.

Таблица 14

Модель kpls_linear_29				Модель kpls_linear_24				Модель kpls_linear_13				Модель kpls_linear_45			
Ranking score = 0.877464				Ranking score = 0.846138				Ranking score = 0.865381				Ranking score = 0.833899			
Тренировочная выборка		Тестовая выборка		Тренировочная выборка		Тестовая выборка		Тренировочная выборка		Тестовая выборка		Тренировочная выборка		Тестовая выборка	
S.D.	R ²	RMSE	Q ²	S.D.	R ²	RMSE	Q ²	S.D.	R ²	RMSE	Q ²	S.D.	R ²	RMSE	Q ²
0.1364	0.8737	0.0922	0.7852	0.1435	0.8397	0.0957	0.8968	0.1435	0.8397	0.0957	0.8968	0.1562	0.8295	0.0591	0.9378
Оптимальное количество факторов = 1				Оптимальное количество факторов = 1				Оптимальное количество факторов = 1				Оптимальное количество факторов = 1			
ID Set	Y(Obs)	Y(Pred)	Error	ID Set	Y(Obs)	Y(Pred)	Error	ID Set	Y(Obs)	Y(Pred)	Error	ID Set	Y(Obs)	Y(Pred)	Error
1 train	5.4737	5.4118	-0.0618	1 train	5.4737	5.3283	-0.1453	1 train	5.4737	5.3283	-0.1453	1 train	5.4737	5.3577	-0.1160
2 train	5.2933	5.1915	-0.1018	2 train	5.2933	5.2907	-0.0026	2 train	5.2933	5.2907	-0.0026	2 test	5.2933	5.2095	-0.0838
3 train	5.2132	5.2223	0.0091	3 train	5.2132	5.2764	0.0631	3 train	25.132	5.2764	0.0631	3 train	5.2132	5.1680	-0.0452
4 train	5.2161	5.3787	0.1626	4 train	5.2161	5.3129	0.0968	4 train	5.2161	5.3129	0.0968	4 train	5.2161	5.3093	0.0932
5 train	5.3307	5.413	0.0828	5 train	5.3307	5.5241	0.1934	5 train	5.3307	5.5241	0.1934	5 train	5.3307	5.4766	0.1459
6 train	5.4260	5.5120	0.0861	6 train	5.4260	5.4194	-0.0066	6 test	5.4260	5.5241	-0.0066	6 train	5.4260	5.4758	0.0498
7 train	5.7773	5.6577	-0.1196	7 train	5.7773	5.6471	-0.1302	7 train	5.7773	5.6471	-0.1302	7 train	5.7773	5.6084	-0.1689
8 train	5.5935	5.5082	-0.0852	8 train	5.5935	5.5663	-0.0272	8 train	5.5935	5.5663	-0.0272	8 train	5.5935	5.5149	-0.0785
9 train	4.6029	4.8943	0.2914	9 train	4.6029	4.8435	0.2405	9 train	4.6029	4.8435	0.2405	9 train	4.6029	4.9326	0.3296
10 train	4.7373	4.5860	-0.1513	10 test	4.7373	4.8876	0.1503	10 test	4.7373	4.8876	0.1503	10 test	4.7373	4.7949	0.0576
11 test	4.8729	4.7480	-0.1249	11 test	4.8729	4.8034	-0.0695	11 test	4.8729	4.8034	-0.0695	11 train	4.8729	4.6785	-0.1944
12 test	4.9574	4.9023	-0.0551	12 train	4.9574	4.7698	-0.1877	12 train	4.9574	4.7698	-0.1877	12 train	4.9574	4.8639	-0.0935
13 train	5.1599	5.1179	-0.0420	13 train	5.1599	5.0113	-0.1485	13 train	5.1599	5.0113	-0.1485	13 test	5.1599	5.1475	-0.0124
14 train	5.1349	5.1994	0.0645	14 train	5.1349	5.1553	0.0204	14 train	5.1349	5.1553	0.0204	14 train	5.1349	5.2139	0.0790
15 train	4.6761	4.6242	-0.0520	15 train	4.6761	4.7033	0.0272	15 train	4.6761	4.7033	0.0272	15 train	4.6761	4.6751	-0.0010

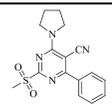
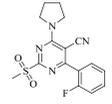
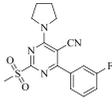
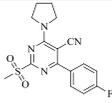
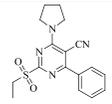
Модель kpls_molprint2D_45				Модель kpls_radial_24				Модель kpls_linear_34				Модель kpls_molprint2D_13			
Ranking score = 0.848916				Ranking score = 0.867026				Ranking score = 0.840516				Ranking score = 0.843220			
Тренировочная выборка		Тестовая выборка		Тренировочная выборка		Тестовая выборка		Тренировочная выборка		Тестовая выборка		Тренировочная выборка		Тестовая выборка	
S.D.	R^2	RMSE	Q^2												
0.1495	0.8439	0.0702	0.9123	0.1329	0.6826	0.0758	0.9352	0.1518	0.8337	0.0908	0.8974	0.1495	0.8366	0.0926	0.8571
Оптимальное количество факторов = 1				Оптимальное количество факторов = 1				Оптимальное количество факторов = 1				Оптимальное количество факторов = 1			
ID Set	Y(Obs)	Y(Pred)	Error	ID Set	Y(Obs)	Y(Pred)	Error	ID Set	Y(Obs)	Y(Pred)	Error	ID Set	Y(Obs)	Y(Pred)	Error
1 train	5.4737	5.3150	-0.1587	1 train	5.4737	5.3183	-0.1553	1 train	5.4737	5.3631	-0.1105	1 train	5.4737	5.3864	-0.0872
2 test	5.2933	5.2607	-0.0325	2 train	5.2933	5.2309	-0.0623	2 train	5.2933	5.2710	-0.0223	2 train	5.2933	5.3177	0.0244
3 train	5.2132	5.1608	-0.0525	3 train	5.2132	5.2162	0.0029	3 train	5.2132	5.2089	-0.0043	3 train	5.2132	5.1105	-0.1028
4 train	5.2161	5.3603	0.1442	4 train	5.2161	5.2079	-0.0082	4 train	5.2161	5.3203	0.1042	4 train	5.2161	5.3647	0.1486
5 train	5.3307	5.3615	0.0308	5 train	5.3307	5.4847	0.1541	5 train	5.3307	5.5160	0.1854	5 test	5.3307	5.2366	-0.0941
6 train	5.4260	5.5374	0.1114	6 test	5.4260	5.4517	0.0257	6 test	5.4260	5.4304	0.0044	6 train	5.4260	5.5330	0.1071
7 train	5.7773	5.7377	-0.0396	7 train	5.7773	5.7283	-0.0490	7 train	5.7773	5.6077	-0.1695	7 train	5.7773	5.6587	-0.1186
8 train	5.5935	5.3811	-0.2124	8 train	5.5935	5.5392	-0.0543	8 train	5.5935	5.5580	-0.0355	8 train	5.5935	5.3955	-0.1980
9 train	4.6029	4.8390	0.2361	9 train	4.6029	4.8582	0.2553	9 train	4.6029	4.9130	0.3101	9 train	4.6029	4.7785	0.1755
10 train	4.7373	4.7882	0.0509	10 test	4.7373	4.8647	0.1274	10 test	4.7373	4.8465	0.1091	10 test	4.7373	4.8515	0.1142
11 train	4.8729	4.6896	-0.1833	11 test	4.8729	4.8540	-0.0189	11 train	4.8729	4.6956	-0.1773	11 train	4.8729	4.6395	-0.2334
12 train	4.9574	4.8904	-0.0670	12 train	4.9574	4.7849	-0.1725	12 train	4.9574	4.8391	-0.1183	12 test	4.9574	4.8956	-0.0618
13 test	5.1599	5.2654	0.1055	13 train	5.1599	5.1392	-0.0207	13 test	5.1599	5.0467	-0.1132	13 train	5.1599	5.3128	0.1529
14 train	5.1349	5.2663	0.1314	14 train	5.1349	5.2847	0.1498	14 train	5.1349	5.1489	0.0140	14 train	5.1349	5.1845	0.0496
15 train	4.6761	4.7355	0.0594	15 train	4.6761	4.6363	-0.0398	15 train	4.6761	4.7003	0.0242	15 train	4.6761	4.7580	0.0818

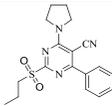
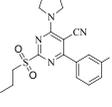
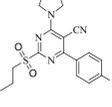
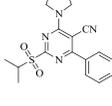
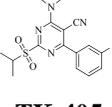
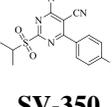
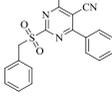
Модельkpls_molprint2D_24				Модельkpls_linear_39			
Ranking score = 0.844009				Ranking score = 0.831142			
Тренировочная выборка		Тестовая выборка		Тренировочная выборка		Тестовая выборка	
S.D.	R ²	S.D.	R ²	RMSE	Q ²	RMSE	Q ²
0.1344	0.8594	0.1495	0.8366	0.0926	0.8571	0.1275	0.8168
Optimum number of factors = 1				Optimum number of factors = 1			
ID Set	Y(Obs)	ID Set	Y(Obs)	Y(Pred)	Error	Y(Pred)	Error
1 train	5.4737	1 train	5.4737	5.3864	-0.0872	5.2635	-0.2102
2 train	5.2933	2 train	5.2933	5.3177	0.0244	5.3756	0.0823
3 train	5.2132	3 train	5.2132	5.1105	-0.1028	5.2329	0.0196
4 train	5.2161	4 train	5.2161	5.3647	0.1486	5.4324	0.2163
5 train	5.3307	5 test	5.3307	5.2366	-0.0941	5.4522	0.1215
6 test	5.4260	6 train	5.4260	5.5330	0.1071	5.4944	0.0684
7 train	5.7773	7 train	5.7773	5.6587	-0.1186	5.7171	-0.0601
8 train	5.5935	8 train	5.5935	5.3955	-0.1980	5.3902	-0.2032
9 train	4.6029	9 train	4.6029	4.7785	0.1755	4.7250	0.1221
10 test	4.7373	10 test	4.7373	4.8515	0.1142	4.8369	0.0996
11 test	4.8729	11 train	4.8729	4.6395	-0.2334	4.6880	-0.1849
12 train	4.9574	12 test	4.9574	4.8956	-0.0618	4.8874	-0.0701
13 train	5.1599	13 train	5.1599	5.3128	0.1529	5.1502	-0.0097
14 train	5.1349	14 train	5.1349	5.1845	0.0496	5.1683	0.0334
15 train	4.6761	15 train	4.6761	4.7580	0.0818	4.6342	-0.0419

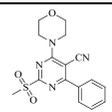
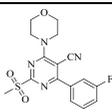
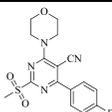
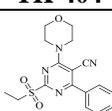
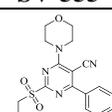
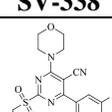
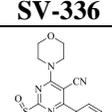
Для выявления основных особенностей зависимости влияния трехмерной структуры исследуемых соединений на их биологическую активность был осуществлён 3D-QSARc применением метода сравнительного анализа молекулярных полей. Анализ проводили с применением двух тренировочных выборок, состоящих из 50 и 75% от исследуемых лигандов.

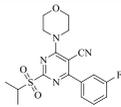
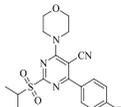
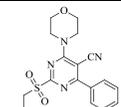
Основные количественные характеристики полученных моделей представлены в таблицах 15-16. Стоит отметить, что при тренировочной выборке в 75% достигаются наилучшие значения коэффициента корреляции перекрёстного контроля Q^2 , что в свою очередь делает модели, полученные на ее основе более предпочтительными для дальнейшего обсуждения.

