**ГОУ ВПО «Пятигорская государственная фармацевтическая академия Федерального агентства по здравоохранению и социальному развитию»**

**Научно-популярная лекция**
**на тему:**

# ИСТОРИЯ И ПЕРСПЕКТИВЫ КОМПЬЮТЕРНОГО КОНСТРУИРОВАНИЯ ЛЕКАРСТВ

**Автор:**
*преподаватель кафедры физической и коллоидной химии ПятГФА*
*к.ф.н. Глушко А.А.*

**Пятигорск 2010 г.**

## 1. Введение

Если спросить двух ученых - химика и физика о том, что такое молекула, они могут дать два совсем разных ответа.

Что такое молекула **с точки зрения химии** мы с Вами хорошо знаем – это группа атомов разного сорта, связанных химическими связями различного типа.

У специалиста в области физики представление о молекулярной системе принципиально иное. Физическая модель молекулы (Рис. 1) – это единая система ядер атомов и электронов, при этом все они взаимодействуют друг с другом электростатически. Каждый электрон в молекуле взаимодействует со всеми другими электронами в молекуле и со всеми ядрами атомов. Каждое ядро атома взаимодействует со всеми ядрами атомов и с каждым из электронов в молекуле.

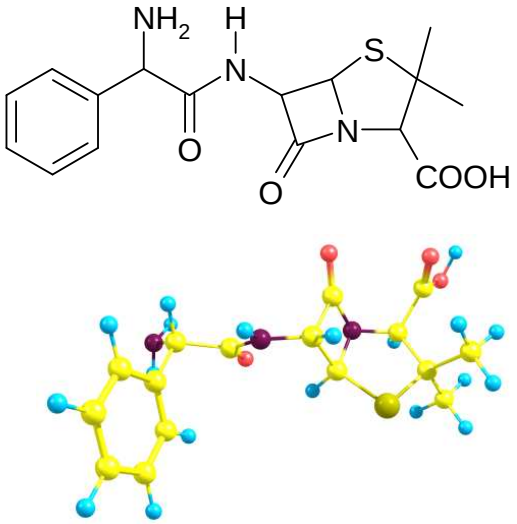



**Рисунок 1. Способы представления молекулярной структуры.**

## 2. История компьютерной химии

Построение моделей молекул может на первый взгляд показаться довольно простой задачей. Кажущаяся простота данной концепции обусловлена тем, что наше мышление базируется на законах классической механики. Однако в начале XX столетия по мере открытия новых фактов в физике становилась все очевиднее неприменимость законов классической механики Ньютона для описания процессов, происходящих в масштабах молекул, атомов и элементарных частиц на расстояниях порядка нанометров и менее.

При попытке применить классическую механику для описания структуры атома возник ряд парадоксальных вопросов[2]:

1. Почему электроны не падают на ядро атома, теряя энергию в виде электромагнитного излучения? Ведь в соответствии с законами электродинамики заряженное тело при ускоренном движении излучает электромагнитные волны. Неужели электроны в атоме неподвижны?
2. Почему при нагревании вещество начинает светиться, и цвет свечения зависит от температуры?
3. Почему излучение происходит квантами строго фиксированной величины, а не произвольными порциями?

В поисках ответов на эти и многие другие вопросы в начале XX века ученые пришли к новой и оригинальной физической теории, получившей название «квантовая механика». Элементарная частица с точки зрения квантовой механики имеет свойства волны, а её поведение описывается не траекторией, а так называемой волновой функцией. Результат взаимодействия элементарных частиц не предсказуем. Причем непредсказуемость поведения частиц не является недостатком теории, а есть фундаментальное свойство материи. При переходе от квантовых систем к более сложным, состоящим из большого числа частиц, все ярче проявляются статистические закономерности и мы наблюдаем привычное нашему сознанию предсказуемое поведение материи, когда причина однозначно определяет следствие. Квантовая механика, таким образом, включает в себя классическую механику как частный случай для большого числа частиц.

Главная сложность практического применения законов квантовой механики для моделирования молекул, заключается в неопределенности координат электронов и непредсказуемости их поведения.

**В 1924 году** французский физик Луи де Бройль высказал гипотезу о том, что установленный ранее для фотонов корпускулярно-волновой дуализм присущ всем частицам —электронам, протонам, атомам[3].

**27 января 1926 года** в редакцию немецкого физического журнала "Annalen der Physik" поступила первая из целой серии работ австрийского физика Эрвина Шредингера (Рис. 2), в которой, содержалась последовательная теория, придававшая квантовой механике наглядность и физический смысл [3]. Это была так называемая волновая механика. Эта новая – более общая механическая теория позволяла описывать непредсказуемые процессы, происходящие в микромире основываясь на представлениях Луи Де Бройля о волнах материи, так называемых волновых функциях. Например, для электронов квадрат амплитуды этих волн имеет понятный физический смысл – электронная плотность.

**Рисунок 2. Эрвин Шрёдингер (1887 – 1961г.)**

В 1929г. Шредингеру была присуждена Нобелевская премия по физике. В книге Э.Шредингера «Что такое жизнь с точки зрения физика?» (1945) впервые была обоснована мысль о молекулярной природе наследственности и возможности физико-химического истолкования жизни. Уже тогда ученые понимали, что это путь к познанию, того, что происходит в атомах, малых молекулах, белках, ДНК и более того молекулярной основы живых систем. Однако в то время еще не были созданы столь мощные вычислительные машины, необходимые для

проверки этой теории и это была лишь теория на бумаге, которую не было возможности проверить на практике.

В начале 50-х годов XX века был совершен еще один прорыв, который изменил мир – создание первых электронных вычислительных машин. На заре компьютерной эпохи машины позволяли моделировать лишь очень простые молекулы с невысокой степенью точности. Но уже тогда стало ясно, что в обозримом будущем химия получит новый эффективный инструмент и возникнут новые ветви науки – квантовая химия и вычислительная химия.

При описании экспериментальных исследований широко используются два термина *in vitro* (в пробирке), *in vivo* (в живом организме). С появлением вычислительной химии возник еще один вид эксперимента - *in silico* (эксперимент в компьютере).

Производительность ЭВМ росла лавинообразно. Стали появляться компьютерные программы, специально предназначенные для вычисления структуры и свойств молекул.

Первые попытки моделирования пространственной структуры белковых молекул были предприняты еще в конце 1960-х годов. Первые модели структуры белков конструировали из большого числа шариков, проволочек, втулок, винтиков и других деталей (Рис. 3). Они были очень громоздки, хрупки и требовали огромного времени и усердия для изготовления, даже при использовании специальных «конструкторов» — наборов стандартных деталей для сборки[8]. В настоящее время компьютеры почти полностью заменили такие конструкторы (Рис. 4).

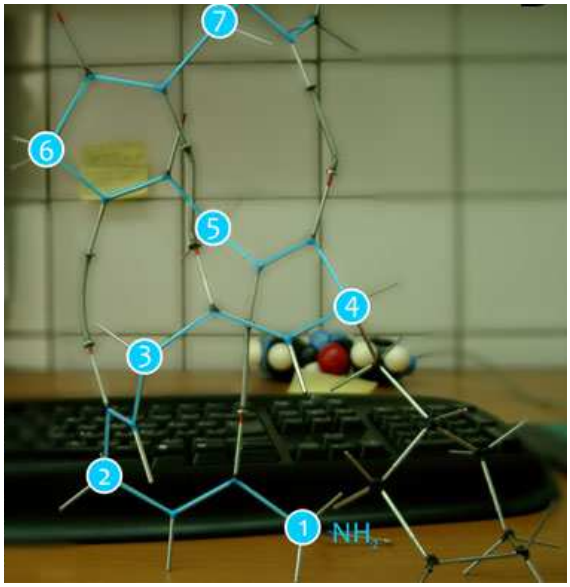
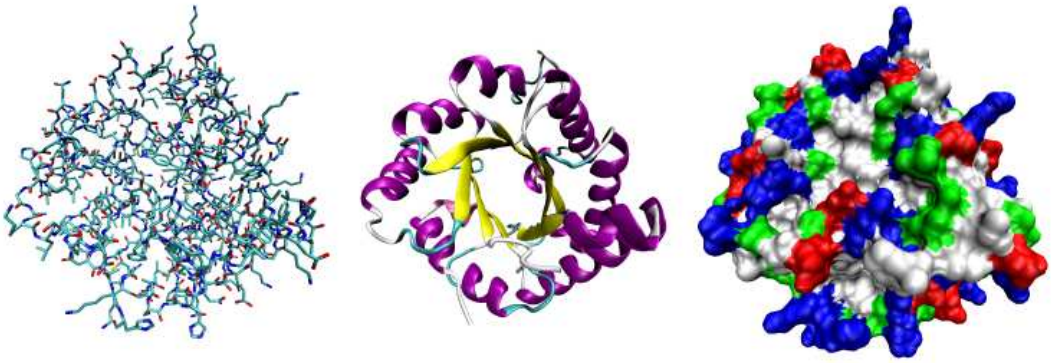



