

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
НАУКИ ИНСТИТУТ БИОХИМИИ И ФИЗИОЛОГИИ РАСТЕНИЙ И
МИКРООРГАНИЗМОВ РОССИЙСКОЙ АКАДЕМИИ НАУК

Базовая кафедра биофизики

Моделирование белков и их комплексов (на примере белка VirE2)

М. И. Чумаков

Методическое пособие для студентов факультета нелинейных процессов и биологического факультета Саратовского государственного университета по специальностям «Физика открытых нелинейных систем» (010710) и «Биоинженерия и биоинформатика» (020501)

Рекомендовано базовой кафедрой биофизики факультета нелинейных процессов Саратовского государственного университета, протокол заседания №3 от 7 мая 2013 г.

Саратов 2013

Содержание

Содержание	Ошибка! Закладка не определена.
Сокращения.....	3
Введение	4
1. Обзор литературы.....	5
1.1. Роль компьютерной биологии в изучении белков.....	5
1.2. Компьютерное моделирование белковых структур и базы данных по белкам.....	6
2. Доступные в сети Интернет методы и ресурсы моделирования белковых структур.....	10
2.1. База данных Swiss-Prot.....	10
2.2. Пакет программ SAM-T06.....	11
2.3. Ресурс Predict Protein.....	12
2.4. Ресурс GWIDD.....	13
2.5. Программа Phyre.....	16
3 Белок VirE2 описание, функции.....	17
3.1. Роль белка VirE2 в процессе трансформации растений.....	17
3.2. Механизмы формирования белковых пор в липидных мембранах.....	20
3.3.1 Примеры использования баз и программ для анализа порообразующих белков.....	21
3.3.1.1 Базы белков NCBI и PDB.....	21
3.3.2. Программа PHYRE.....	22
3.3.3. Программы Deep View Swiss-PdbViewer.....	22
4. 4. Использование баз и программ для анализа порообразующих белков.....	233
4.1. Получение модели белка VirE2 по его аминокислотной последовательности.....	234
4.2. Методика построения и анализа модели белка VirE2.....	234
4.3. Моделирование комплекса белков VirE1 и VirE2.....	288
4.4. Моделирование комплекса белков VirE2 и VIP1.....	33
4.5. Оценка возможного взаимодействия VirE2 с T-ДНК.....	38
4.6. Проверка корректности построенных моделей белков.....	39
4.7. Проверка корректности построенных моделей белковых комплексов.....	41
4.8. Компьютерное и экспериментальное исследование свойств белка VirE2 формировать поры в биологических мембранах.....	43
Литература	45
Использованные веб-сайты	50
Приложения	51
Приложение 1. Обозначения аминокислот.....	51
Приложение 2. Аминокислотная последовательность белка VirE2 (в однобуквенном формате).....	52
Приложение 3. Аминокислотная последовательность белка VirE1 (в однобуквенном формате).....	52
Приложение 4. Аминокислотная последовательность белка Bacteriorhodopsin (в однобуквенном формате).....	

Сокращения

T-ДНК – транспортируемая дезоксирибонуклеиновая кислота

ЯМР (NMR) – ядерный магнитный резонанс

HMM – скрытая модель Маркова (Hidden Markov model)

Латинские названия

in vitro – в искусственных условиях

in vivo – в организме

vir – вирулентный

Единицы измерения

Å – ангстрем

mV – милливольт

nSм – наносименс

M – моль

s – секунда

Введение

Agrobacterium является обитающим в почве патогеном, который вызывает образование опухолей на стеблях, листьях растений (главным образом двудольных). Опухоль у растений образуется под влиянием продукции дополнительных фитогормонов, которые контролируются фрагментом бактериальной ДНК, встраиваемой в хромосому растительной клетки (Чумаков, 2001).