Таблица 15 - 3DQSAR тренировочная выборка 50%

№	Соединения	EC ₅₀ (A431) μM	pEC ₅₀ (A431)	QSAR выборка	Активность	# Фактор	Предсказанная активность	Ошибка предсказания	% Экстраполирования
1	 SV-260	12.15	4.915	тренировочная	4.915	1	5.09	0.176	2.55
						2	5.074	0.158	
2	 TK-387	3.92	5.407	тренировочная	5.407	1	5.286	-0.121	0.41
						2	5.391	-0.016	
3	 TK-388	3.36	5.474	тренировочная	5.474	1	5.281	-0.193	0.18
						2	5.387	-0.086	
4	 TK-403	4.72	5.326	тренировочная	5.326	1	5.279	-0.047	0.00
						2	5.370	0.044	
5		5.09	5.293	тестовая	5.293	1	5.285	-0.008	12.37
						2	5.331	0.038	

	SV-340								
6	 SV-332	6.12	5.213	тренировочная	5.213	1	5.262	0.048	0.81
						2	5.337	0.124	
7	 SV-334	6.08	5.216	тестовая	5.216	1	5.336	0.119	15.01
						2	5.419	0.203	
8	 SV-335	4.67	5.333	тестовая	5.331	1	5.275	-0.056	2.96
						2	5.357	0.026	
9	 TK-406	3.75	5.426	тестовая	5.426	1	5.341	-0.085	3.35
						2	5.367	-0.059	
10	 TK-405	3.34	5.476	тестовая	5.476	1	5.351	-0.125	4.58
						2	5.387	-0.090	
11	 SV-350	1.67	5.777	тестовая	5.777	1	5.356	-0.421	4.34
						2	5.389	-0.388	
12	 TK-411	2.55	5.593	тренировочная	5.593	1	5.707	0.114	0.87
						2	5.542	-0.052	

13	 SV-261	19.03	4.721	тренировочная	4.721	1	4.761	0.040	0.04
						2	4.719	-0.001	
14	 TK-391	6.13	5.213	тренировочная	5.213	1	5.049	-0.164	0.73
						2	5.102	-0.111	
15	 TK-404	24.95	4.603	тренировочная	4.603	1	4.775	0.173	0.04
						2	4.742	0.139	
16	 SV-333	18.31	4.737	тестовая	4.737	1	5.071	0.333	20.48
						2	5.081	0.344	
17	 SV-338	13.4	4.873	тренировочная	4.873	1	5.034	0.162	0.02
						2	4.859	-0.013	
18	 SV-336	11.03	4.957	тренировочная	4.957	1	4.769	-0.188	0.48
						2	4.770	-0.188	
19	 SV-337	12.09	4.918	тестовая	4.918	1	4.957	0.039	16.37
						2	4.772	-0.145	

20	 SV-348	6.92	5.160	тестовая	5.160	1	5.036	-0.124	12.16
						2	5.099	-0.061	
21	 SV-347	7.33	5.135	тестовая	5.135	1	5.096	-0.039	18.07
						2	5.101	-0.034	
22	 SV-341	21.08	4.676	тестовая	4.676	1	5.147	0.471	22.38
						2	5.138	0.462	

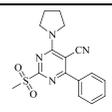
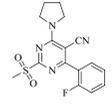
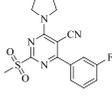
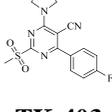
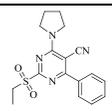
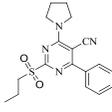
3DQSARstatistics: тренировочная выборка 50%

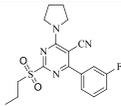
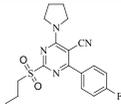
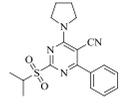
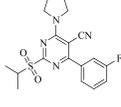
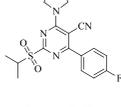
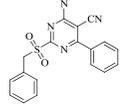
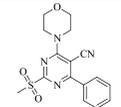
# Факторы	SD	R ²	R ² CV	R ² Scramble	Stability	F	P	RMSE	Q ²	Pearson-r
1	0.1563	0.7902	0.3699	0.3973	0.538	33.9	0.000252	0.23	0.4638	0.7413
2	0.1222	0.8859	0.5103	0.5551	0.788	31.1	0.00017	0.23	0.4690	0.6947

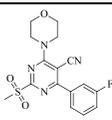
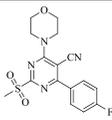
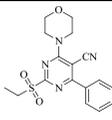
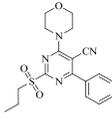
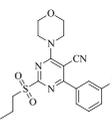
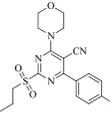
Fied fractions:

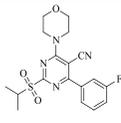
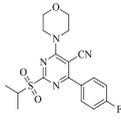
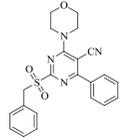
# Factors	Gaussian Steric	Gaussian Electrostatic	Gaussian Hydrophobic	Gaussian Hbond Acceptor
1	0.505314	0.056349	0.211673	0.226664
2	0.443765	0.072140	0.230801	0.253295

Таблица 16 - 3DQSAR тренировочная выборка 75%

№	Соединения	EC ₅₀ (A431) μM	pEC ₅₀	QSAR выборка	Активность	# Фактор	Предсказанная активность	Ошибка предсказания	% Экстраполирования
1	 SV-260	12.15	4.915	тренировочная	4.915	1	5.178	0.263	0.96
						2	4.983	0.068	
2	 TK-387	3.92	5.407	тренировочная	5.407	1	5.332	-0.075	0.02
						2	5.342	-0.065	
3	 TK-388	3.36	5.474	тренировочная	5.474	1	5.347	-0.127	0.07
						2	5.353	-0.121	
4	 TK-403	4.72	5.326	тренировочная	5.326	1	5.357	0.031	0.02
						2	5.366	0.041	
5	 SV-340	5.09	5.293	тестовая	5.293	1	5.267	-0.026	2.22
						2	5.267	-0.026	
6		6.12	5.213	тестовая	5.213	1	5.421	0.207	5.91
						2	5.378	0.165	

	SV-332								
7	 SV-334	6.08	5.216	тренировочная	5.216	1	5.369	0.154	1.43
						2	5.401	0.185	
8	 SV-335	4.67	5.333	тренировочная	5.331	1	5.449	0.119	0.07
						2	5.434	0.103	
9	 TK-406	3.75	5.426	тренировочная	5.426	1	5.444	0.018	0.45
						2	5.473	0.048	
10	 TK-405	3.34	5.476	тестовая	5.476	1	5.465	-0.011	3.45
						2	5.511	0.034	
11	 SV-350	1.67	5.777	тренировочная	5.777	1	5.472	-0.305	0.07
						2	5.531	-0.246	
12	 TK-411	2.55	5.593	тренировочная	5.593	1	5.582	-0.011	0.87
						2	5.689	0.096	
13	 SV-261	19.03	4.721	тренировочная	4.721	1	4.977	0.257	0.02
						2	4.803	0.083	

14	 TK-391	6.13	5.213	тренировочная	5.213	1	5.003	-0.209	0.27
						2	4.998	-0.214	
15	 TK-404	24.95	4.603	тестовая	4.603	1	5.013	0.409	8.45
						2	4.865	0.262	
16	 SV-333	18.31	4.737	тестовая	4.737	1	4.912	0.175	0.07
						2	4.908	0.171	
17	 SV-338	13.4	4.873	тренировочная	4.873	1	4.948	0.075	0.02
						2	4.889	0.016	
18	 SV-336	11.03	4.957	тестовая	4.957	1	5.108	0.151	17.23
						2	4.937	-0.021	
19	 SV-337	12.09	4.918	тренировочная	4.918	1	4.944	0.027	1.38
						2	4.802	-0.115	

20	 SV-348	6.92	5.160	тренировочная	5.160	1	5.115	-0.045	0.56
						2	5.174	0.014	
21	 SV-347	7.33	5.135	тренировочная	5.135	1	4.913	-0.221	0.19
						2	4.978	-0.157	
22	 SV-341	21.08	4.676	тренировочная	4.676	1	4.551	-0.125	0.24
						2	4.769	0.094	

QSAR statistics: тренировочная выборка 75%

# Factors	SD	R ²	R ² CV	R ² Scramble	Stability	F	P	RMSE	Q ²	Pearson-r
1	0.1707	0.7379	0.4527	0.4730	0.749	42.2	1e-005	0.22	0.4882	0.9150
2	0.1393	0.8372	0.4822	0.6279	0.77	36.0	3.04e-006	0.14	0.7858	0.9342

Field fractions:

# Factors	Gaussian Steric	Gaussian Electrostatic	Gaussian Hydrophobic	Gaussian Hbond Acceptor	Gaussian Hbond Donor
1	0.500982	0.070046	0.215911	0.213061	0.000000
2	0.475753	0.086796	0.217545	0.219906	0.000000

На рисунке 10 представлена визуализация влияния стерических факторов на биологическую активность исследуемых соединений. В представленной визуализации выделяются три области стерического влияния. В первую очередь это влияние алкильного заместителя со стороны сульфонильной группы (например соединения SV348, SV350, ТК411) и арильного заместителя с заместителями в орто- и мета-положениях.

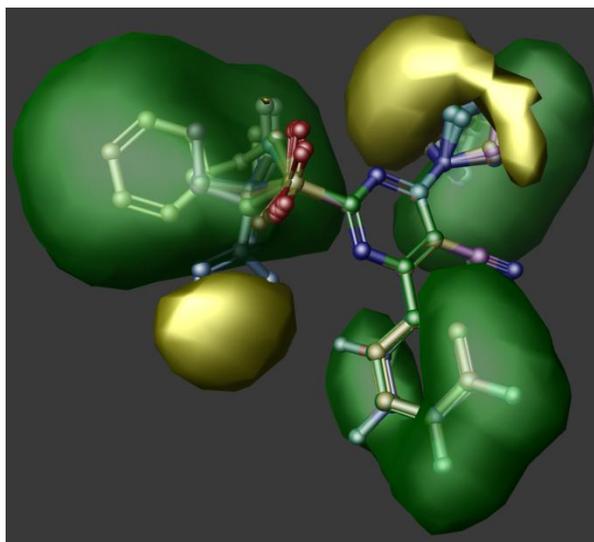


Рисунок 10 – Влияние стерических факторов на биологическую активность.

Зеленый контур – положительный вклад, желтый контур – отрицательный вклад.

На рисунке 11 приведена карта электростатических контуров. Интерпретация стерических факторов как правило проста, тогда как интерпретация электростатических карт более сложна из-за требования электронейтральности и того, что к увеличению биологической активности могут приводить как положительные, так и отрицательные заряды.

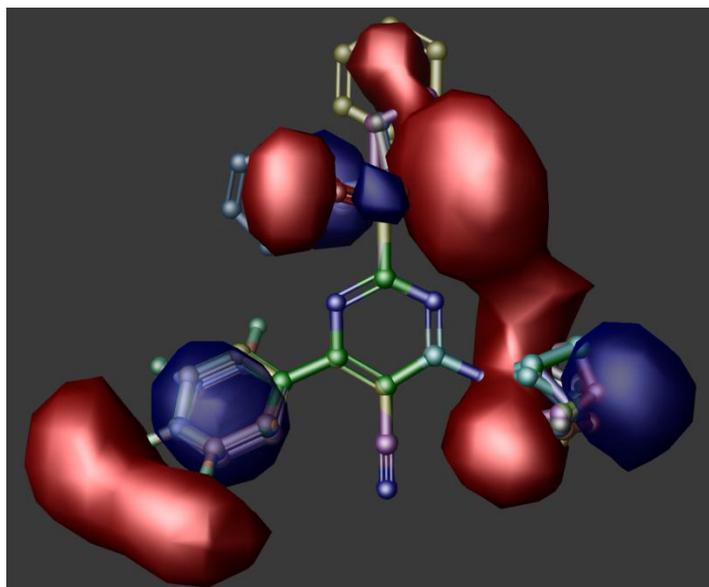


Рисунок 11 – Влияние электростатических факторов на биологическую активность. Синий контур – положительный вклад, красный контур – отрицательный вклад.