**Рисунок 3. Участок α-спирали, построенный из «проволочного» конструктора.**



А Б В

**Рисунок 4. Компьютерные изображения структуры белка. А. Проволочная графика. Б. Лента. В. Поверхность доступная растворителю окрашенная условными цветами аминокислот.**

## 3. Современные направления в биофизике и поиск новых биологически активных веществ

Если мы спросим двух ученых - химика и физика о том, что такое живой организм, то столкнемся с двумя принципиально различными мнениями.

С точки зрения биохимии – живой организм очень сложная система химических реакций, катализируемых ферментами под управлением ДНК.

Внимание биофизики сфокусировано, прежде всего, на движении и взаимодействии отдельных частей молекулярных систем. Поэтому с точки зрения физики – живой организм есть сложная система атомов и молекул, а также элементарных частиц, взаимодействующих между собой.

Одна из основных практических задач, которая объединяет биофизику и биохимию – это изучение глубинных механизмов функционирования живых организмов на различных уровнях организации с целью поиска путей сохранения жизни. Поэтому создание новых лекарств – это всегда междисциплинарное научное исследование.

Основная проблема в процессе поиска лекарственных веществ – это выбор объектов исследования в огромном многообразии химических структур.

В настоящее время значительный массив экспериментальных данных и фундаментальных знаний, накопленных учеными о живых системах, используется для направленного синтеза новых лекарств [17].

С возникновением теории, позволяющей описывать поведение атомов и молекул, появилась возможность создавать математические модели белков, нуклеиновых кислот, моделировать ферментативные реакции, с помощью компьютерных расчетов определять – как именно будет взаимодействовать с биохимическими системами вещество, попавшее в организм извне.

По мере того как методы моделирования молекул совершенствуются, а вычислительная мощность компьютеров увеличивается, все шире внедряются вычислительные методы в процесс создания новых лекарств.

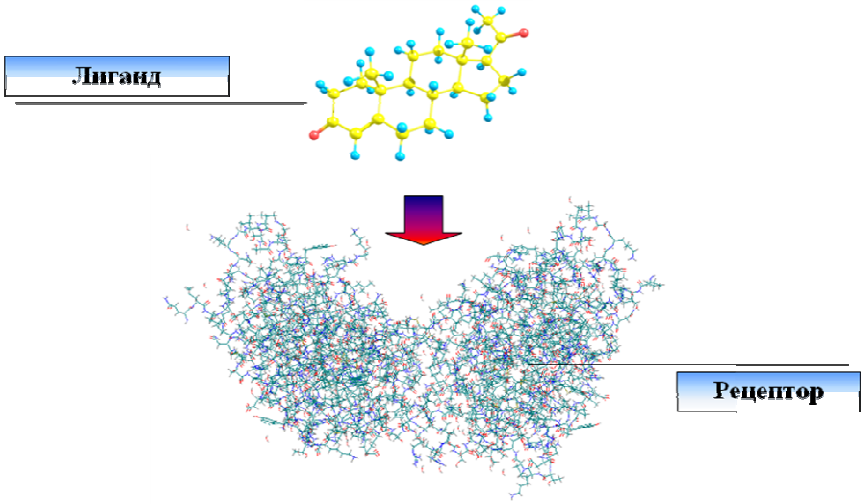



**Рисунок 5. На рисунке изображена трехмерная модель рецептора и модель молекулы лиганда.**

Создание лекарства (*drug design*) требует от исследователя глубокого понимания молекулярного механизма действия активного вещества. В основе конструирования лекарств лежит модель «ЛИГАНД-РЕЦЕПТОР» (Рис. 5). В качестве «ЛИГАНДА» рассматривают низкомолекулярное вещество (лекарство), а в качестве «РЕЦЕТОРА» - высокомолекулярное вещество (биомолекула) рецептор, ионный канал, фермент, ДНК, РНК.

Лиганд образует комплекс с рецептором и, если его структура соответствует рецептору комплементарно (как ключ к замку), то рецептор, изменяя свою конформацию определенным образом, реагирует на присоединение лиганда, вызывая цепь биохимических процессов, приводящих к некоторому биологическому ответу (понижение артериального давления, усиление сокращений сердечной мышцы, снотворное действие и др.).

В случае, если пространственная молекулярная структура рецептора известна, возникает возможность проведения молекулярного докинга - компьютерного моделирования взаимодействия лиганда с рецептором и определения прочности их связывания. Для многих рецепторов (н-холинорецептор, родопсин) и ферментов (циклооксигеназа) в настоящее время определены пространственные структуры методами рентгеноструктурного анализа (РСА)[9]. Эти структуры доступны в виде файлов формата pdb (Protein Data Base) в базе данных белковых структур Protein Data Base [16], представленной в сети Интернет в свободном доступе.

Однако для изучения структуры макромолекулы методом РСА необходимо выделить данное вещество в кристаллическом виде. Для рецепторов, встроенных в клеточную мембрану и содержащихся в организме в относительно низких концентрациях этого сделать до сих пор не удалось, а следовательно нет и экспериментальных данных об их пространственной структуре. Среди таких рецепторов большинство G-протеин ассоциированных рецепторов (адренорецепторы, дофаминовые рецепторы и др.)

В настоящее время геном человека полностью прочитан. База данных GenBank и визуальная карта генома человека представлены в сети Интернет в свободном доступе на сайте Национального центра биотехнологической информации США [15] (Рис. 6).