Традиционно считается, что нить Т-ДНК (Т-нить) может переноситься внутрь растительной клетки. Перенос Т-нити в растительную цитоплазму происходит с помощью пилотирующего белка VirD2, где происходит формирование Т-комплекса: Т-нить покрывается несколькими сотнями агробактериальных белковых молекул VirE2 (Citovsky et al. 1989, 1994; Tinland et al. 1993; Tinland, 1996; Rossi et al., 1996). Комплекс Т-ДНК-VirD2 и белок VirE2 попадают в растительную клетку независимо (Gelvin, 1998) и, возможно, разными каналами. Не исключено, что процесс переноса Т-ДНК может осуществляться по пути, сходному с проникновением вирусов через мембрану клетки-хозяина. Для интеграции в хромосому растения Т-нити требуется попасть в ядро и для этого пересечь клеточную и ядерную мембраны. Одним из предполагаемых способов переноса Т-ДНК через клеточную мембрану является перенос через белковую пору, образуемую белком VirE2 (Чумаков, 2001, Dumas et al., 2001).

К сожалению, в настоящее время пространственные структуры белков VirE2 и VirD2 неизвестны. Для их расчёта требуются мощные компьютерные технологии.

Целью данной работы является моделирование двух-, трехмерных структур белка VirE2, изучение его свойств и образуемых им комплексов с другими белками и Т-ДНК.

1. Обзор литературы

1.1. Роль компьютерной биологии в изучении белков

В биологических исследованиях компьютеры и суперкомпьютеры становятся всё более самостоятельным инструментом познания и получения прикладных результатов.

К числу основных вычислительных задач компьютерной биологии в настоящее время относятся (Лахно, 2000):

1. Распознавание белок-кодирующих участков в первичной структуре биополимеров. Сравнительный анализ первичных структур биополимеров.
2. Расшифровка пространственной структуры биополимеров и их комплексов. (Рентгеноструктурный анализ, методы ЯМР).
3. Пространственное сворачивание белков (3D-фолдинг).
4. Моделирование структуры и динамики биомакромолекул.
5. Создание и сопровождение специализированных баз данных (баз белковых структур, нуклеотидных последовательностей, путей метаболизма, клеточных ансамблей и др.).

Соответствие между молекулярной и компьютерной биологией показано на рис. 1.3.1.

Молекулярная биология

Компьютерная биология

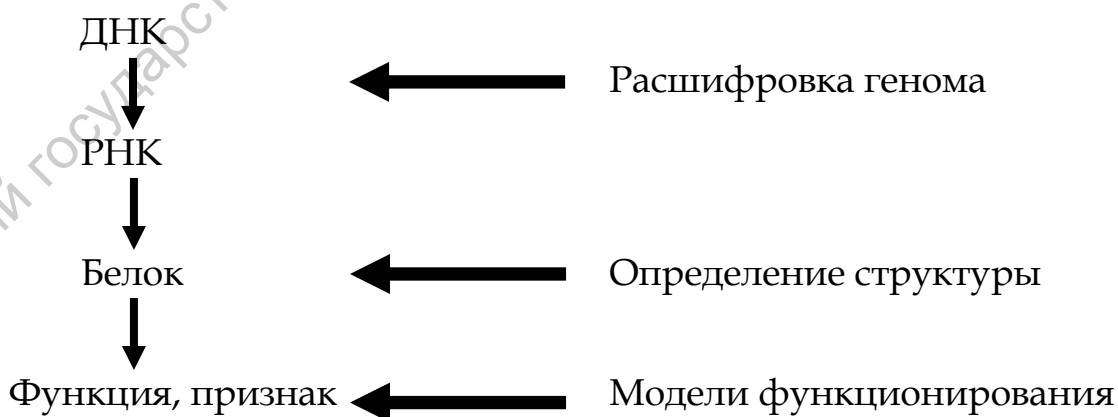


Рис. 1.3.1. Связь между молекулярной и компьютерной биологией (Лахно, 2000)

Одной из центральных задач компьютерной биологии является предсказание пространственной структуры белков по

аминокислотной последовательности, т. е. фактически по последовательности нуклеотидных пар в ДНК. Важность решения этой задачи состоит еще и в том, что число известных первичных белковых структур, установленных по известным нуклеотидным последовательностям ДНК, намного превосходит число известных пространственных белковых структур.