На рисунке 12 приведена карта гидрофобных контуров. Как видно из представленных данных наибольший вклад в активность вносят арильные гидрофобные фрагменты и фрагменты бензильного заместителя сульфонильной группы.

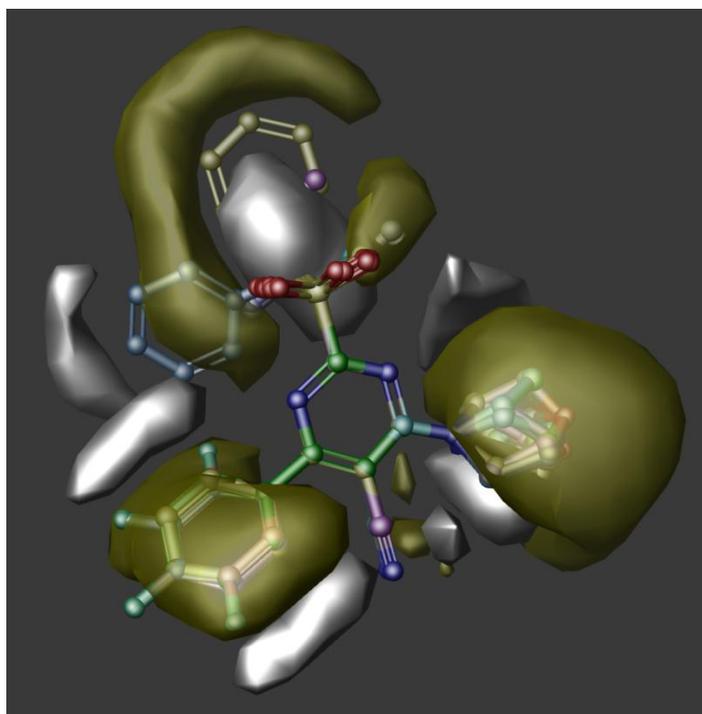


Рисунок 12 – Влияние факторов гидрофобности на биологическую активность. Желтый контур – положительный вклад, белый контур – отрицательный вклад.

Анализ данных карты акцепторов водородной связи (рисунок 13) выявил область расположения сульфонильной группы, дающей положительный вклад в биологическую активность и область расположения морфолинового фрагмента дающей отрицательный вклад. Представленные данные хорошо коррелируют со значениями цитотоксичности.

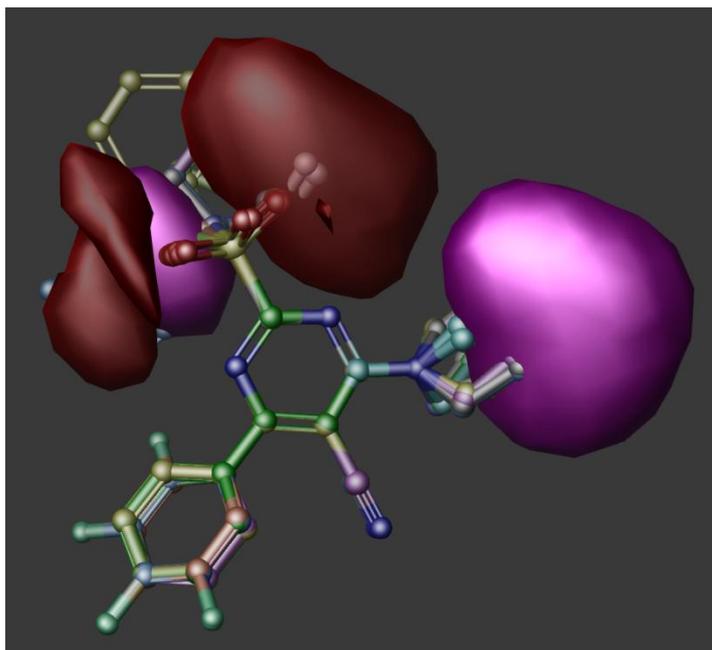


Рисунок 13 – Влияние факторов акцептора водородной связи на биологическую активность. Красный – положительный вклад, пурпурный – отрицательный вклад.

3. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Все расчеты в рамках выполненной работы проводили с применением программного пакета SchrodingerSuite 2017-1.

3.1. Подготовка комбинаторной библиотеки

Трехмерные структуры исследуемых соединений были построены с использованием панели сборки Maestro в пакете Schrödinger. Модуль Ligprep был использован для генерации всех возможных состояний в физиологическом диапазоне pH 7 ± 2 с учетом наименьшей потенциальной энергии конформации лиганда, с использованием силового поля OPLS_2005. Все полученные структуры соединений были использованы для исследований молекулярной стыковки и свойств ADME с использованием модуля Qikprop.

3.2. QSAR

Построение классических PSL моделей производили с помощью инструмента AutoQSAR.

3.3. 3D-QSAR

Для исходных соединений были назначены заряды Gasteiger-Huckel30. Для анализа CoMFA и CoMSIA использовались стандартные силовые поля. Создана трехмерная кубическая решетка размерностью 3 Å в каждом направлении с каждым пересечением решетки регулярно распределенной сетки размерностью 2.0 Å. Стерео- и электростатические параметры были рассчитаны в полях CoMFA, в то время как гидрофобные, акцепторные и донорные параметры в дополнение к стерическим и электростатическим были рассчитаны в полях CoMSIA на каждой решетке. Атом углерода (sp^3 -гибридизация) использовался в качестве зондового атома для генерирования пространственных энергий (Леннарда-Джонса) и заряд +1 для генерации энергий электростатического потенциала (кулоновского потенциала). Использовалась диэлектрическая проницаемость, зависящая от расстояния. Стерические и электростатические поля были сконденсированы до +30,00 ккал/моль.

Анализ PLS проводился для корреляции вычисленных полей с биологической активностью ($pIC_{50} = -\log IC_{50}$) для набора исследуемых лигандов. Молекулы были случайным образом разделены на тренировочные и тестовые комплекты в соотношении 1:1 и 3:1. Учебный набор использовался для генерации модели QSAR, а тестовый набор использовался для проверки модели генератора. В статистическом анализе PLS в качестве зависимых переменных использовались значения биологической активности лигандов. Для определения оптимального количества компонентов (ONC) и коэффициента q^2_{LOO} были выполнены перекрестные проверки LOO. Валидация моделей CoMFA и CoMSIA проводилась путем прогнозирования активности соединений с установленными критериями.

ЗАКЛЮЧЕНИЕ

По результатам проделанной работы можно сделать следующие выводы:

1. С применением алгоритма AutoQSAR получен ряд валидных моделей количественного соотношения структура-свойства для серии производных 5-цианопиримидина и их цитотоксичности в отношении эпидермоидной карциномы. Установлено, что наилучшие значения коэффициента детерминации и оценочной (скоринг) функции удается получить с применением линейных корреляционных моделей (kpls_linear).
2. Методами сравнительного анализа молекулярных полей осуществлен 3D-QSAR. Установлены основные зависимости между трехмерным строением исследуемых соединений и проявляемой ими биологической активностью. Определены количественные характеристики представленной модели, с применением набора лигандов для валидации, показано что при тренировочной выборке в 75% достигаются наилучшие значения коэффициента корреляции перекрёстного контроля Q^2 достигающие значения выше 0.75.
3. Полученные в работе модели могут быть рекомендованы для использования в оптимизации комбинаторной библиотеки на основе 5-цианопиримидина, с целью изыскания более активных соединений-лидеров.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Crum-Brown A., Fraser T. R. On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the ammonium bases, derived from Strychia, Brucia, Thebaia, Codeia, Morphia and Nicotia //Trans. R. Soc. Edinburgh. – 1868. – Vol. 25. – P.151-203.
2. Richet M. C. Note sur le rapport entre la toxicité et les propriétés physiques des corps //Compt Rend Soc Biol (Paris). – 1893. – Vol. 45. – P. 775-6.
3. Meyer H. Zur theorie der alkoholnarkose //Archiv für experimentelle Pathologie und Pharmakologie. – 1899. – Vol. 42. – №. 2-4. – P. 109-118.
4. Overton C. E. Studien über die Narkose zugleich ein Beitrag zur allgemeinen Pharmakologie. – Fischer, 1901.
5. Lipnick R. L. Charles Ernest Overton: Narcosis studies and a contribution to general pharmacology //Trends in Pharmacological Sciences. – 1986. – Vol. 7. – P.161-164.
6. Fuhner H., and Neubauer, E., Arch. Exp. Path. Pharm. 56, 333-345 (1907)
7. Ferguson J. The use of chemical potentials as indices of toxicity //Proc. R. Soc. Lond. B. – 1939. – Vol. 127. – №. 848. – P. 387-404.
8. Bruice T. C., Kharasch, N., and Winzler, R. J., Arch. Biochem. Biophys. 62,305 -317 (1956).
9. Zahradnik R., Chvapil M. Study of the relationship between the magnitude of biological activity and the structure of aliphatic compounds //Experientia. – 1960. – Vol. 16. – №. 11. – P. 511-512.
10. Zahradnik R. Influence of the structure of aliphatic substituents on the magnitude of the biological effect of substances //Archives internationales de pharmacodynamie et de therapie. – 1962. – Vol. 135. – P. 311-329.
11. Zahradnik R. Correlation of the biological activity of organic compounds by means of the linear free energy relationships //Experientia. – 1962. – Vol. 18. – №. 11. – P. 534-536.

12. Hammett L. P. // *Physical Organic Chemistry. Reaction Rates, Equilibria and Mechanism*, 2nd Ed., McGraw-Hill. – P. 197.
13. Hansen O. R. Hammett series with biological activity // *Acta chem. scand.* – 1962. – Vol. 16. – №. 7.
14. Hansch C. et al. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients // *Nature*. – 1962. – Vol. 194. – №. 4824. – P. 178.
15. Hansch C., Fujita T. p - σ - π Analysis. A method for the correlation of biological activity and chemical structure // *Journal of the American Chemical Society*. – 1964. – Vol. 86. – №. 8. – P. 1616-1626.
16. Free S. M., Wilson J. W. A mathematical contribution to structure-activity studies // *Journal of Medicinal Chemistry*. – 1964. – Vol. 7. – №. 4. – P. 395-399.
17. Hansch C. Quantitative approach to biochemical structure-activity relationships // *Accounts of Chemical Research*. – 1969. – Vol. 2. – №. 8. – P. 232-239.
18. Hansch C. Quantitative structure-activity relationships in drug design // *Drug design*. – 1971. – Vol. 1. – P. 271-342.
19. Hansch C., Clayton J. M. Lipophilic character and biological activity of drugs II: The parabolic case // *Journal of Pharmaceutical Sciences*. – 1973. – Vol. 62. – №. 1. – P. 1-21.
20. Fujita T., Ban, T., Structure-Activity Study of Phenethylamines as Substrates of Biosynthetic Enzymes of Sympathetic Transmitters, *J. Med. Chem.* – 1971. – Vol. 14, P. 148 – 152.
21. Hansch C., Yoshimoto M. Structure-activity relations in immunochemistry. 2. Inhibition of complement by benzamidines // *Journal of medicinal chemistry*. – 1974. – Vol. 17. – №. 11. – P. 1160-1167.
22. Kubinyi H. Quantitative structure-activity relationships. 2. A mixed approach, based on Hansch and Free-Wilson analysis // *Journal of medicinal chemistry*. – 1976. – Vol. 19. – №. 5. – P. 587-600.