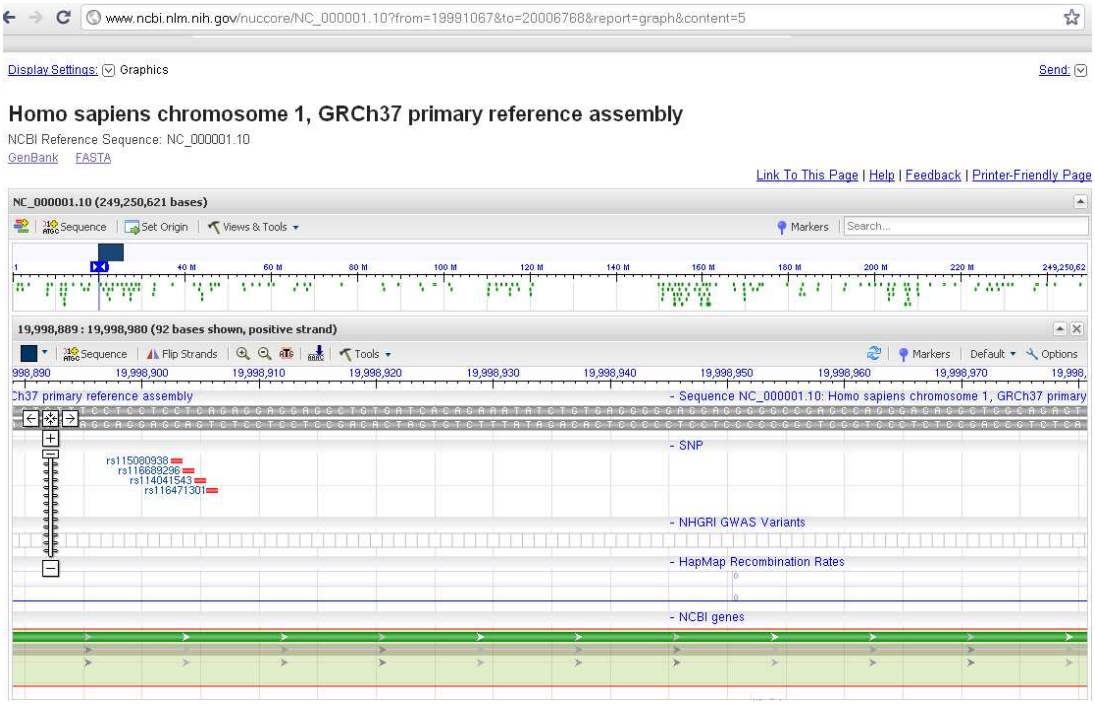



**Рисунок 6. Часть схемы нуклеотидной последовательности кодирующей серотониновый рецептор человека (5-HT$_6$), представленной на сайте Национального центра биотехнологической информации США.**

В базе данных GenBank представлены также геномы многих других организмов. По результатам проекта «Геном человека» в геноме человека содержится около 30000 генов. Это лишь небольшая доля информации, содержащейся в ДНК человека (около 1,5%), которая кодирует белки и РНК. Остальная (большая) часть генетической информации представляет собой малоизученную «темную материю». На основе карты генома можно узнать аминокислотные последовательности записанных в ней структур белков. Однако аминокислотная последовательность (первичная структура белка) еще не дает представления о третичной и четвертичной структуре, а о том какую функцию выполняет данный белок можно лишь догадываться.

Для определения пространственных структур и функций этих неизвестных белков в настоящее время ведутся активные исследования. Основной метод – виртуальный фолдинг белка (Рис. 7). Фолдинг – это процесс сворачивания белка в его нативную вторичную и далее в третичную структуру, который происходит самопроизвольно в процессе синтеза[11].

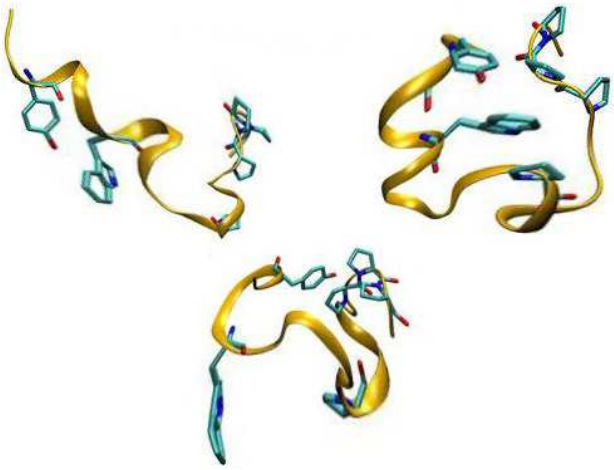


**Рисунок 7. Схема процесса фолдинга белка.**

Виртуальный фолдинг осуществляется с применением высокопроизводительных вычислительных систем на основе математической модели белка методом молекулярной динамики. Молекулярная динамика позволяет моделировать движение всех частей молекулы, наблюдать за эволюцией молекулярной системы во времени. В процессе моделирования молекулярной динамики применяют так называемое термостатирование – виртуальный термостат, алгоритм, следящий за тем, чтобы отдельные части молекулы слишком сильно не нагревались, а другие не охлаждались. Термостат при необходимости разгоняет атомы, движущиеся слишком медленно или, наоборот тормозит слишком быстрые, поддерживая заданную температуру системы.

В процессе фолдинга применяют так называемую имитацию отжига. Сначала динамику белка моделируют при более высоких температурах, при этом для макромолекулы становятся доступны множество разнообразных конформаций. Затем постепенно температуру системы снижают, и количество конформаций, занимаемых белком начинает уменьшаться. В результате такого медленного снижения температуры белок постепенно приходит к состоянию с наименьшей потенциальной энергией. С большой вероятностью это и есть его нативная структура. Однако такой подход не достаточно надежен, т.к. нет возможности проверить результат.

Для того, чтобы с большей достоверностью определить нативную структуру белка применяется метод фолдинга по гомологии [12]. В биохимии существует так называемая структурная классификация белков, которая в огромном многообразии белков выделяет структурные мотивы, группы родственных белков сходных по структуре. Если для одного из белков такой группы пространственная структура определена экспериментально, для других родственных белков есть возможность определить нативную структуру методом фолдинга по гомологии. Такой фолдинг отличается от простого тем, что моделирование динамики сворачивания белка направляется экспериментально определенной структурой – гомологом. И если в результате полученная структура исследуемого белка сходна со структурой гомолога, считают, что цель фолдинга достигнута. Так были определены структуры многих G-протеин ассоциированных рецепторов[18] (адренорецепторы, дофаминорецепторы, морфинановые и др.) на основе экспериментально изученной структуры бычьего родопсина – зрительного рецептора. Также с применением фолдинга по гомологии определена структура ГАМК-А рецептора человека[19]. Для этого в качестве гомолога

был использован Н-холинорецептор электрического ската. Оба этих рецептора являются ионными каналами и состоят из пяти субъединиц.

Пространственные структуры более 2000000 белков (данные на сентябрь 2010г.), полученные с применением виртуального фолдинга доступны в сети Интернет в базе данных Swiss-model[14]. Количество структур в этой базе данных постоянно увеличивается, т.к. фолдинг новых белков проводится в автоматическом режиме круглосуточно.

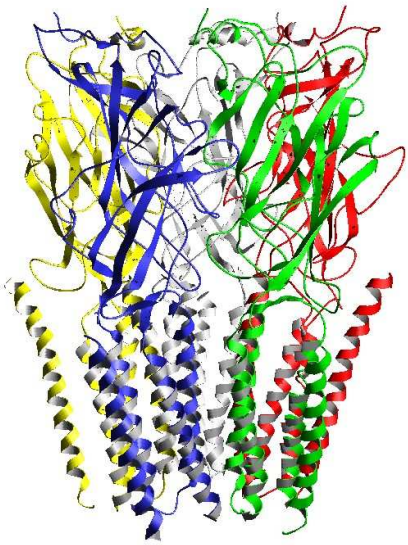


**Рисунок 8. Пространственная структура рецептора гамма-аминомасляной кислоты (ГАМК) А-типа человека, полученная методом фолдинга по гомологии с Н-холинорецептором электрического ската.**

Взаимодействие лекарственного вещества с биологической мишенью (рецептором) весьма структурно специфично. Малейшее изменение в структуре вещества может привести к потере его активности в отношении данной мишени.

В случае наличия информации о пространственном строении рецептора для теоретического определения биологической активности веществ используют метод молекулярного докинга – моделирования взаимодействия «лиганд-рецептор». Основная задача докинга – поиск наиболее устойчивых расположений лиганда в сайте связывания рецептора (Рис. 9).