Формально, зная взаимодействие между отдельными атомами в полипептидной цепи с известной первичной структурой и окружающими такую цепь молекулами растворителя, путем минимизации свободной энергии всей системы можно было бы найти искомую структуру. Даже для небольшого белка это задача поиска глобального минимума функции десятков тысяч переменных. Бесперспективность точного решения подобной задачи с помощью любого мыслимого суперкомпьютера вполне очевидна.

В настоящее время для решения проблемы сворачивания разработано большое число приближенных подходов (Fischer et al., 1996). Один из наиболее эффективных – использование информации о гомологии, т.е. о пространственной структуре белков, обладающих первичной структурой, близкой к исследуемому белку (Финкельштейн, 1999). Известная пространственная структура белка, гомологичного исследуемому белку, берется в качестве начального приближения, а затем производится ее уточнение.

1.2. Компьютерное моделирование белковых структур и базы данных по белкам

Достоверно предсказать структуру белка, основываясь на его аминокислотной последовательности, в настоящее время всё ещё невозможно. Однако попытки сделать это продолжаются, и в сети Интернет существует множество баз данных, где можно найти первичную структуру практически любого известного или даже гипотетического белка в виде аминокислотной последовательности. Для предсказания структуры белка (главным образом, вторичной) по его первичной последовательности существует множество различных программ. Используемые ими методы можно разделить на четыре категории (Ouali, King, 2000):

1. Простая линейная статистика, основанная на соответствии остатка определенной структуре или на физико-химических свойствах аминокислоты
2. Ближайшее соседство (k-nearest-neighbor method)
3. Обучение машины, например, использование нейронных сетей

4. Методы, использующие сложную нелинейную статистику, основанную на соответствии остатка определенной структуре и физико-химических свойствах аминокислоты (например, скрытая модель Маркова)

Все эти методы основаны на информации об известных структурах и зависят от текущего содержания баз данных. Наиболее ранние методы предсказания вторичной структуры относились к первой категории, используя свойства остатков (Garnier et al., 1978) и их предрасположенность (Chou, Fasman, 1974). В дальнейшем они получили свое развитие путем использования множественного выравнивания (Gibrat et al., 1987).

Приближения для ближайших соседей – это методы, происходящие от гипотезы, что одинаковые первичные последовательности будут иметь одинаковые структуры, и использовавшие эмпирически полученные данные (Levin, Garnier, 1988), искусственный интеллект (Yi, Lander, 1993), и выравнивания нескольких последовательностей (Salamov, Solovyev, 1995). При символическом обучении машин, вырабатываются правила на основе тренировочного набора данных, что и формирует базу для большинства исследований искусственного интеллекта. Эти правила могут быть применены к последовательностям для предсказания структуры (King, Sternberg, 1990, 1996). В скрытых моделях Маркова используются вероятностные модели, и предполагается, что вероятность нахождения остатка в структуре определенного конкретного типа зависит от структуры для предшествующего остатка (Krogh et al., 1994). Нейронные сети – одно из наиболее популярных приложений сложной нелинейной статистики в предсказании вторичной структуры белка; они принимают как аминокислотную последовательность, так и множественное выравнивание на входе и выдают предсказанную вторичную структуру (Rost, Sander, 1993; Cuff et al., 1998; Jones, 1999).

К примеру, в 2004 году PDB (<http://www.pdb.org/>) содержал около 50 оригинальных (без учета гомологов) структур спиральных мембранных белков (При этом, что лишь в человеческом организме их предположительно более 7000.) и около 50 структур белков в виде β -цилиндров с похожими последовательностями (причем все они относятся к грамотрицательным бактериям). Такой недостаток информации связан с трудностями рентгеноструктурного анализа мембранных белков. Поэтому возможность моделирования играет решающую роль, тем более что наилучшие методы предсказания

трансмембранных спиралей не намного уступают в точности экспериментально получаемым результатам низкого разрешения.