23. Kubinyi H. Quantitative structure-activity relations. 7. The bilinear model, a new model for nonlinear dependence of biological activity on hydrophobic character //Journal of medicinal chemistry. – 1977. – Vol. 20. – №. 5. – P. 625-629.
24. Brugger W. E., Stuper A. J., Jurs P. C. Generation of descriptors from molecular structures //Journal of Chemical Information and Computer Sciences. – 1976. – Vol. 16. – №. 2. – P. 105-110.
25. Kirschner G. L., and Kowalski, B. R., Drug Design// Academic Press. 1979. Vol.8. P. 73-131.
26. Wold. S., Ruhe. A., Wold. H., Dunn III, W. J. // SIAM J. Sci. Stat. Comput. - Vol.5. P. 735-743.
27. Dunn III, W. J., Wold, S., Edlund, U., Hellberg, S., and Gasteiger, J. // Quant. Struct.-Act.Relat. – 1984.- Vol. 3. - P. 131- 137
28. Fukunaga J. Y., Hansch C., Steller E. E. Inhibition of dihydrofolate reductase. Structure-activity correlations of quinazolines //Journal of medicinal chemistry. – 1976. – Vol. 19. – №. 5. – P. 605-611.
29. Hoeltje H. D., Kier L. B. Theoretical approach to structure-activity relations of chloramphenicol and congeners //Journal of medicinal chemistry. – 1974. – Vol. 17. – №. 8. – P. 814-819.
30. Höltje H. D., Tintelnot M. Theoretical Investigations on Interactions Between Pharmacological Molecules and Receptor Models V: Construction of a Model for the Ribosomal Binding Site of Chloramphenicol //Molecular Informatics. – 1984. – Vol. 3. – №. 1. – P. 6-9.
31. Crippen G. M. Distance geometry approach to rationalizing binding data //Journal of medicinal chemistry. – 1979. – Vol. 22. – №. 8. – P. 988-997.
32. Ghose A. K., Crippen G. M. Quantitative structure-activity relationship by distance geometry: quinazolines as dihydrofolate reductase inhibitors //Journal of medicinal chemistry. – 1982. – Vol. 25. – №. 8. – P. 892-899.

33. Goodford P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules // *Journal of medicinal chemistry*. – 1985. – Vol. 28. – №. 7. – P. 849-857.
34. Wise M., *QSAR and Strategies in the Design of Bioactive Compounds* // *European Symposium on Quantitative Structure-Activity Relationships 1984: Bad Segeberg, Germany*. – Distribution, USA and Canada, VCH Publishers. 1985. P. 19-29.
35. Cramer III, R. D., and Bunce, J. D., *QSAR in Drug Design and Toxicology* (Proceedings of the 6th European QSAR Symposium, Portoroz, 1986), Hadzi, D., and Jerman-Blazic, B., (ed.) // *Pharmacochem. Libr.* 10, Elsevier, Amsterdam. –1987.–P. 3- 12.
36. Cramer R. D., Patterson D. E., Bunce J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins // *Journal of the American Chemical Society*. – 1988. – Vol. 110. – №. 18. – P. 5959-5967.
37. Cramer R. D., and Wold, S. B., U.S. Pat. 5 025 388 (June 18, 1991).
38. Kubinyi H. // *3D QSAR in drug design: volume 1: theory methods and applications*. – Springer Science & Business Media.–1993. – Vol. 1.
39. Ramsden C. A. Hansch, C., Sammes, P. G., and Taylor, J. B // *Quantitative Drug Design, of Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds*.–1990.–Vol.4.
40. Purcell W. P., Bass G. E., Clayton J. M. *Strategy of drug design: a guide to biological activity*. – John Wiley & Sons.–1973.
41. Martin Y. C. // *Quantitative Drug Design. A Critical Introduction (Medicinal Research Series)*.–1978.–Vol. 8.
42. Seydel J. K., Schaper K. J. // *Chemische Struktur und biologische Aktivität von Wirkstoffen: Methoden der quantitativen Struktur-Wirkung-Analyse*. Verlag Chemie.–1979.
43. Franke R. // *Optimierungsmethoden in der WirkstoffForschung. Quantitative Struktur-Wirkungs-Analyse*, Akademie-Verlag, Berlin, 1980; revised English translation:

- Franke, R., Theoretical Drug Design Methods, Pharmacochem. Libr. 7, Elsevier, Amsterdam.–1984.
44. Topliss J. G. // Quantitative Structure-Activity Relationships of Drugs, Academic Press.–1983.
45. Golberg. L. // Structure- Activity Correlation as a Predictive Tool in Toxicology. Fundamentals, Methods and Applications, –1983.
46. Jolles, G., Woolridge, K. R. // Drug Design: Fact or Fantasy, Academic Press,–1984.
47. Karcher W., Devillers J. (ed.). Practical applications of quantitative structure-activity relationships (QSAR) in environmental chemistry and toxicology // Springer Science & Business Media, 1990. – Vol. 1.
48. Kier L. B., Hall L. H. Molecular Connectivity in Chemistry and Drug Research Academic //New York. – 1976.
49. Rekker R. F. The Hydrophobic Fragmental Constant. Its Derivation and Application. A Means of Characterizing Membrane Systems // Pharmacochem. – Libr. 1.– Elsevier, Amsterdam. –1977.
50. Hansch C., Leo A. // Substituent constants for correlation analysis in chemistry and biology. – Wiley.– 1979.
51. Yalkowsky S. H., Sinkula, A. A., Valvani, S. C. // Physical Chemical Properties of Drugs (Medicinal Research Series), Marcel Dekker, New York.– 1980. Vol. 10.
52. Balaban A. T. Steric Fit in Quantitative Structure-activity Relations: Topological Indices. II. – Universitatea din Timişoara. Facultatea de Ştiinţe de Naturii, –1980.
53. Perrin D. D., Dempsey B., Serjeant E. P. pKa prediction for organic acids and bases. – London; New York: Chapman and Hall.–1981. – Vol. 1.
54. Dunn III, W. J., Block, J. H., Pearlman, R. S. (ed.)// Partition Coefficient: Determination and Estimation, Pergamon, New York.–1986.
55. McGowan J. C., Mellors, A. // Molecular Volumes in Chemistry and Biology: Applications Including Partitioning and Toxicity, Ellis Horwood Ltd., Chichester, – 1986.

56. Rekker R. F., Mannhold R. Calculation of drug lipophilicity: the hydrophobic fragmental constant approach. – Weinheim: Vch. –1992. – Vol. 112.
57. Chapman N. B., and Shorter, J., (eds.) Correlation Analysis in Chemistry: Recent Advances, Plenum Press, New York.. – 1978.
58. Stuper A. J., Brugger, W. E., Jurs, P. C. // Computer Assisted Studies of Chemical Structure and Biological Function. Wiley.– 1979.
59. Dean P. M. // Molecular Foundations of Drug-Receptor Interaction, Cambridge University Press, Cambridge.–1987.
60. Martin Y. C., (ed.)// Modern Drug Research: Path to Better and Safer Drugs. – Informa Health Care, 1989. – Vol. 12.
61. Richards W. G. // Computer-Aided Mofimlar Design, IBC Technical Services. –1989.
62. Perun T. J., Propst, C. L.// Computer-Aided Drug Design. Methods and Applications.– 1989.
63. Johnson M. A., Maggiora, G. M. // Concepts and Applications of Molecular Similarity, Wiley. –1990.
64. Wermuth, C. G., Koga, N., Konig, H., Metcalf, B. W., (ed.) // Medicinal Chemistry, for the 21st Century. –1992.
65. Ariens E. J., (ed.) // Drug Design.–1971.–Vol. 1.
66. Jucker E., (ed.) // Fortschr. Arzneimittelforsch. (Prog. Drug Res.). –1959.–Vol. 1.
67. Pharmacochem.– Libr. 1.–1977.
68. Gould R. F., (ed.) // Biological Correlations- The Hansch Approach, Advan. Chem. Ser. 114. –1972.
69. Tichy M., (ed.)// Quantitative Structure-Activity Relationships (Proceedings of the Confercncc on Chemical Structure-Biological Activity Relationships: Quantitative Approaches, Prague, 1973). Experientia Suppl. 23. Birkhauser Verlag, Basel.–1976.
70. Keverling Buisman J. A.,(ed.) // Biological Activity and Chemical Structure (Proceedings of the IUPAC-IUPHAR Symposium, Noordwijkerhout, 1977), Phurmacochem. Lihr. 2, Elsevier, Amsterdam.–1977.

71. Franke R., Oehme P. // Quantitative Structure-Activity Analysis (Proceedings of the Second Symposium on Chemical Structure-Biological Activity Relationships: Quantitative Approaches, Suhl, 1976), Akademie-Verlag, Berlin.– 1978.
72. Olsen, E. C., Christofferson, P. E. // Computer-Assisted Drug Design (ACS Symp. Series 112), Am. Chem. SOC., Washington. 1979.
73. Knoll J., and Darvas, F. (ed.) // Chemical Structure-Biological Activity Relationships. Quantitative Approaches (Proceedings of the 3rd European QSAR Symposium, Budapest, 1979), Akadtmiai Kiado, Budapest,– 1980.
74. Keverling Buisman, J.A.(ed.) // Strategy in Drug Research (Proceedings of the Second IUPAC-IUPHAR Symposium, Noordwijkerhout, 1981), Pharmacochem. Libr. 4.– 1982.
75. Dearden S. C., (ed.) // Quantitative Approaches to Drug Design (Proceedings of the 4th European Symposium on Chemical Structure- Biological Activity: Quantitative Approaches, Bath, 1982), Pharmacochem. Libr. 6.–1983.
76. Kuchar M. QSAR in Design of bioactive compounds //Proceedings of the 1st International Telesymposium on Medicinal Chemistry, Prous, Barcelona, Spain. – 1984.
77. Tichy M. QSAR in toxicology and xenobiochemistry: proceedings of a symposium, Prague, Czechoslovakia, September 12-14, 1984 (a satellite symposium to the 5th European Symposium on Chemical Structure-Biological Activity, Bad Segeberg, FRG, September 17-21, 1984). – Elsevier. – 1985.
78. Seydel J. K. QSAR and Strategies in the Design of Bioactive Compounds //European Symposium on Quantitative Structure-Activity Relationships 1984: Bad Segeberg, Germany). – Distribution, USA and Canada, VCH Publishers, 1985.
79. Harms A. F. et al. Innovative approaches in drug research //Noordwijkerhout Symposium on Medicinal Chemistry 1985). – Elsevier. – 1986.

80. Hadzi D., Jerman-Blazic B. QSAR in drug design and toxicology //European Symposium on Quantitative Structure-Activity Relationships 1986: Portoroz, Slovenia). – Elsevier. – 1987.
81. Kaiser K. L. E., Palabrica V. S., Ribo J. M. QSAR of acute toxicity of mono-substituted benzene derivatives to *Photobacterium phosphoreum* //QSAR in Environmental Toxicology-II. – Springer, Dordrecht, 1987. – P.153-168.
82. Fauchère J. L. QSAR: quantitative structure-activity relationships in drug design: proceedings of the 7th European Symposium on QSAR, Interlaken, Switzerland, 5-9 September 1988. – Alan R. Liss, 1989. – Vol. 291.
83. Silipo C., Vittoria A. QSAR, rational approaches to the design of bioactive compounds //European Symposium on Quantitative Structure-Activity Relationships 1990: Sorrento, Italy). – Distributors for the US and Canada, Elsevier Science.–1991.
84. Wermuth C. G. (ed.) // Trends in QSAR and Molecular Modelling 92: Proceedings of the 9th European Symposium on Structure-Activity Relationships: QSAR and Molecular Modelling, September 7-11, 1992, Strasbourg, France. – Escom, 1993.
85. Darvas F. (ed.) // Abstracts of Publications Related to QSAR. - Quant. Struct.-Act. Relat. 1.–1982.
86. CAS Selects “Structure-Activity Relationships”, Chemical Abstracts Service (service code 04S), Columbus, Ohio, U. S. A.
87. Ringdoc Drug Information, Profile 38, “Structure Activity”, Derwent Publications Ltd., London, England.
88. Molecular Modelling & Computational Chemistry Results 1 (1 992) ff., MMCC Publishing, Cambridge, MA, U. S. A.
89. a) J. Med. Chem. b) Eur. J. Med. Chem. c) Med. Chem. Res. d) J. Comput.-Aided Mol. Des.
90. Craig P. N. // Drug № 18. –1984.–P.123-30.
91. Tute M. S., ref. 39, –P. 1-31.
92. Rekke R. F. // Quant. Struct.-Act. Relat. 11.–1992. –P. 195- 199.