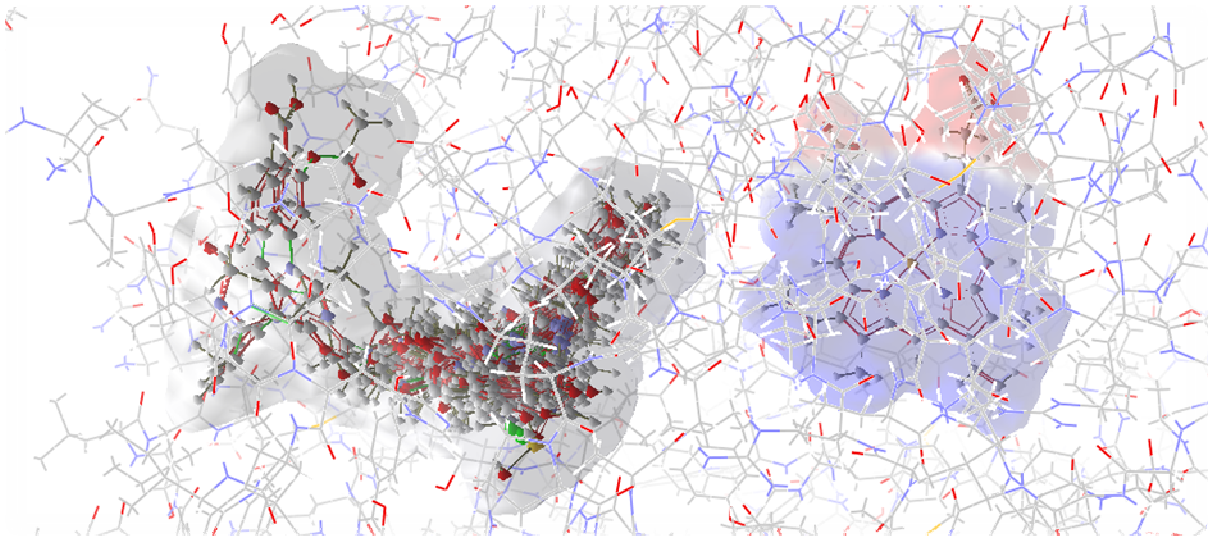


**Рисунок 9. Множество различных конформаций лиганда в активном центре фермента циклооксигеназы-2, полученных в результате гибкого молекулярного докинга программой Autodock 4.0.**

Существуют различные модификации докинга [10]. Наиболее точный, хотя и ресурсоемкий подход – моделирование динамики взаимодействия лиганда с рецептором. При этом в процессе вычислительного эксперимента записываются различные параметры образующегося комплекса, такие как энергия взаимодействия между лигандом и рецептором, диаметр поры ионного канала, дистанции между отдельными группами атомов. При этом исследователи выбирают такие параметры, которые являются индикаторами активации рецептора и следят за их изменением во времени.

Однако молекулярная динамика требует длительных расчетов. В случаях, когда требуется исследовать большое количество структур, применяют упрощенные процедуры молекулярного докинга: гибкий докинг и жесткий докинг. При жестком докинге рецептор и лиганд неподвижны, алгоритм лишь выбирает оптимальные с точки зрения энергии положения лиганда в сайте связывания. В процессе гибкого докинга лиганду и аминокислотам сайта связывания придается конформационная подвижность. Гибкий докинг значительно точнее жесткого, однако более ресурсоемок. Как показывает практика, в т.ч. и наши исследования, применение гибкого докинга дает удовлетворительные количественные результаты для рядов соединений со сходными структурами.

В случаях, когда не известна структура рецептора, существует ряд методов позволяющих теоретически прогнозировать биологическую активность путем определения сходства с эталонными веществами. При этом полагается, что свойства вещества являются функцией его структуры. Если исследуемое вещество обладает структурным сходством с веществом с известной биологической активностью, можно предположить, что и биологическая активность у этих веществ сходная. Однако при данном подходе перед исследователями всегда встает вопрос, о том каким образом сравнивать между собой структуры.

Если изобразить структуры двух молекул на бумаге в виде химических структурных формул, взгляд способен уловить в них элементы сходства (Рис 10). Наше сознание каким-то таинственным образом способно улавливать визуальное сходство двух объектов, в том числе и молекулярных структур. В настоящее время этот механизм описан на языке теории графов и воплощен в виде компьютерных программ, способных сравнивать между собой структурные формулы веществ. Результат сравнения может быть выражен количественно, например в процентах. Такой способ сравнения называется топологическим и применяется в программе PASS разработанной в НИИ БМХ РАМН [6]. Программа сравнивает исследуемое вещество с обширной базой данных веществ, для которых

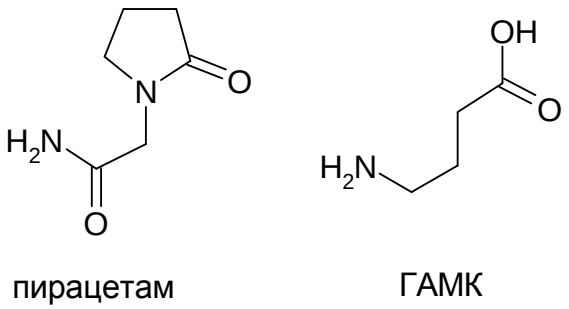



**Рисунок 10. Сходные по структуре и биологическим свойствам вещества: пирацетам и гамма-аминомасляная кислота (ГАМК).**

известно множество различных видов биологической активности. В результате получается спектр биологической активности для исследуемого вещества. Программа PASS позволяет за небольшое время произвести виртуальный скрининг большого количества низкомолекулярных веществ, т.к. сравнение таких молекул топологическим способом не требует длительных вычислений.

Результаты работы программы PASS можно использовать для предварительного широкого скрининга. Более точное сравнение химических структур требует рассмотрения молекулы с позиции квантовой физики.

Сравнение пространственных физических моделей молекул является более сложной задачей. Основная проблема при этом заключается в выборе взаимного расположения молекул, а также в выборе их конформаций. Ведь сочетаний конформаций двух молекул может быть очень много, особенно для сложных гибких молекул.

Для сравнения физических моделей молекул между собой используется сравнительный анализ молекулярного поля [13] (САМП), в английской аббревиатуре CoMFA (Comparative Molecular Field Analysis). В качестве молекулярного поля рассматривают распределение электронной плотности в пространстве, электростатический потенциал, потенциал Леннарда-Джонса, описывающий ванн-дер-ваальсово взаимодействие, а так же др. поля и различные их комбинации.

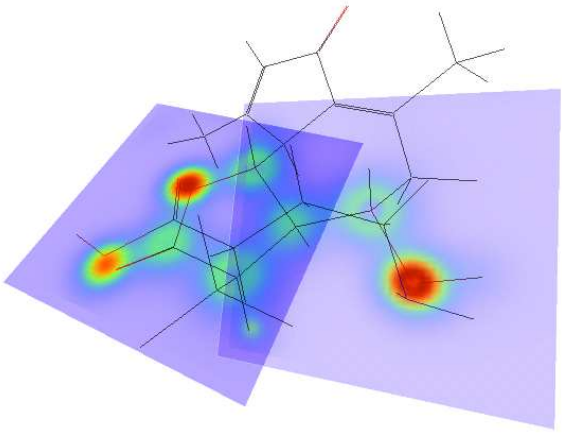


**Рисунок 11. Наложение пространственных структур молекул и их функций распределения электронной плотности в программе MSpace.**

В разработанной нами программе MSpace [1,4] реализован быстрый алгоритм вычисления сходства между функциями распределения электронной плотности молекул (Рис. 11). Проблема выбора взаиморасположения молекул решается с использованием метода молекулярной динамики. При этом в процессе моделирования сравниваемые молекулы приходят к конформациям, в которых их сходство максимально. Для этого не требуется перебирать большое количество конформаций, т.к. применяется метод имитации отжига. Т.е. в программе MSpace используется идея фолдинга по гомологии, только для низкомолекулярных веществ. Однако программа может быть использована и для фолдинга белков.

Еще один способ сравнения молекулярных структур – использование дескрипторов молекулярной структуры [7,5]. Дескриптором может быть любое свойство молекулы, которое можно выразить количественно, например теплота образования, молекулярная масса, дипольный момент, поляризуемость и т.п. Набор достаточно большого количества дескрипторов характеризует молекулу и является своего рода её уникальным «отпечатком

пальцев». Такие числовые наборы можно сравнивать между собой с помощью методов математической статистики.