При этом база PDB постоянно расширяется за счет полученных экспериментальными методами структур (в том числе благодаря развитию электронной микроскопии и методов, основанных на использовании ядерного магнитного резонанса (ЯМР)), а с 15 октября 2006 года теоретические модели в базу PDB вообще не принимаются. Структуры, акцептованные до 15 октября 2006 года, хранятся в отдельном (общедоступном) архиве. Обновления публикуются каждый четверг в 1:00 am (Pacific time). На 20 мая 2008 г. в PDB насчитывалось 50 961 различных структур (см. Табл. 1.4.1.).

Таблица 1.2.1. Состав базы PDB на 20 мая 2008 г.

		Тип молекулы				
		Белки	Нуклеиновые кислоты	Комплексы белок-НК	Другие	Всего
Экспериментальный метод	X-ray	40461	1034	1859	24	43378
	ЯМР	6352	809	138	7	7306
	Электронная микроскопия	122	11	46	0	179
	Другие	88	4	4	2	98
	Всего	47023	1858	2047	33	50961

Самый простой и наиболее распространенный критерий, показывающий точность предсказания – точность предсказания на аминокислоту – процентное количество правильно предсказанных остатков:

$$Q_3 = 100 \cdot \frac{\sum_{i=1}^3 c_i}{N}$$

где c_i – количество остатков, предсказанных правильно в состоянии i (H – спираль, E – тяж, и L – нерегулярная структура), и N – количество аминокислот в белке (или в данной последовательности). Типичная последовательность данных содержит приблизительно 32% H, 21% E, и 47% L. (Kabsch et al., 1983)

Измерения точности предсказания конкретной аминокислоты не отражают качество предсказания. Вот три простых измерения для оценки качества предсказания сегментов вторичной структуры:

1. Количество сегментов в белке
2. Средняя длина сегмента

3. Количественное распределение сегментов по длинам

Эти измерения связаны. Они полезны для характеристики методов предсказания. В действительности методы с высоким предсказанием конкретного сегмента, имеют нереальное распределение.

Универсального и идеально точного метода моделирования структуры белковых молекул до сих пор нет, и далеко не для всех белков структуры выявлены экспериментально. В условиях отсутствия достоверных данных о трёхмерной структуре белка нам остаётся только положиться на построенные модели, тем более большинство существующих методов их построения дают правильные предсказания для большинства белков.

2. Доступные в сети Интернет методы и ресурсы моделирования белковых структур

2.1. База данных Swiss-Prot

База данных белков, аннотированных вручную, **UniProtKB/Swiss-Prot** образована в 1986 году в Швейцарии. С 2003 года базу поддерживает **UniProt Consortium** – совместный проект Швейцарского института биоинформатики (Swiss Institute of Bioinformatics, SIB), кафедры биоинформатики и структурной биологии Женевского университета (the Department of Bioinformatics and Structural Biology of the Geneva University), Европейского института биоинформатики (the European Bioinformatics Institute, EBI) и информационных ресурсов по белкам Медицинского центра университета Джорджтауна (the Georgetown University Medical Center's Protein Information Resource, PIR). База доступна в сети Интернет на ресурсе <http://www.expasy.org/sprot/>, а также имеет несколько зеркальных сайтов. Помимо того, что в аннотациях к представленным в базе белкам по возможности описаны их функции, пост-трансляционные модификации, составляющие их домены и сайты, вторичная (а иногда даже четвертичная) структура, сходства с другими белками, болезни вызываемые отсутствием или неполноценностью белка, а также различные «конфликтные» последовательности и варианты, база содержит также перекрёстные ссылки на другие ресурсы: базы нуклеотидных последовательностей (EMBL/GenBank/DDBJ), трёхмерных структур (PDB), белковых доменов, семейств и проч. Всего около 60 ресурсов всемирной паутины (подробный перечень на <http://ca.expasy.org/cgi-bin/lists?dbxref.txt>).