93. Van de Waterbeemd, H. // *Quant. Struct.-Act. Relat.* 11.–1992.–P. 200-204.
94. Ehrlich P. *Chemotherapeutics: scientific principles, methods, and result* // *Lancet.* – 1913. – Vol. 16. – P.445-451.
95. Langley J. N. *On the physiology of the salivary secretion* // *The Journal of physiology.* – 1878. – Vol. 1. – №. 4-5. – P. 339-369.
96. Fischer E. // *Ber. Dtsch. Chem. Ges.* 27.– 1894. – P. 2985-2993.
97. Ariens E. J. // *Drug Design.*–1971.–Vol. 1.–P. 1-270.
98. Emmett, J. C., ed., *Membranes and Receptors Comprehensive Medicinal Chemistry // The Rational Design, Mechanistic Study & Therapeutic Application of Chemical Compounds*, Hansch, C., Sammes, P. G., and Taylor, J. B., (eds.) Pergamon Press, Oxford.–1990.–Vol. 3.
99. Dickerson R. E., Geis I. // *Struktur und Funktion der Proteine.* – Weinheim: Chemie. – 1971.
100. De Stevens, G. // *Fortschr. Arzneimittelforsch. (Prog. Drug. Res.)* 30.–1986. – P.189-203
101. Williams R. J. P. *Flexible drug molecules and dynamic receptors* // *Angewandte Chemie International Edition in English.* – 1977. – Vol. 16. – №. 11. – P. 766-777.
102. Roberts G. C. K., ref. 75, P. 91-98.
103. Deisenhofer, J., Michel, H. // *Science.* – 1989.– P.1463-1480.
104. Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., and Downing, K. H. // *Mol. Biol.*– 1990. – P. 899-929.
105. Weiss M. S. et al. *Molecular architecture and electrostatic properties of a bacterial porin* // *Science.* – 1991. – Vol. 254. – №. 5038. – P. 1627-1630.
106. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard., and Tasumi, M. // *Mol. Biol.* –1977, – P. 535-542 .
107. Blundell, T. L., and Johnson, L. N., *Protein Crystallography*, Academic Press, London, 1976.

108. Brandcn, C., and Tooze, J., Introduction to Protein Structure, Garland Publishing Company, New York, 1991.
109. Bode W., and Huber R. // Biochem. – 1992.–P. 433-451.
110. Bode W., and Huber R.– ref. 64.–P. 73- 118.
111. Findlay J., Eliopoulos E. Three-dimensional modelling of G protein-linked receptors //Trends in pharmacological sciences. – 1990. – Vol. 11. – №. 12. – P. 492-499.
112. a) Trumpp-Kallmeyer, S., Hoflack, J., Bruinvels A. and Hibert M., J. Med. Chem. 35, 3448-3462 (1992); b) Hibert, M. F., Trumpp-Kallmeyer, S., Hoflack, J., and Bruinvels.A, Trends Pharmacol. Sci. 14, 7- 12 (1993).
113. Humblet C., and Mirzadegan T. // Ann. Rep. Med. Chem. -1992.–Vol. 27, -P. 291 -300.
114. Herbette L. G., Chester D. W., and Rhodes D. G // Biophys. -1986. -Vol.49, -P. 91 -94.
115. Mason R. P., Rhodes D. G., Herbette L. G. Reevaluating equilibrium and kinetic binding parameters for lipophilic drugs based on a structural model for drug interactions with biological membranes //Journal of medicinal chemistry. – 1991. – Vol. 34. – №. 3. – P. 869-877.
116. Andrews P. R., and Tintelnot M., ref. 39, C. 321 -347.
117. Tintelnot M., and Andrews P. R. // J. Comput. -Aided Mol. Des. –1989, -Vol.3. –P. 67-84.
118. Williams, D. H., Aldrichimicu Acta –1991. –Vol.24. –P. 71-80.
119. Williams, D. H., Aldrichimicu Acta 25, 9 (1992).
120. Murray-Rust P., Glusker J. P. Directional hydrogen bonding to sp²-and sp³-hybridized oxygen atoms and its relevance to ligand-macromolecule interactions //Journal of the American Chemical Society. – 1984. – Vol. 106. – №. 4. – P. 1018-1025.

121. Fersht A. R. et al. Hydrogen bonding and biological specificity analysed by protein engineering // Nature. – 1985. – Vol. 314. – №. 6008. – P. 235.
122. Fersht A. R. et al. Hydrogen bonding and biological specificity analysed by protein engineering // Nature. – 1985. – Vol. 314. – №. 6008. – P. 235.
123. Fersht A. R. The hydrogen bond in molecular recognition // Trends in Biochemical Sciences. – 1987. – Vol. 12. – P. 301-304.
124. Ward W. H. J., Timms D. and Fersht A. R. // Trends Pharmacol.Sci. –1990. Vol.11. –P.280-284.
125. Wolfenden R., Kati W. M. Testing the limits of protein-ligand binding discrimination with transition-state analogue inhibitors //Accounts of chemical research. – 1991. – Vol. 24. – №. 7. – P. 209-215.
126. a)Doig A. J., Williams D. H. Binding energy of an amide-amide hydrogen bond in aqueous and nonpolar solvents //Journal of the American Chemical Society. – 1992. – Vol. 114. – №. 1. – P. 338-343; b) Searle, M. S., and Williams, D. H., ref. 84, P. 86-92.
127. Sharp K. A. et al. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models //Biochemistry. – 1991. – Vol. 30. – №. 40. – P. 9686-9697.
128. Hitzemann R. Thermodynamic aspects of drug-receptor interactions //Trends in pharmacological sciences. – 1988. – Vol. 9. – №. 11. – P. 408-411.
129. Andrews P. R., Craik D. J., Martin J. L. Functional group contributions to drug-receptor interactions //Journal of medicinal chemistry. – 1984. – Vol. 27. – №. 12. – P. 1648-1657.
130. Andrews P. Functional groups, drug-receptor interactions and drug design //Trends in Pharmacological Sciences. – 1986. – Vol. 7. – C.148-151.
131. Goodford P. J.– ref. 83. –P. 49-55.

132. Hansch C. et al. Aromatic substituent constants for structure-activity correlations //Journal of medicinal chemistry. – 1973. – Vol. 16. – №. 11. – P. 1207-1216.
133. Hansch C. et al. Substituent constants for correlation analysis //Journal of medicinal chemistry. – 1977. – Vol. 20. – №. 2. – P. 304-306.
134. Norrington F. E. et al. Physicochemical-activity relations in practice. 1. Rational and self-consistent data bank //Journal of medicinal chemistry. – 1975. – Vol. 18. – №. 6. – P. 604-607.
135. Hansch C., Unger S. H., Forsythe A. B. Strategy in drug design. Cluster analysis as an aid in the selection of substituents //Journal of medicinal chemistry. – 1973. – Vol. 16. – №. 11. – P. 1217-1222.
136. Verloop A., Drug Design, Ariens E. J. et al. Academic Press // New York. – 1972.–Vol. 111.– P. 133- 187.
137. Martin Y.C., Drug Design, Ariens E. J. et al. Academic Press // New York. – 1979.–Vol. 8.– P. 1-72.
138. Franke R.– ref. 78.–P. 59-78.
139. Buchwald P., Bodor N. Octanol-water partition: searching for predictive models //Current medicinal chemistry. – 1998. – Vol. 5. – №. 5. – P. 353-380.
140. Van de Waterbeemd H., Testa B., Carrupt P.-A. et al.– ref. 82. –P. 123- 126.
141. Van de Waterbeemd H. et al. Pattern recognition study of QSAR substituent descriptors //Journal of computer-aided molecular design. – 1989. – Vol. 3. – №. 2. – P. 111-132.
142. Van de Waterbeemd H., Carrupt P.A. and El Tayar N. ref. 82. – P. 101-103.
143. Boyd D. B. and Seward C. M. – ref. 83.– P. 167- 170.
144. Streich W. J., Dove S., Franke R. Rational selection of test series. 1. Principal component method combined with multidimensional mapping //Journal of medicinal chemistry. – 1980. – Vol. 23. – №. 12. – P. 1452-1456.

145. Cramer III R. D. BC (DEF) parameters. 1. The intrinsic dimensionality of intermolecular interactions in the liquid state //Journal of the American Chemical Society. – 1980. – Vol. 102. – №. 6. – P. 1837-1849.
146. Cramer III R. D. Quont. Struct. // Act. Relut.– 1983. – Vol. 2. – P. 7- 12.
147. Cramer III R. D.– ref. 75.– P. 3- 14.
148. Hyde R.– ref. 82.– P. 91 -95.
149. Alunni S., Clementi S., Edlund. et al. Acta Chewt. // SCLVZ~. B37. – 1983. – P. 47-53.
150. Skagerberg B., Bonelli D., Clementi S., et al. Quont. Szruct. // Act. Relut. – 1989. – Vol. 8.– P. 32-38.
151. Sjostrom M. and Wold S. // J. Mof. Evof. – 1985. – Vol. 22.– P. 272-277.
152. Jonsson J., Eriksson L., Hellber, S. et al. Quont. Struct. //Act. Relut.–1989.– Vol. 8.– P. 204-209.
153. Dill K. A. et al. The meaning of hydrophobicity //Science. – 1990. – Vol. 250. – №. 4978. – P. 297-299.
154. Taylor P. J. ref. – 39. – P. 241 -294.
155. Leo A., Hansch C., Elkins D. Partition coefficients and their uses // Chemical reviews. – 1971. – Vol. 71. – №. 6. – P. 525-616.
156. Kubinyi, H., Fortschr. Arzneimittelfbrsch. (Prog. Drug Res.) 23, 97- 198 (1979).
157. Coats E.A. – ref. 51.– P. 111-139.
158. Chou J. T. and Jurs P. C. – ref. 51. – P. 163-199.
159. Dearden J. C., Enuiron. Hc.alth Pwspect. 61, 203-228 (1985).
160. Leo A. J.– ref. 39.– P. 295-319.
161. Smith R. N., Hansch C., Ames M. M. Selection of a reference partitioning system for drug design work //Journal of pharmaceutical sciences. – 1975. – Vol. 64. – №. 4. – P. 599-606.
162. Hansch, C., and Leo, A., Partition Coefficient Data Bank. // MEDCHEM Project. Pomona College. Claremont. CA. USA.