Для наглядного примера использования дескрипторов давайте возьмем набор из нескольких сот структур лекарственных препаратов с различной биологической активностью и вычислим с помощью квантово-химического метода три свойства для каждого вещества: теплоту образования, дипольный момент и энергию сольватации в воде. Теперь построим трехмерную систему координат, где каждое из вычисленных нами свойств будет измерением. Расположим препараты в этом пространстве в виде точек и раскрасим каждую точку условным цветом в соответствии с биологической активностью. В результате мы сможем увидеть на пространственной диаграмме области разных цветов соответствующие различным видам активности (Рис. 12). Каждая область имеет свою достаточно сложную пространственную форму.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Name | *Bond MM2* | *Angle MM2* | *Dihedral MM2* | *VdwMM2* | *Stretch MM2* |
| 2 | Moclobemide-aurorix | 0,73 | 4,32 | -0,69 | 12,93 | 0,33 |
| 3 | Fluvoxamine | 0,96 | 4,63 | -5,29 | 11,71 | 0,25 |
| 4 | Paroxetine | 1,23 | 10,77 | -3,66 | 7,35 | 0,26 |
| 5 | Sertraline-zoloft | 1,04 | 3,15 | -8,74 | 11,61 | 0,18 |
| 6 | Sydnophenum | 0,59 | 9,30 | -5,29 | 3,75 | -0,17 |
| 7 | Sydnocarbum | 0,77 | 14,28 | -18,35 | 9,23 | 0,29 |
| 8 | Cytisinum | 0,85 | 2,44 | 9,18 | 11,11 | 0,17 |
| 9 | Aethylmorphini_hydroc | 1,64 | 14,09 | 3,96 | 11,02 | 0,56 |
| 10 | Morphinum | 2,91 | 20,01 | -1,49 | 11,61 | 0,74 |
| 11 | Nalorphini_hydrochloric | 4,33 | 40,88 | 10,06 | 25,02 | 0,54 |
| 12 | Piritramidum-dipidolor | 3,21 | 7,20 | -1,97 | 20,96 | 0,83 |
| 13 | Amidopyrinum | 1,01 | 21,60 | -2,49 | 2,61 | 0,00 |
| 14 | Analginum | 1,09 | 23,00 | -0,81 | 1,93 | -0,50 |
| 15 | Antipyrinum | 0,69 | 16,35 | 2,12 | 3,31 | -0,10 |
| 16 | Salicylamidum | 0,34 | 2,37 | 0,70 | 5,89 | 0,06 |
| 17 | Butadionum | 0,92 | 9,36 | -2,27 | 6,01 | -0,05 |
| 18 | Domperidone | 1,25 | 28,39 | 7,46 | 9,35 | 0,16 |
| 19 | Galantharminum | 1,42 | 14,44 | 1,09 | 10,73 | 0,42 |
| 20 | Platyphyllini_hydrotartr | 2,43 | 19,08 | 20,58 | 15,63 | 0,55 |
| 21 | Pirenzepine-gastrozepi | 1,19 | 11,72 | 5,70 | 15,20 | 0,49 |
| 22 | Dimecolinum | 4,06 | 8,16 | 2,48 | 20,50 | 1,36 |
| 23 | Bopindolol | 1,21 | 21,38 | -4,39 | 7,81 | 0,12 |
| 24 | Pindolol | 0,76 | 21,25 | -2,03 | 6,26 | -0,05 |

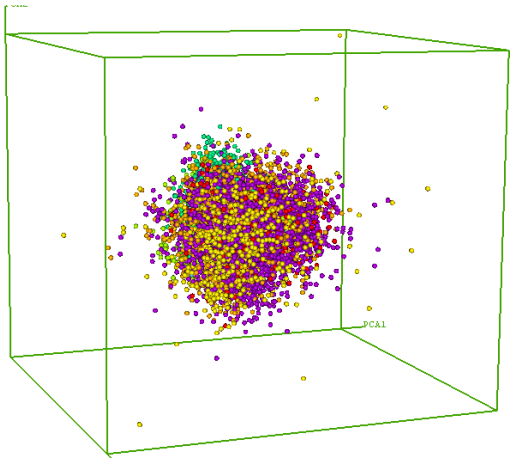



**Рисунок 12. Пространственная диаграмма деск дескрипторов молекулярной структуры с цветовым кодом для различных видов биологической активности.**

У нас в руках инструмент для прогнозирования биологической активности. Теперь если мы вычислим эти три свойства для нового исследуемого вещества и расположим для него точку на нашей диаграмме, то в зависимости от того в какую зону эта точка попадет мы сделаем вывод о биологической активности исследуемого соединения. Однако на практике для достоверного прогнозирования биологической активности трех дескрипторов обычно не достаточно. Используют десятки и сотни разнообразных свойств веществ. При этом молекулы представлены точками уже не в трехмерном, а в многомерном пространстве, которое визуализировать наглядно не представляется возможным. Однако можно вычислить расстояния между такими точками в многомерном пространстве по их многомерным координатам. Это расстояние и используют как меру сходства между структурами.

Прежде чем достигнуть рецептора, попадая в организм человека лекарственное вещество проходит ряд преград и может подвергаться превращениям.

Поэтому помимо структурных особенностей свойства лекарства зависят еще и от физико-химичеких свойств. Так растворимость связана с высвобождением из лекарственной формы и скоростью выведения из организма. Коэффициент липофильности определяет проницаемость через биомембраны и накопление в жировых тканях. Размеры молекулы и молекулярная масса связаны с проницаемостью через поры. Также важную роль играют потенциал ионизации, дипольный момент, поляризуемость и др. свойства.

Для вычисления физико-химических свойств веществ используются квантово-химические методы моделирования молекул (программы MOPAC, GAMESS, GAUSSIAN).

Для прогнозирования биологической активности веществ следует учитывать в совокупности и тонкости пространственной структуры молекулы и физико-химические свойства вещества. Поэтому метод дескрипторов молекулярной структуры при возможности используют в совокупности с исследованиями лиганд-рецепторного взаимодействия.

На практике поиск новых биологически активных веществ происходит в тесном сотрудничестве исследователей из разных дисциплин: химических, физических, математических и биологических. Ученые – химики предлагают возможные для синтеза структуры, при этом возникает вопрос о том, какие из них синтезировать, т.к. выбор огромен. Компьютерная химия позволяет в кратчайшие сроки дать ответ на этот вопрос и выбрать вещества наиболее перспективные с точки зрения биологической активности. Далее выбранные соединения синтезируют и подвергают биологическим исследованиям. После чего из них выбирают наиболее активные и наименее токсичные. Из лучших веществ химики вновь предлагают синтезировать огромное количество производных. Вновь производиться виртуальный скрининг, затем синтез и фармакологический скрининг небольшого количества веществ. Такой цикл повторяется до тех пор, пока не будет достигнуто приемлемое соотношение эффективная доза/токсичная доза.

Компьютер может не только анализировать виртуальные вещества, но и «придумывать» новые. Такой виртуальный синтез называют латинским термином «de novo», т.е. «с самого начала». При этом в компьютере происходит процесс эволюции химических структур. В качестве исходных данных исследователь задает программе некоторую начальную структуру. Затем программа автоматически случайным образом изменяет структуру, производя небольшие её «мутации». После этого полученный набор структур пропускается через процедуру виртуального скрининга, и алгоритм выбирает из них наиболее активную структуру. После этого снова производится мутация уже этой структуры и снова скрининг. Такие циклы повторяются длительное время в автоматическом режиме. В результате компьютер выдает все более и более активные вещества. Алгоритмы с таким принципом называют эволюционными [20].

## 4. Перспективы развития компьютерного конструирования лекарств

Благодаря успехам в биоинформатике можно говорить о том, что исследованиями в области компьютерного конструирования лекарств проложен путь от гена к лекарству, который включает в себя следующие этапы:

-Анализ генома человека и/или анализ генома болезнетворных микроорганизмов.

-Выявление составляющих генома (генов), кодирующих макромолекулы, которые могут быть использованы в качестве мишеней для новых лекарств.

-Экспериментальное определение или компьютерное моделирование структуры макромолекул-мишеней.