Дополнением к вручную аннотированной Swiss-Prot является база **UniProtKB/TrEMBL**, в которую включены трансляции EMBL, ещё не интегрированные в Swiss-Prot. Обе базы являются составными частями проекта **UniProt Knowledgebase**, который постоянно обновляется, и обновлённый 20 мая 2008 года выпуск 13.4 включил в себя UniProtKB/Swiss-Prot Release 55.4 (385 721 запись) и UniProtKB/TrEMBL Release 38.4 (5 814 087 записей).

UniProt предоставляет своим пользователям множество ресурсов, в числе которых программа **SwissModel** (<http://swissmodel.expasy.org/>), которая позволяет смоделировать трёхмерную структуру белка с заданной аминокислотной последовательностью, используя библиотеку структур белков, уже изученных методами рентгеноструктурного анализа. Программа имеет обычный для

работающих с базами приложений веб-интерфейс в виде формы, содержащей поля для ввода самой исследуемой последовательности (в формате FASTA), баз, по которым предполагается поиск гомологов, а также адреса электронной почты, на который должен быть отправлен отчёт о выполнении задания с гиперссылкой на файлы результатов. Для коммерческих пользователей также предусмотрено поле ввода пароля, что предотвратит несанкционированный доступ к результатам исследований. Если построение модели успешно, программа генерирует файл в формате PDB (некоторые приложения для визуализации трёхмерной структуры молекул работают с файлами этого же формата, но с расширением *.ENT), в противном случае возвращает сообщение о невозможности построения модели из-за отсутствия в библиотеке подходящих доменов и сайтов.

Для визуализации полученного результата предлагается использовать Deep View Swiss-PdbViewer (программа распространяется свободно и текущая версия 3.5 доступна на <http://www.expasy.ch/spdbv/text/download.htm>). Авторы Nicolas Guex, Alexandre Diemand, Torsten Schwede и Manuel C. Peitsch, предлагают вниманию пользователей обширный справочный материал. Также сайт Deep View имеет зеркало в

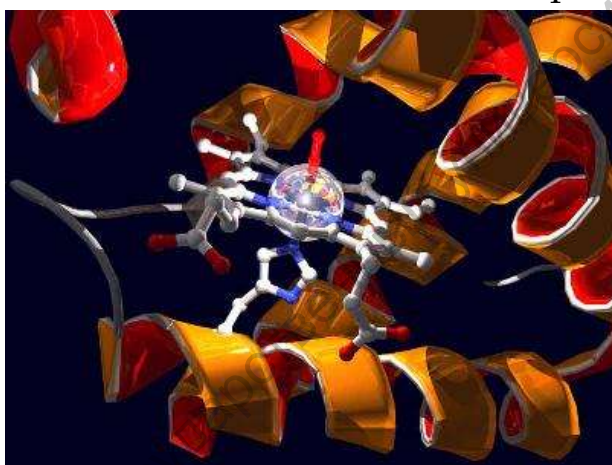


Рис. 2.1.1. Пример изображения трёхмерной структуры белка, взятый

Рунете (www.genebee.msu.ru/spdbv/). В программу уже добавлены инструменты расширенной работы с графикой и для пользователей открыты галереи наиболее интересных графических работ (на рисунке 2.1.1 приведен пример такой работы из Swiss-PdbViewer Art Gallery, расположенной на <http://www.expasy.ch/spdbv/text/gal.htm>) Также существует

независимая галерея А. Терпера:

<http://chemb16.leidenuniv.nl/~armand/>.