163. Hansch C. and Dunn III W. J. // J. Pharm.– 1972. – Vol. 61.–P. 1- 19.
164. Collander R., // Acta Chem. Scand. – 1951. – Vol. 5. –P.774-780.
165. Yeh K. C. and Higuchi W. I. // J. Pharm.– 1976. – Vol. 65,–P. 80-86.
166. Leo A. and Hansch C. // J. Org. Chem. – 197. – Vol. 36.–P. 1539- 1544.
167. Leo A.– ref. 68.– P. 51 -60.
168. Seeman P., Roth S., Schneider H. The membrane concentrations of alcohol anesthetics //Biochimica et Biophysica Acta (BBA)-Biomembranes. – 1971. – Vol. 225. – №. 2. – P. 171-184.
169. DUNN III W. J., WOLD S. Statistical Analysis of the Partition Coefficient //Acta Chemica Scandinavica B. – 1978. – Vol. 32. – P. 536-542.
170. Franke R., Kuhne R. and Dove S.– ref. 75.– P. 15-32.
171. Koehler M. G., Grigoras S. and Dunn III W. J. Quant. Struc.r. // Act.Relut. – 1988. – Vol. 7. – P. 150-159.
172. Seiler, P., Eur. J . Med. Chem. 9, 473-479 (1974).
173. El Tayar, N., Tsai, R.-S., Testa, B., Carrupt, P.-A., and Leo, A., J . Pharm. Sci. 80, 590- 598 (1991).
174. Fan, W., El Tayar, N., Testa, B., and Kier, L. B., J . Pkys. Chem. 94, 4764-4766 (1990).
175. Fan, W., El Tiyar, N., Testa, B., Carrupt, P.-A., and Kicr, L. B., ref. 83, p. 67-70.
176. Wolfenden R. Waterlogged molecules //Science. – 1983. – Vol. 222. – №. 4628. – P. 1087-1093.
177. Fujita T., Nishioka T., Nakajima M. Hydrogen-bonding parameter and its significance in quantitative structure-activity studies //Journal of medicinal chemistry. – 1977. – Vol. 20. – №. 8. – P. 1071-1081.
178. Hadzi, D., Kidric, J., Koller, J., and Mavri, J.,?. Mol. Struct. (Theochem) 237, 139- I50 (1990).

179. Raevsky, O.A., Grigor'ev, V. U., Solov'ev, V. P., Kireev, D. B., Sapegin, A. M., and ZeGrov, N. S., ref. 83, p. 135-138.
180. Young R. C. et al. Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H₂ receptor histamine antagonists //Journal of medicinal chemistry. – 1988. – Vol. 31. – №. 3. – P. 656-671.
181. Ganellin, C. R., Brown, T. H., Griffiths, R., Jones, M., Mitchell, R. C., Rana, K., Saunders, D., Smith, I. R., Sore, N. E., Wilks, T. J., and Young, R. C., ref. 83, p. 103-110.
182. Leeson P. D. et al. Role of hydrogen bonding in ligand interaction with the N-methyl-D-aspartate receptor ion channel //Journal of medicinal chemistry. – 1990. – Vol. 33. – №. 5. – P. 1296-1305.
183. Abraham, M. H., Whiting, G. S., Alarie, Y., Morris, J. J., Taylor, P. J., Doherty, R. M., Taft, R. W., and Niclsen, G. D., *Quint. Struct.-Act. Relat.* 9, 6- 10 (1990).
184. Altomare, C., Tsai, R.-S., El Tayar, N., Testa, B., Carrupt, P.-A., Carotti, A., and De Benedetti, P. G., ref. 83, p. 139-142.
185. Leahy, D. E., Taylor, P. J., and Wait, A. R., *Quint. Struct.-Act. Relat.* 8, 17-31 (1989).
186. Leahy, D. E., Morris, J. J., Taylor, P. J., and Wait, A. R., *J. Chem. SOC. Perkin Trans. 2*, 705-722, 723-731 (1992).
187. Rhodcs, D. G., Sarmiento, J. G., and Herbette, L. G., *Mol. Plzarinacof.* 27, 612-623 (1985).
188. Herbette, L. G. and Gruner, S. M., *Deu. Curdiovusc.* 68, 353-365 (1987).
189. Herbette L. G., Erve Y. M. H. V., Rhodes D. G. Interaction of 1, 4 dihydropyridine calcium channel antagonists with biological membranes: lipid bilayer partitioning could occur before drug binding to receptors //Journal of molecular and cellular cardiology. – 1989. – Vol. 21. – №. 2. – P. 187-201.

190. Mason R. P. et al. Partitioning and location of Bay K 8644, 1, 4-dihydropyridine calcium channel agonist, in model and biological membranes // *Biophysical journal*. – 1989. – Vol. 55. – №. 4. – P. 769-778.
191. Moring, J., and Herbette, L. G., *Biophys. J.* 57, 523a (1990).
192. Herbette L. G., Rhodes D. G., Mason R. P. New approaches to drug design and delivery based on drug-membrane interactions // *Drug design and delivery*. – 1991. – Vol. 7. – №. 2. – P. 75-118.
193. Seydel J. K. et al. QSAR and multivariate data analysis of amphiphilic benzylamines and their interaction with various phospholipids determined by different methods // *Molecular Informatics*. – 1989. – Vol. 8. – №. 4. – P. 266-278.
194. Seydel, J. K., Wiese, M., Cordes, H. P., Chi, H. L., Schaper, K.-J., Coats, E. A., Kunz, B., Engel, J., Kutscher, B., and Emig, H., ref. 83, p. 367-376.
195. Seydel J. K. et al. The Importance of Drug-Membrane Interaction in Drug Research and Development // *Molecular Informatics*. – 1992. – Vol. 11. – №.2. – P. 205-210.
196. Schwyzer R. New principle in QSAR: membrane requirements // *Journal of receptor research*. – 1991. – Vol. 11. – №. 1-4. – P. 45-57.
197. Schwyzer R. Peptide-membrane interactions and a new principle in quantitative structure-activity relationships // *Biopolymers*. – 1991. – Vol. 31. – №. 6. – P. 785-792.
198. Dunn W. J. Molar refractivity as an independent variable in quantitative structure-activity studies // *European Journal of Medicinal Chemistry*. – 1977. – Vol. 12. – №. 2. – P. 109-112.
199. Silipo, C., and Vittoria, A., ref. 39, p. 153-204.
200. Blaney, J. M., and Hansch, C., ref. 39, p. 459-496.
201. Gryllaki, M., van de Waterbeemd, H., and Testa, B., ref. 78, p. 273-276.

202. Ahmad P., Fyfe C. A., Mellors A. Parachors in drug design //Biochemical pharmacology. – 1975. – Vol. 24. – №. 10. – P. 1103-1109.
203. Dearden, J. C., Bradburne, S. J. A., and Abraham, M. H., ref. 83, p. 143-150.
204. Hansch C., Gao H., Hoekman D. A generalized approach to comparative QSAR // Comparative Qsar. – 1998. – C.285-368.
205. Shah K. J., Coats E. A. Design, synthesis, and correlation analysis of 7-substituted 4-hydroxyquinoline-3-carboxylic acids as inhibitors of cellular respiration //Journal of medicinal chemistry. – 1977. – Vol. 20. – №. 8. – P. 1001-1006.
206. Chapman N. B., and Shorter, J. eds.// Advances in Linear Free Energy Relationships, Plenum Press, London.– 1972.
207. Perrin, D. D., ref. 51, p. 1-48.
208. Bowden, K., ref. 39, p. 205-239.
209. Loew, G. H., and Burt, S. K., ref. 39, p. 105-123.
210. Swain C. G., Lupton E. C. Field and resonance components of substituent effects //Journal of the American Chemical Society. – 1968. – Vol. 90. – №. 16. – P. 4328-4337.
211. Swain C. G. et al. Substituent effects on chemical reactivity. Improved evaluation of field and resonance components //Journal of the American Chemical Society. – 1983. – Vol. 105. – №. 3. – P. 492-502.
212. Swain C. G. Substituent and solvent effects on chemical reactivity //The Journal of Organic Chemistry. – 1984. – Vol. 49. – №. 11. – P. 2005-2010.
213. Exner, O., ref. 294, p. 1-69.
214. Shorter, J., ref. 57, p. 119- 173.
215. Reynolds W. F., Topsom R. D. Field and resonance substituent constants for aromatic derivatives: limitations of Swain's revised F and R constants for predicting aromatic substituent effects //The Journal of Organic Chemistry. – 1984. – Vol. 49. – №. 11. – P. 1989-1992.

216. Hoefnagel A. J., Oosterbeek W., Wepster B. M. Substituent effects. 10. Critique of the "improved evaluation of field and resonance effects" proposed by Swain et al //The Journal of Organic Chemistry. – 1984. – Vol. 49. – №. 11. – P. 1993-1997.
217. Charton M. The validity of the revised F and R electrical effect substituent parameters //The Journal of Organic Chemistry. – 1984. – Vol. 49. – №. 11. – P. 1997-2001.
218. Hansch C., Leo A., Taft R. W. A survey of Hammett substituent constants and resonance and field parameters // Chemical Reviews. – 1991. – Vol. 91. – №. 2. – P. 165-195.
219. Unger S. H., Hansch C. Model building in structure-activity relations. Reexamination of adrenergic blocking activity of beta-halo-beta-arylalkylamines // Journal of medicinal chemistry. – 1973. – Vol. 16. – №. 7. – P. 745-749.
220. Taft R. W. Separation of polar, steric and resonance effects in reactivity //Steric effects in organic chemistry. – 1956. – P. 556-675.
221. Silipo, C., and Vittoria, A., Farmaco, Ed. Sci. 34, 858-868 (1979).
222. Seydel, J. K., Drug Design, Vol. I, Ariens, E. J., ed., Academic Press, New York, 1971, p. 343-379.
223. Fauchere, J.-L., and Lauterwein, J., Quant. Struct.-Act. Relat. 4, 11 - 13 (1985).
224. De Benedetti P. G. Structure activity relationships and mechanism of action of antibacterial sulphanilamides and sulphones //Advances in Drug Research. – 1987. – Vol. 16. – P. 227-279.
225. Hetnarski B., O'Brien R. D. Charge-transfer constant. New substituent constant for structure-activity relations //Journal of medicinal chemistry. – 1975. – Vol. 18. – №. 1. – P. 29-33.

226. Foster R., Hyde R. M., Livingstone D. J. Substituent constant for drug design studies based on properties of organic electron donor-acceptor complexes //Journal of pharmaceutical sciences. – 1978. –Vol. 67. –№. 9. –P. 1310-1313.
227. Livingstone D. J., Hyde R. M., Foster R. Further study of an organic electron-donor-acceptor related substituent constant //EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY. – 1979. – Vol. 14. – №. 5. – P. 393-397.
228. Lien E. J., Liao R. C. H., Shinouda H. G. Quantitative structure - activity relationships and dipole moments of anticonvulsants and CNS depressants //Journal of pharmaceutical sciences. – 1979. –Vol. 68. –№. 4. –P. 463-465.
229. Lien E. J. et al. Use of dipole moment as a parameter in drug -receptor interaction and quantitative structure - activity relationship studies //Journal of pharmaceutical sciences. – 1982. –Vol. 71. –№. 6. –P. 641-655.
230. Li W. Y., Guo N. R., Lien E. J. Examination of the interrelationship between aliphatic group dipole moment and polar substituent constants //Journal of pharmaceutical sciences. – 1984. – Vol. 73. – №. 4. – P. 553-558.
231. Hansch,C.,Vittoria,A.,Silipo,C.,andJow,P. Y. C.,J. Med. Chem.18,546-548(1975).
232. Taft R. W. et al. Studies of hydrogen-bonded complex formation with p-fluorophenol. V. Linear free energy relationships with OH reference acids //Journal of the American Chemical Society. – 1969. – Vol. 91. – №. 17. – P. 4801-4808.
233. Mitsky J., Joris L., Taft R. W. Hydrogen-bonded complex formation with 5-fluoroindole. Applications of the pKHB scale //Journal of the American Chemical Society. – 1972. – Vol. 94. – №. 10. – P. 3442-3445.