-Поиск или конструирование (низкомолекулярных) органических веществ, которые способны взаимодействовать с молекулами-мишенями. Моделирование взаимодействия лигандов с макромолекулами-мишенями и сравнительная оценка прочности связывания.

-Синтез перспективных веществ или их выделение из природных объектов.

-Выбор оптимальной лекарственной формы.

Развитие вычислительной техники позволяет решать задачи все большей сложности. Создание параллельных вычислительных систем (суперкомпьютеров), увеличило производительность вычислений до сотен триллионов операций с плавающей запятой в секунду (рис. 13). Это открывает возможности для решения новых задач. В настоящее время в сети Интернет с целью повышения производительности вычислительных систем создаются так называемые GRID-системы – распределенные вычислительные системы, состоящие из тысяч компьютеров и вычислительных центров по всему Миру. Такие системы способны давать наибольшую вычислительную мощность и используются в настоящее время для решения задач исследования генома человека, виртуального фолдинга, исследования динамики рецепторов и ферментов, а также сложных надмолекулярных систем.

**Рисунок 13. Суперкомпьютер Columbia компании SGI с производительнстью 42,7 трлн. операций в секунду.**

Международный научно-исследовательский проект по расшифровке генома человека достиг полного прочтения генетического кода человека, однако работа проекта продолжается и направлена на расшифровку полученной информации, определение функций генов. Для этого

используются мощнейшие вычислительные системы, работающие в круглосуточном режиме, а также новые алгоритмы расшифровки генетической информации, использующие методы искусственного интеллекта. В ближайшие несколько лет ожидается множество новых открытий в области исследований генома человека, что может оказаться полезным в разработке новых эффективных лекарств против таких заболеваний как СПИД, рак, болезнь Альцгеймера и др.

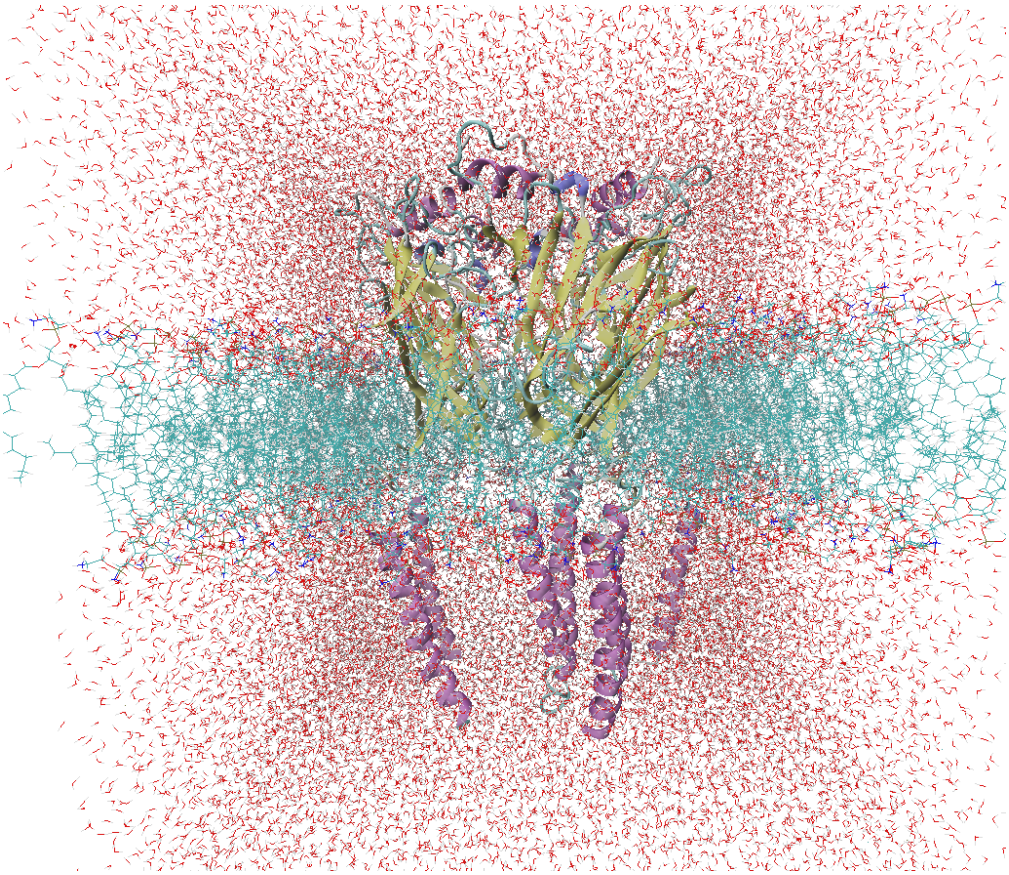


**Рисунок 14. Модель системы состоящей из ГАМК-А рецептора человека встроенного в фосфолипидную мембрану в окружении молекул воды.**

Еще одним активно развивающимся направлением в компьютерном конструировании лекарств является моделирование динамики сложных надмолекулярных структур: фрагментов клеточной мембраны со встроенными рецепторами (рис. 14) или ферментами в ячейке с водой и растворенными в ней веществами и даже целых клеточных органелл, например, моделирование динамики рибосомы, созданы молекулярные модели некоторых вирусов и фагов. Существует проект создания динамической молекулярной модели одноклеточного организма E. coli.

## 5. Заключение

Подавляющее большинство работ по созданию лекарственных препаратов было сделано без всякого «осознанного конструирования»,

методом «проб и ошибок», когда химики-органики в значительной степени произвольно заменяли одни химические группы другими.

99,98% вновь выделенных и синтезированных веществ отсеивались по причинам низкой эффективности или высокой токсичности.

Однако постепенно на основании полученных результатов возникало понимание того, какой стратегии и тактики нужно придерживаться, чтобы сконструировать лекарственное соединение.

Учитывая колоссальную сложность зависимости структуры вещества и его биологической активности, огромное разнообразие химических структур и сложность процессов, происходящих в живом организме, использование вычислительной техники на различных этапах создания новых лекарств очевидно является незаменимым инструментом.

**Используемая литература**

1. А.В. Погребняк, А.А. Глушко, Свидетельство об официальной регистрации программы для ЭВМ «Molecular Space» №2003612547 / Заявка №2003612314 от 10.11.2003. Зарегистрировано в Реестре программ для ЭВМ 21.11.2003.
2. Гейзенберг В., Шредингер Э., Дирак П. А. М. Современная квантовая механика. Три нобелевских доклада. Л.- М.: Гостехиздат, 1934.
3. Джеммер М. Эволюция понятий квантовой механики. М.: Наука, 1985. — 384с.
4. Погребняк А.В., Глушко А.А. Комбинированный метод прогнозирования биологической активности на основе сравнительного анализа молекулярного поля и метода потенциальных функций (MSPACE) // Биомедицинская химия. – 2004. – Т. 50, приложение №1. - С.38-41
5. Погребняк А.В., Глушко А.А. Основные направления практического использования результатов многомерного анализа молекулярных дескрипторов лекарственных веществ // IX российский национальный конгресс «Человек и лекарство»: тезисы докладов - Москва, 2002
6. Филимонов Д.А., Поройков В.В. (2006). Прогноз спектров биологической активности органических соединений. Российский химический журнал, 50 (2), 66-75.
7. Хэнч К. Об использовании количественных соотношений структура-активность при конструировании лекарств // Хим.-фармац. журн., 1980, №10, С.15-30
8. Чугунов А.О. // Изваяние невидимого (Молекулярная скульптура). Компьютерра, 2007, 712, С. 24–26
9. Anna L. Blobaum and Lawrence J. Marnett // Structural and Functional Basis of Cyclooxygenase Inhibition. Journal of Medicinal Chemistry, 2007, Vol. 50, N 7, P. 1425-1141.
10. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function / Morris, G. M. et all. // Computational Chemistry, 1998. – V.19. – P. 1639-1662.