234. Gerhards, J., and Mehler, E. L., ref. 78, p. 153- 161.
235. Dearden J. C., Nicholson R. M. The prediction of biodegradability by the use of quantitative structure–activity relationships: correlation of biological oxygen demand with atomic charge difference//Pest Management Science. – 1986. – Vol. 17. – №. 3. – P. 305-310.
236. Dearden, J. C., and Nicholson, R. M., ref. 80, p. 307-312.
237. Shusterman A. J. et al. Mutagenicity of dimethyl heteroaromatic triazenes in the Ames test: the role of hydrophobicity and electronic effects //Molecular pharmacology. – 1989. – Vol. 36. – №. 6. – P. 939-944.
238. Lopez de Compadre R. L. et al. LUMO energies and hydrophobicity as determinants of mutagenicity by nitroaromatic compounds in Salmonella typhimurium //Environmental and molecular mutagenesis. – 1990. – Vol. 15. – №. 1. – P. 44-55.
239. Hansch, C., ref. 83, p. 3- 10.
240. Debnath A. K. et al. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity //Journal of medicinal chemistry. – 1991. – Vol. 34. – №. 2. – P. 786-797.
241. a) Ebert C., Katritzky A. R., Musumarra G. Effects of the Heteroaromatic Moiety on the Antisecretory Activities of Heteroarylmethyl Heteroaryl Sulfides [1] //Quantitative Structure-Activity Relationships. – 1991. –Vol. 10. –№. 2. –P. 107-109.b) Caruso L., Musumarra G., Katritzky A. R. “Classical” and “Magnetic” Aromaticities as new Descriptors for Heteroaromatics in QSAR. Part 3 [1]. Principal Properties for Heteroaromatics // Molecular Informatics. – 1993. – Vol. 12. – №. 2. – P. 146-151.
242. Gasteiger J., Marsili M. Tetrahedron 36: 3219 //Guillen MD, Gasteiger J (1983) Tetrahedron. – 1980. – Vol. 39. – P. 1331.
243. Gasteiger, J., Hutchings, M. G., Marsili, M., and Saller, H., ref. 78, p. 90-97.

244. Gasteiger J. et al. A new treatment of chemical reactivity: Development of EROS, an expert system for reaction prediction and synthesis design // Organic Synthesis, Reactions and Mechanisms. – Springer, Berlin, Heidelberg, 1987. – P. 19-73.
245. Unger, S. H., and Hansch, C., Prog. Phys. Org. Chem. 12, 91-118 (1976).
246. Fujita, T., Pure & Appl. Chem. 50, 987-994 (1978).
247. Verloop, A., Hoogenstraaten, W., and Tipker, J., Drug Design, Vol. V11, Ariens, E. J., ed., Academic Press, New York, 1976, p. 165-207.
248. Verloop, A., The STERZMOL Approach to Drug Design, Marcel Dekker, New York, 1987.
249. Charton M. Steric effects. I. Esterification and acid-catalyzed hydrolysis of esters //Journal of the American Chemical Society. – 1975. – Vol. 97. – №. 6. – P. 1552-1556.
250. Charton M. Steric effects. 7. Additional V constants //The Journal of Organic Chemistry. – 1976. – Vol. 41. – №. 12. – P. 2217-2220.
251. Charton M. The epsilon steric parameter—definition and determination //Steric Effects in Drug Design. – Springer, Berlin, Heidelberg, 1983. –P. 57-91.
252. Hancock C. K., Mcyers E. A. and Yager B. J. // J. Am. Chem. – 1961. – Vol. 83.– P. 4211 - 4213.
253. Fujita, T., and Iwamura, H., Topics Curr. Chem. 114, 119- 157 (1983).
254. Austel, V., Kutter, E., and Kalbfleisch, W., Arzneim.-Forsch. (Drug Res.) 29,585-587 (1979).
255. Verloop A., Tipker J. Physical basis of sterimol and related steric constants // Pharmacochem Libr. – 1987. – Vol. 10. – P. 97-102.
256. Fujita, T., ref. 39, p. 497-560.
257. Graham J. D. P., Karrar M. A. Structure-action relations in N, N-dimethyl-2-halogenophenethylamines //Journal of medicinal chemistry. – 1963. – Vol. 6. – №. 2. – P. 103-107.

258. Hansch C., Lien E. J. An analysis of the structure-activity relationship in the adrenergic blocking activity of the β -haloalkylamines //Biochemical pharmacology. – 1968. – Vol. 17. – №. 5. – P. 709-720.
259. Cammarata A. Interrelation of the regression models used for structure-activity analyses //Journal of medicinal chemistry. – 1972. – Vol. 15. – №. 6. – P. 573-577.
260. Topliss J. G., Costello R. J. Chance correlations in structure-activity studies using multiple regression analysis //Journal of Medicinal Chemistry. – 1972. – Vol. 15. – №. 10. – P. 1066-1068.
261. Topliss J. G., Edwards R. P. Chance factors in studies of quantitative structure-activity relationships //Journal of Medicinal Chemistry. – 1979. – Vol. 22. – №. 10. – P. 1238-1244.
262. Jha, T., Debnath, A. K., Mazumdar, A., Sengupta, C., and De, A. U., Ind. J. Chem. 25 B,169- I74 (1986).
263. Pleiss, M. A., and Unger, S. H., ref. 39, p. 561 -587.
264. Cammarata A. et al. Cautions regarding the physical interpretation of statistically based structure-activity relationships //Journal of pharmaceutical sciences. – 1970. – Vol. 59. – №. 10. – P. 1496-1499.
265. Craig P. N. et al. Minimal statistical data for structure-function correlations //Journal of medicinal chemistry. – 1971. – Vol. 14. – №. 5. – P. 447-447.
266. Purcell W. P. Some Guidelines for Application of Quantitative Structure-Activity-Relationships in Drug Design //European Journal of Medicinal Chemistry. – 1975. – Vol. 10. – №. 4. – P. 335-339.
267. Kubinyi, H., ref. 39, p. 589-643.
268. Kubinyi H. Free Wilson analysis. Theory, applications and its relationship to Hansch analysis //Molecular Informatics. – 1988. – Vol. 7. – №. 3. – P. 121-133.

269. Kubinyi H., Kehrhahn O. H. Quantitative structure-activity relationships. 1. The modified Free-Wilson approach //Journal of medicinal chemistry. – 1976. – Vol. 19. – №. 5. – P. 578-586.
270. Duewer, D. L., J. Chemometrics4, 299-321 (1990).
271. Hansch, C., Silipo, C., and Steller, E. E., J. Plzarm. Sci.64, 1 186- I191 (1975).
272. Kelder, J., and Greven, H. M., Rec. Tr-au. Chim. Pays-Bus 98, 168- 172 (1979).
273. Mager P. P. et al. Substituent constants for structure–activity relationships in pharmacochemistry //Sci Pharma. – 1979. – Vol. 47. – P. 265-297.
274. Mager, P. P., Med. Res. Ret. 3, 435-498 (1983).
275. Mager P. P. Reduced Free-Wilson approach applied to HSV-1 antiviral 7-amino-quinolines // Die Pharmazie. – 1990. – Vol. 45. – №. 5. – P. 359-360.
276. Darvas, F., Rohricht, J., Budai, Z., and_Bordas, B., Aduunces in Pharmacological Researchand Practice (Proceedings of the 3rd Congress of the Hungarian Pharmacological Society,Budapest, 1979), Vol. 3, Knoll, J., ed., Pergamon Press, 1980, p. 25-38.
277. Dubois J.-E., Laurent, D., and ArandaA.J., Chim. Phy.~. 70, 1608-1615, 1616-1624(1973).
278. Aranda, A., Compt. Rend. Hebd. Seances Acad. Sci. Paris, Ser. C 276, 1301-1304(1973).
279. Dubois, J.-E., Laurent, D., Bost, P., Chambaud, S., and Mercier, C., Eur. J . Med. C'hem.
280. Duperray, B., Chastrette, M., Makabeh, M. C., and Pacheco, H., Eur. J . Mrd. Clzcm. 11,225-236 (1976).
281. Mercier C., Dubois J. E. Comparison of molecular connectivity and Darc-Pelco methods-performance in anti-microbial, halogenated phenol QSARS //European Journal of Medicinal Chemistry. – 1979. – Vol. 14. – №. 5. – P. 415-423.

282. Dubois J. E., Mercier C., Panaye A. DARC TOPOLOGICAL SYSTEM AND COMPUTER-AIDED-DESIGN //ACTA PHARMACEUTICA JUGOSLAVICA. – 1986. – Vol. 36. – №. 2. – P. 135-169.
283. Hall L. H., Kier L. B. COMPARATIVE ANALYSIS OF MOLECULAR CONNECTIVITY, HANSCH, FREE-WILSON AND DARC-PELCO METHODS IN SAR OF HALOGENATED PHENOLS //EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY. – 1978. – Vol. 13. – №. 1. – P. 89-92.
284. Dunn III W. J., and Wold, S., ref. 39, p. 691-714.
285. Fukunaga, K., Introduction to Statistical Pattern Recognition, Academic Press, New York, 1972.
286. Varmuza, K., Pattern Recognition in Chemistry (Lecture Notes in Chemistry 21), Springer-Verlag, Berlin, 1980.
287. Kowalski B. R., Bender C. F. Pattern recognition. Powerful approach to interpreting chemical data //Journal of the American Chemical Society. – 1972. – Vol. 94. – №. 16. – P. 5632-5639.
288. Stuper A. J., Jurs P. C. Classification of psychotropic drugs as sedatives or tranquilizers using pattern recognition techniques //Journal of the American Chemical Society. – 1975. – Vol. 97. – №. 1. – P. 182-187.
289. Cammarata A., Menon G. K. Pattern recognition. Classification of therapeutic agents according to pharmacophores //Journal of medicinal chemistry. – 1976. – Vol. 19. – №. 6. – P. 739-748.
290. Menon G. K., Cammarata A. Pattern recognition II: Investigation of structure—activity relationships //Journal of pharmaceutical sciences. – 1977. – Vol. 66. – №. 3. – P. 304-314.
291. Wijnne, H. J., ref. 70, p. 211-229.
292. Dunn III W. J., Wold S., Martin Y. C. Structure-activity study of beta-adrenergic agents using the SIMCA method of pattern recognition // Journal of medicinal chemistry. – 1978. – Vol. 21. – №. 9. – P. 922-930.

293. Stuper A. J., Jurs P. C. Structure-activity studies of barbiturates using pattern recognition techniques //Journal of pharmaceutical sciences. – 1978. –Vol. 67. –№. 6. –P. 745-751.
294. Jurs, P. C., Chou, J. T., and Yuan, M., ref. 72, p. 103-129.
295. Jurs P. C., Chou J. T., Yuan M. Computer-assisted structure-activity studies of chemical carcinogens. A heterogeneous data set //Journal of medicinal chemistry. – 1979. – Vol. 22. – №. 5. – P. 476-483.
296. Dunn III W. J., Wold S. Structure-activity analyzed by pattern recognition: the asymmetric case //Journal of medicinal chemistry. – 1980. – Vol. 23. – №. 6. – P. 595-599.
297. Rose S. L., Jurs P. C. Computer-assisted studies of structure-activity relationships of N-nitroso compounds using pattern recognition //Journal of medicinal chemistry. – 1982. – Vol. 25. – №. 7. – P. 769-776.
298. Jurs P. C. et al. Computer-assisted studies of molecular structure biological activity relationships //Journal of chemical information and computer sciences. – 1985. – Vol. 25. – №. 3. – P. 296-308.
299. a) Livingstone, D. J., Methods Enzymol. 203, 613-638 (1991); b) Miyashita, Y., Li, Z., and Sasaki, S.-I., Trends Anal. Chem. 12, 50-60 (1993).
300. Mathews R. J. Structure-activity correlations obtained using pattern recognition methods //Journal of the American Chemical Society. – 1975. – Vol. 97. – №. 4. – P. 935-936.
301. Craig, P. N., ref. 39, p. 645-666.
302. Enslein, K., and Craig, P. N., J. Environ. Path. Toxicol. 2, 115- 121 (1978).

303. Enslein K., Craig P. N. Carcinogenesis: A predictive structure-activity model //Journal of Toxicology and Environmental Health, Part A Current Issues. – 1982. – Vol. 10. - №. 4-5. –P. 521-530.
304. Enslein K., Lander T. R., Strange J. R. Teratogenesis: A statistical structure-activity model //Teratogenesis, Carcinogenesis, and Mutagenesis. – 1983. –Vol. 3. – №. 3. –P. 289-309.
305. Enslein, K., Blake, B. W., Tomb, M. E., and Borgstedt, H. H, I n Mtro Toxicol. 1, 33-44 (1986).
306. Klopman G., Raychaudhury C. A novel approach to the use of graph theory in structure–activity relationship studies. Application to the qualitative evaluation of mutagenicity in a series of nonfused ring aromatic compounds //Journal of computational chemistry. – 1988. – Vol. 9. – №. 3. – P. 232-243.
307. Blake B. W. et al. Salmonella mutagenicity and rodent carcinogenicity: quantitative structure-activity relationships //Mutation Research/Genetic Toxicology. – 1990. – Vol. 241. – №. 3. – P. 261-271.
308. Gombar V. K. et al. A QSAR model of teratogenesis //Molecular Informatics. – 1991. – Vol. 10. – №. 4. – P. 306-332.
309. Tennant R. W. et al. Prediction of the outcome of rodent carcinogenicity bioassays currently being conducted on 44 chemicals by the National Toxicology Program //Mutagenesis. – 1990. – Vol. 5. – №. 1. – P. 3-14.
310. Bristol, D. W., personal comunication; Workshop “Predicting Chemical Carcinogenesis inRodents”, National Institute of Environmental Health, Research Triangle Park, NC, U. S. A.May 1993.

311. Klopman G. Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules //Journal of the American Chemical Society. – 1984. – Vol. 106. – №. 24. – P. 7315-7321.
312. Klopman G., Rosenkranz H. S. Structural requirements for the mutagenicity of environmental nitroarenes //Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. – 1984. – Vol. 126. – №. 3. – P. 227-238.
313. Klopman G. Predicting toxicity through a computer automated structure evaluation program //Environmental Health Perspectives. – 1985. – Vol. 61. – P. 269.
314. Klopman, G., Frierson, M. R., and Rosenkranz, H. S., Environ. Mutagen. 7, 625-644 (1985).
315. Frierson, M. R., Klopman, G., and Rosenkranz, H. S., Environ. Mutagen. 8,283-327 (1986).
316. Klopman G., Macina O. T. Computer-automated structure evaluation of antileukemic 9-anilinoacridines //Molecular pharmacology. – 1987. – Vol. 31. – №. 4. – P. 457-476.
317. Klopman, G., Kalos, A. N., and Rosenkranz, H. S., Mol. Toxicol. 1, 61-81 (1987).
318. Rosenkranz H. S., Klopman G. CASE, the computer-automated structure evaluation method, correctly predicts the low mutagenicity for Salmonella of nitrated cyclopenta-fused polycyclic aromatic hydrocarbons //Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. – 1988. – Vol. 199. – №. 1. – P. 95-101.
319. Rosenkranz H. S., Klopman G. CASE, the computer-automated structure evaluation system, as an alternative to extensive animal testing //Toxicology and industrial health. – 1988. – Vol. 4. – №. 4. – P. 533-540.
320. Klopman, G., and Dimayuga, M. L., J. Comput.-Aided Mol. Des. 4, 117- 130 (1990).

321. a) Klopman G. MULTICASE 1. A hierarchical computer automated structure evaluation program // *Molecular Informatics*. – 1992. – Vol. 11. – №. 2. – P. 176-184.;
b) Klopman, G., and Ptchelintsev, *Med. Chem. Res.* 3, 67-77 (1993).
322. King, R. D., Muggleton, S., Lewis, R. A., and Sternberg, M. J. E., *Proc. Natl. Acad. Sci. U S A* 89, 11322-11326 (1992).
323. Mercier, C., Troullier, G., and Dubois, J.-E., *Quant. Struct.-Act. Relat.* 9, 88-93 (1990).
324. Huebel S., Roesner T., Franke R. The evaluation of topological pharmacophores by heuristic approach // *Die Pharmazie*. – 1980. – Vol. 35. – №. 7. – P. 424-433.
325. Franke R., Huebel S., Streich W. J. Substructural QSAR approaches and topological pharmacophores // *Environmental health perspectives*. – 1985. – Vol. 61. – P. 239.
326. Streich W. J., Franke R. Topological Pharmacophores New Methods and Their Application to a Set of Antimalarials Part 1: The Methods LOGANA and LOCON // *Molecular Informatics*. – 1985. – Vol. 4. – №. 1. – P. 13-18.
327. Franke, R., and Streich, W. J., *Quant. Struct.-Act. Relat.* 4, 51-63, 63-69 (1985).
328. Hubel, S., and Franke, R., *ibid.* 83, p. 177- 180.
329. Simon Z., Szabadai Z. MINIMAL STERIC DIFFERENCE PARAMETER AND IMPORTANCE OF STERIC FIT FOR STRUCTURE-BIOLOGICAL ACTIVITY CORRELATIONS // *Studia Biophysica*. – 1973. – Vol. 39. – №. 2. – P. 123-132.
330. Simon Z. Specific interactions. Intermolecular forces, steric requirements, and molecular size // *Angewandte Chemie International Edition*. – 1974. – Vol. 13. – №. 11. – P. 719-727.
331. Simon Z. Minimum steric difference: the MTD method for QSAR studies. – Research Studies Press, 1984. – Vol. 7.

332. Simon Z. SUBSTRATE-SPECIFICITY IN ENZYMATIC-REACTIONS CORRELATED WITH MINIMAL STERIC DIFFERENCES FOR SUBSTRATE MOLECULES //STUDIA BIOPHYSICA. – 1975. – Vol. 51. – №. 1. – P. 49-57.
333. Motoc I. et al. MINIMAL STERIC DIFFERENCE COMPUTED AS NONOVERLAPPING VOLUMES CORRELATION WITH ENZYMATIC-HYDROLYSIS OF RIBONUCLEOSIDES //STUDIA BIOPHYSICA. – 1977. – Vol. 66. – №. 1. – P. 75-78.
334. Motoc I. Molecular shape descriptors //Steric effects in drug design. – Springer, Berlin, Heidelberg, 1983. – P. 93-105.
335. Simon Z. et al. RECEPTOR-SITE MAPPING-SEARCH STRATEGY OF STANDARD FOR CORRELATIONS WITH MINIMAL STERIC DIFFERENCES //studia biophysica. – 1976. – Vol. 55. – №. 3. – P. 217-226.
336. Simon, Z., Holban, S., Motoc, I., Mrdcec, M., Chiriac, C., Kerek, F., Ciubotariu, D., Szabadai, Z., Pop, R. D., and Schwartz, I., Stud. Biophys. 59, 181- 197 (1976).
337. Simon Z., Badilescu I., Racovitan T. Mapping of dihydrofolate-reductase receptor site by correlation with minimal topological (steric) differences //Journal of theoretical biology. – 1977. – Vol. 66. – №. 3. – P. 485-495.
338. Popoviciu V. et al. STERIC MAPPING OF ESTROGENIC RECEPTOR-SITE BY A MINIMAL STERIC DIFFERENCE METHOD //STUDIA BIOPHYSICA. – 1978. – Vol. 69. – №. 1. – P. 75-76.
339. Simon, Z., and Bohl, M., Quant. Struct.-Act. Relat. 11, 23-28 (1992).
340. Motoc, I., Quant. Struct.-Act. Relat. 3, 43-47, 47-51 (1984).
341. Magee, P. S., ref. 83, p. 549-552.
342. Magee P. S. A new Approach to Active-Site Binding Analysis. Inhibitors of Acetylcholinesterase //Molecular Informatics. – 1990. – Vol. 9. – №.3. – P. 202-215. Hopfinger A. J. A QSAR investigation of dihydrofolate reductase inhibition by

- Baker triazines based upon molecular shape analysis //Journal of the American Chemical Society. – 1980. – Vol. 102. – №. 24. – P. 7196-7206.
343. Battershell C., Malhotra D., Hopfinger A. J. Inhibition of dihydrofolate reductase: Structure-activity correlations of quinazolines based upon molecular shape analysis //Journal of medicinal chemistry. – 1981. – Vol. 24. – №. 7. – P. 812-818.
344. Hopfinger A. J. Inhibition of dihydrofolate reductase: structure-activity correlations of 2, 4-diamino-5-benzylpyrimidines based upon molecular shape analysis //Journal of medicinal chemistry. – 1981. – Vol. 24. – №. 7. – P. 818-822.
345. Hopfinger A. J. A general QSAR for dihydrofolate reductase inhibition by 2, 4-diaminotriazines based upon molecular shape analysis //Archives of biochemistry and biophysics. – 1981. – Vol. 206. – №. 1. – P. 153-163.
346. Hopfinger A. J., Potenzzone R. Ames test and antitumor activity of 1-(X-phenyl)-3, 3-dialkyltriazenes. Quantitative structure-activity studies based upon molecular shape analysis //Molecular pharmacology. – 1982. – Vol. 21. – №. 1. – P. 187-195.
347. Hopfinger A. J. Theory and analysis of molecular potential energy fields in molecular shape analysis: A QSAR study of 2, 4-diamino-5-benzylpyrimidines as DHFR inhibitors //J. Med. Chem. – 1983. – Vol. 26. – P. 990-996.
348. Hopfinger, A. J., Quunt. Struct.-Act. Relac. 3, 1--5 (1984).
349. Mabilia M., Pearlstein R. A., Hopfinger A. J. Molecular shape analysis and energetics-based intermolecular modelling of benzylpyrimidine dihydrofolate reductase inhibitors //European journal of medicinal chemistry. – 1985. – Vol. 20. – №. 2. – P. 163-174.
350. Walters, D. E., and Hopfinger, A. J., J . Mol. Struct. (Theochem) 134, 317-323 (1986).
351. Hopfinger, A. J., and Burke, B. J., ref. 82, 151- 159.
352. Hopfinger, A. J., and Burke, B. J., ref. 63, p. 173-209.

353. Aoyama T., Suzuki Y., Ichikawa H. Neural networks applied to structure-activity relationships //Journal of medicinal chemistry. – 1990. – Vol. 33. – №. 3. – P. 905-908.
354. Andrea, T. A., and Kalayeh, H., ref. 83, p. 209-212.
355. Rose, V. S., MacFie, H. J. H., and Croall, I. F., ref. 83, p. 213-216.
356. Rose V. S., Croall I. F., Macfie H. J. H. An Application of Unsupervised Neural Network Methodology Kohonen Topology-Preserving Mapping to QSAR Analysis //Molecular Informatics. – 1991. – Vol. 10. –№.1. – P. 6-15.
357. Wiese M. Application of neural networks in the analysis of percent effect biological data //Quantitative Structure-Activity Relationships. – 1991. – Vol. 10. – №.4. –P. 369-371.
358. Andrea T. A., Kalayeh H. Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors //Journal of Medicinal Chemistry. – 1991. – T. 34. – №. 9. – C. 2824-2836.
359. Aoyama, T., and Ichikawa, H., Chem. Pharm. Bull. 39, 358-366, 372-378 (1991).
360. Chastrette M., De Saint Laumer J. Y. Structure-odor relationships using neural networks //European journal of medicinal chemistry. – 1991. – Vol. 26. – №. 8. – P. 829-833.
361. Livingstone D. J., Hesketh G., Clayworth D. Novel method for the display of multivariate data using neural networks //Journal of molecular graphics. – 1991. – Vol. 9. – №. 2. – P. 115-118.
362. Livingstone D. J., Salt D. W. Regression analysis for QSAR using neural networks //Bioorganic & medicinal chemistry letters. – 1992. – Vol. 2. – №. 3. – P. 213-218.
363. Salt D. W. et al. The use of artificial neural networks in QSAR //Pest Management Science. – 1992. – Vol. 36. – №. 2. – P. 161-170.

364. a)Manallack, D. T., and Livingstone, D. J., *Med. Chem. Res.* 2,181- 190 (1992);b) Manallack,D. T., and Livingstone, D. J., ref. 84, p. 128- 131.
365. Livingstone, D. J., and Manallack, D. T., *J . Med. Chem.* 36, 1295- 1297 (1992).