11. Dill K.A., Ozkan S.B., Weikl T.R., Chodera J.D., Voelz V.A. // The protein folding problem: when will it be solved? Curr. Opin. Struct. Biol., 2007, **17**, 342–346
12. Ginalski K. // Comparative modeling for protein structure prediction. Curr. Opin. Struct. Biol., 2006, **16**, 172–177
13. H. Kubinyi // Comparative Molecular Field Analysis (CoMFA), in: The Encyclopedia of Computational Chemistry. Vol. 1, pp. 448-460.
14. http://swissmodel.expasy.org/repository/
15. http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606
16. http://www.pdb.org
17. J. Drews, Drug Discovery: A Historical Perspective. Science, 2000, **287**, P. 1960-1964.
18. S. Shacham, M. Topf, N. Avisar,F. Glaser, Y. Marantz, S. Bar-Haim, S. Noiman, Z. Naor, O. M. Becker // Modeling the 3D structure of GPCRs from sequence. Medicinal Research Reviews, Special Issue: Fourth Winter Conference on Medicinal and Bioorganic Chemistry. Part 1, 2001, Vol. 21, Issue 5, P. 472–483.
19. Sieghart W. // Structure and pharmacology of γ-aminobutyric acid$_A$receptor subtypes. Pharmacol Rev 47, 1995, P. 181-234.
20. Willett P. Genetic algorithms in molecular recognition and design. Trends Biotechnol. 1995, Vol. 13, P. 516–521.

## Литература, рекомендуемая студентам

1. Погребняк, А.В. Молекулярное моделирование и дизайн биологически активных веществ /А.В. Погребняк. - Ростов-на-Дону: Изд-во СКНЦ ВШ, 2003. – 229 с.
2. Чугунов А.О. // Изваяние невидимого (Молекулярная скульптура). Компьютерра, 2007, 712, С. 24–26
3. Хэнч К. Об использовании количественных соотношений структура-активность при конструировании лекарств // Хим.-фармац. журн., 1980, №10, С.15-30
4. http://biomolecula.ru/insilico/
5. http://www.moldyn.ru/library/moldyn/moldyn.htm

## Глоссарий

**De novo - (**лат. «от нового») - латинская фраза, означающая начало действия от нового.

**Drug design -** (англ. «дизайн лекарств») - рациональный поиск новых лекарственных веществ, основанный на знании о биологической мишени.

**GRID -** это форма распределённых вычислений, в которой «виртуальный суперкомпьютер» представлен в виде кластера соединённых с

помощью сети, слабосвязанных компьютеров, работающих вместе для выполнения огромного количества заданий (операций, работ). Эта технология применяется для решения научных, математических задач, требующих значительных вычислительных ресурсов.

**In silico** - термин, обозначающий компьютерное моделирование (симуляцию) эксперимента, чаще биологического. Фраза была создана по аналогии с фразами in vivo (в живом организме) и in vitro (в пробирке), которые часто используются в биологии, и сама не является латинской. Термин по написанию близок к латинскому выражению in silicio- «в кремнии», поскольку кремний как полупроводниковый материал играет важную роль в производстве компьютерного оборудования.

**In vitro –** (лат. «в стекле») — это технология выполнения экспериментов, когда опыты проводятся в пробирке, либо, в более общем смысле, вне живого организма.

**In vivo – (**лат. «в живом» ) in vivo обозначает проведение экспериментов на (или внутри) живой ткани при живом организме.

**Биоинформатика** - одна из дисциплин биологии, изучающая молекулярные процессы in silico, при помощи компьютеров.

**Виртуальный скрининг** - это вычислительная процедура, которая включает автоматизированный просмотр базы данных химических соединений и отбор тех из них, для которых прогнозируется наличие желаемых свойств.

**Виртуальный фолдинг белка** - моделирование процесса фолдинга белка методом молекулярной динамики с целью установления его третичной структуры.

**Волновая механика** - вариант квантовой механики, разработанный в 1926 г. Эрвином Шрёдингером. Объясняет поведение электронов с точки зрения их волновых свойств, и хотя ее быстро вытеснил более общий вариант, сформулированный физиком Полем Дираком, уравнения волновой механики до сих пор широко используются в вычислениях.

**Волновая функция** - комплекснозначная функция, используемая в квантовой механике для описания чистого состояния квантовомеханической системы. Физический смысл приписывается квадрату её модуля, который интерпретируется как плотность вероятности обнаружить систему в данной точке.

**Вторичная структура белка** - локальное упорядочивание фрагмента полипептидной цепи, стабилизированное водородными связями.

**Вычислительная химия** - ветвь химии, которая использует компьютеры для решения химических проблем. Вычислительная химия использует результаты классической и квантовой теоретической химии, реализованные в виде эффективных компьютерных программ, для вычисления свойств и определения структуры молекулярных систем.

**Геном** - совокупность всех генов организма; его полный хромосомный набор.

**Гибкий докинг -** разновидность молекулярного докинга, при котором учитывается конформационная подвижность лиганда и рецептора.

**Дескриптор молекулярной структуры -** свойство молекулярной структуры или вещества, которое может быть выражено в виде числа. Дескрипторы применяются для изучения количественной связи «структура-активность».

**ДНК – дезоксирибонуклеиновая кислота -** один из двух типов нуклеиновых кислот, обеспечивающих хранение, передачу из поколения в поколение и реализацию генетической программы развития и функционирования живых организмов.

**Жесткий докинг -** разновидность молекулярного докинга, при котором не учитывается конформационная подвижность молекул.

**Имитация отжига –** общий алгоритмический метод решения задачи глобальной оптимизации. Алгоритм основывается на имитации физического процесса, который происходит при кристаллизации вещества из жидкого состояния в твёрдое, в том числе при отжигеметаллов.

**Ионный канал -** порообразующие белки (одиночные либо целые комплексы), поддерживающие разницу потенциалов, которая существует между внешней и внутренней сторонами клеточной мембраны всех живых клеток. Относятся к транспортным белкам. С их помощью ионы перемещаются согласно их электрохимическим градиентам сквозь мембрану.

**Квант** - неделимая порция какой-либо величины в физике. В основе понятия лежит представление квантовой механики о том, что некоторые физические величины могут принимать только определённые значения.

**Квантовая механика** - раздел теоретической физики, описывающий квантовые системы и законы их движения.

**Квантовая химия** - это направление химии, рассматривающее строение и свойства химических соединений, реакционную способность, кинетику и механизм химических реакций на основе квантовой механики. Вследствие того, что сложность изучаемых объектов не позволяет в полной мере математически формализовать процессы, происходящие в химических системах на микроуровне, применяют приближенные методы расчета.

**Классическая механика Ньютона** - вид механики (раздела физики, изучающей законы изменения положений тел и причины, это вызывающие), основанный на законах Ньютона и принципе относительности Галилея. Поэтому её часто называют «Ньютоновской механикой».

**Количество операций с плавающей запятой в секунду –** единица измерения производительности комьютера, обозначающая количество операций с десятичными числами с плавающей запято в единицу времени.

**Комплементарность -** взаимное соответствие молекул биополимеров или их фрагментов, обеспечивающее образование связей между пространственно взаимодополняющими (комплементарными) фрагментами молекул или их структурных фрагментов вследствиесупрамолекулярных взаимодействий.

**Компьютерная химия** (Математическая химия) — сравнительно молодая область химии, основанная на применении теории графов к химическим задачам фундаментального и прикладного характера.

**Компьютерное конструирование лекарств** – использование компьютерных вычислительной техники на различных этапах создания лекарственных препаратов с целью оптимизации научных исследований.

**Конформация** - геометрическая форма, которую может принимать молекула при вращении атомов или групп атомов (заместителей) вокруг простых связей при сохранении неизменным порядка химической связи атомов (химического строения), длины связей и валентных углов.

**Корпускулярно-волновой дуализм** - принцип, согласно которому любой объект может проявлять как волновые, так и корпускулярныесвойства. Был введён при разработке квантовой механики для интерпретации явлений, наблюдаемых в микромире, с точки зрения классических концепций.

**Коэффициент липофильности** - свойство вещества, означающее его химическое сродство к органическим веществам. Величина, которая определяется экспериментально, а также может быть рассчитана при помощи таблицы инкрементов групп атомов для органических соединений. Для определения коэффициента липофильноси рассчитывают десятичный логарифм коэффициента распределения вещества в системе октанол-вода.

**Лекарственная форма** - придаваемое лекарственному средству или лекарственному растительному сырью удобное для применения состояние, при котором достигается необходимый лечебный эффект.

**Лиганд** - низкомолекулярное вещество, имеющее высокое сродство к сайту связывания рецептора.

**Математическая химия** — раздел теоретической химии, область исследований, посвящённая новым применениям математики к химическим задачам. Основная область интересов — это математическое моделирование химических явлений и процессов.

**Метод рентгеноструктурного анализа (РСА)** - один из дифракционных методов исследования структуры вещества. В основе данного метода лежит явление дифракции рентгеновских лучей на трехмерной кристаллической решетке. Рентгеноструктурный анализ является распространенным методом определения структуры вещества.

**Метод фолдинга по гомологии** - виртуальный фолдинг белка, происходящий с учетом информации о пространственной структуре белков - гомологов.

**Методы искусственного интеллекта** - компьютерные алгоримты, направленные на решение трудно формализуемых, творческих задач, таких как распознавание образов, оптимизация, генерация гипотез.

**Молекулярная динамика** - метод, в котором временная эволюция системы взаимодействующих атомов или частиц отслеживается путем решения уравнений движения.

**Молекулярный докинг** - процедура моделирования взаимодействия лиганда с рецептором, проводимая с целью получения набора гипотез об их взаимном расположении в пространстве и конформациях.

**Надмолекулярная система –** система, состоящая из некоторого количества молекул (клеточная мембрана, рибосома, хлоропласт)

**Параллельная вычислительная система** - это физические компьютерные, а также программные системы, реализующие тем или иным способом параллельную обработку данных на многих вычислительных узлах.

**Первичная структура белка** - последовательность аминокислот в полипептидной цепи.

**Потенциал Леннарда-Джонса** - (потенциал 6-12) — простая модель парного взаимодействия неполярных молекул, описывающая зависимость энергии взаимодействия двух частиц от расстояния между ними.

**Проект «Геном человека» - Проект по расшифровке генома человека** (англ. The Human Genome Project, HGP) — международный научно-исследовательский проект, главной целью которого было определить последовательность нуклеотидов, которые составляют ДНК и идентифицировать 20.000-25.000 генов в человеческом геноме. Проект начался в 1990 году, под руководством Джеймса Уотсона под эгидой Национальной организации здравоохранения США. В 2000 году был выпущен рабочий черновик структуры генома, полный геном — в 2003, однако и сегодня дополнительный анализ некоторых участков ещё не закончен.

**Рецептор** - биологическое структурное образование, способное реагировать на разлиные химические и физические факторы. Так хеморецепторы реагируют на присутствие определенных веществ. Чаще всего рецептор является белковой молекулой, состоящей из одной или нескольких субъединиц.

**Рецепторы, сопряженные с G белком** - (англ. G protein-coupled receptors, GPCRs), также известные как **семиспиральные рецепторы**, составляют большое семейство трансмембранных рецепторов. GPCR выполняют функцию активаторов внутриклеточных путей передачи сигнала, приводящими в итоге к клеточному ответу.

**РНК – рибонуклеиновая кислота** - нуклеиновые кислоты, полимеры нуклеотидов, в состав которых входят остаток ортофосфорной кислоты, рибоза (в отличие от ДНК, содержащей дезоксирибозу) и азотистые основания — аденин, цитозин, гуанин и урацил (в отличие отДНК, содержащей вместо урацила тимин). Эти молекулы содержатся в клетках всех живых организмов, а также в некоторых вирусах.

**Сайт связывания рецептора** - часть рецептора, с которой происходит комплементарное связывание его лигандов.

**Соотношение** эффективная **доза**/токсичная **доза** - соотношение между дозой, в которой проявляется определенный вид биологической активности и дозой, при которой проявляется токсичность вещества.

**Сравнительный анализ молекулярного поля** - статистический метод предсказания биологической активности веществ, основанный на анализе распределения различных молекулярных полей (электронная плотность, электростатический потенциал, потенциал Леннарда-Джонса и др.).

**Суперкомпьютер** - вычислительная машина, значительно превосходящая по своим техническим параметрам большинство существующих компьютеров. Как правило, современные суперкомпьютеры представляют собой большое число высокопроизводительных серверных компьютеров, соединённых друг с другом для достижения максимальной производительности в рамках подхода распараллеливания вычислительной задачи.

**Теория графов** - раздел дискретной математики, изучающий свойства графов. В общем смысле граф представляется как множество вершин (узлов), соединённых рёбрами. В химии теория графов применяется для формального описания молекулярных структур.

**Термостатирование в молекулярной динамике –** компьютерный алгоритм предназначенный для контроля за температурой молекулярной системы в процессе моделирования молекулярной динамики

**Третичная структура белка** - пространственное строение полипептидной цепи (набор пространственных координат составляющих белок атомов). Структурно состоит из элементов вторичной структуры, стабилизированных различными типами взаимодействий, в которых гидрофобные взаимодействия играют важнейшую роль.

**Фармакологический скрининг** - экспериментальное исследование биологической активности большой группы веществ, преимущественно с использованием методик, позволяющих оценить наличие биологической активности за короткое время при малых количествах вещества.

**Фермент** - обычно белковые молекулы или молекулы РНК (рибозимы) или их комплексы, ускоряющие (катализирующие) химические реакции в живых системах.

**Фолдинг белка** - процесс спонтанного сворачивания полипептидной цепи в уникальную нативную пространственную структуру.

**Хемоинформатика** - это научная дисциплина, возникшая за последние 40 лет в пограничной области между химией и вычислительной математикой. Было осознано, что во многих областях химии огромный объем информации, накопленный в ходе химических исследований, может быть обработан и проанализирован только с помощью компьютеров. Более того, многие из проблем в химии настолько сложны, что для их решения требуются новые подходы, основанные на применении методов информатики. Исходя из этого, были разработаны методы для построения баз данных по химическим соединениям и реакциям, для прогнозирования физических, химических и биологических свойств соединений и материалов, для поиска новых лекарственных препаратов, анализа спектральной информации, для

предсказания хода химических реакций и планирования органического синтеза.

**Четвертичная структура белка -** взаимное расположение нескольких полипептидных цепей в составе единого белкового комплекса. Белковые молекулы, входящие в состав белка с четвертичной структурой, образуются на рибосомах по отдельности и лишь после окончания синтеза образуют общую надмолекулярную структуру.

**Эволюционный алгоритм -** направление в искусственном интеллекте, которое использует и моделирует биологическую эволюцию. Такие алгоритмы относятся к адаптивным поисковым механизмам.

**Электродинамика** - раздел физики, изучающий электромагнитное поле в наиболее общем случае (то есть, рассматриваются переменные поля, зависящие от времени) и его взаимодействие с телами, имеющими электрический заряд (электромагнитное взаимодействие). Предмет электродинамики включает связь электрических и магнитных явлений, электромагнитное излучение (в разных условиях, как свободное, так и в разнообразных случаях взаимодействии с веществом), электрический ток (вообще говоря, переменный) и его взаимодействие с электромагнитным полем (электрический ток может быть рассмотрен при этом как совокупность движущихся заряженных частиц).

**Электронная плотность** - плотность вероятности обнаружения электрона в данной точке пространства.

**Электростатический потенциал -** скалярная энергетическая характеристика электростатического поля, характеризующая потенциальную энергию поля, которой обладает единичный заряд, помещённый в данную точку поля.

**Элементарная частица** - собирательный термин, относящийся к микрообъектам в субъядерном масштабе, которые (согласно существующим представлениям) невозможно расщепить на составные части.