2.2. Пакет программ SAM-T06

На сайте Инженерной школы Джека Бэскина (Jack Baskin School of Engineering, UC Santa Cruz, Университет Калифорнии, США), основными направлениями работы которой являются биотехнологии, информационные технологии и нанотехнологии, по адресу www.cse.ucsc.edu/research/compbio/HMM-apps/HMM-applications.html

доступен целый ряд интерактивных приложений, основанный на использовании модели Маркова (hidden Markov model, HMM) и предназначенных для обнаружения гомологии и предсказания трёхмерной структуры белков. Весь пакет разработанных программ носит название SAM (Sequence Alignment and Modeling System). Наиболее мощным и точным инструментом является SAM-T06. Программа загружает исходные последовательности белков в формате FASTA и возвращает множественные выравнивания (по принципу SAM-T02 и SAM-T04), предсказания вторичной структуры, эскизы устойчивых локальных структур предсказанного белка, таблицы значений математического ожидания E для результатов сравнения с библиотекой структур SCOP, выравнивания к матрицам и модели PDB с полными координатами. Главное окно программы для ввода исходных данных находится по адресу http://www.soe.ucsc.edu/research/compbio/SAM_T06/T06-query.html

2.3 Ресурс Predict Protein

На веб-сайте Центра биоинформатики кафедры биохимии и молекулярной биологии Колумбийского университета (Columbia University Bioinformatics Center, CUBIC) по адресу <http://cubic.bioc.columbia.edu/> доступен ресурс для анализа последовательностей и предсказания функций и структуры белков под названием PredictProtein (PP). Имеется также выделенный сервер <http://www.predictprotein.org> с зеркалами в Европе, Австралии, Азии и Америке. Основным преимуществом данного ресурса является использование в процессе предсказания не только информации о первичной последовательности, но и эволюционной информации и результатов других предсказаний. Сервис работает также с нейронными сетями, векторными механизмами (support-vector machines) алгоритмами древа и классификаторами Bayesian'a. На странице приветствия пользователь может

- Заполнить форму предсказания белка – простую, расширенную или для экспертов, которые различаются набором настроек
- Обратиться к ресурсу META-PP, сводящему множество сервисов к одному META-интерфейсу
- Получить доступ к документации PredictProtein
- Воспользоваться дополнительными средствами выравнивания (в разделе post-processing tools) и прочими второстепенными функциями, делающими использование ресурса более удобным.

После ввода аминокислотной последовательности PredictProtein в качестве результатов возвращает

1. множественное выравнивание (т. е. результат поиска по базам)
2. эскизы ProSite
3. области малой сложности (SEG)
4. распределение доменов ProDom,
5. сигналы локализации в ядре
6. и предсказание
 - вторичной структуры
 - растворимости
 - глобулярных областей
 - трансмембранных спиралей
 - биспиральных областей
 - областей переключения структуры (structural switch regions)

Также опционально доступны элементы распознавания структур в процессе протяжки (трейдинга) предсказания и проверки точности предсказания.

В процессе работы PP базу данных последовательностей (в настоящее время только SWISSPROT) на предмет сходства последовательностей (при помощи BLASTP). Для получения эскизов используется база ProSite.

2.4. Ресурс GWIDD

Genome Wide Docking Database, доступная на сайте лаборатории профессора Пиа А. Vakser Центра биоинформатики Канзаского университета <http://gwidd.bioinformatics.ku.edu/home> представляет собой обширный ресурс полногеномного моделирования белок-белковых взаимодействий. В нем содержатся данные для множества организмов. Структуры взаимодействующих белков смоделированы либо, если это возможно, получены их кристаллографические координаты и состыкованы при помощи GRAMM-X. Ресурс не ограничен взаимодействиями базы GWIDD - на различных стадиях могут быть добавлены и другие последовательности и структуры. Работа системы GWIDD состоит из трёх этапов: Последовательность, Структура, Докинг (состыковка).

На этапе **Последовательность** (Sequence) можно выполнить поиск взаимодействующих белков и получить их последовательности. Поиск можно выполнить по идентификатору NCBI Geninfo ID, ключевым словам описания или даже по самой последовательности (в последнем случае необходимы также такие параметры как область поиска Organism и его точность E-value threshold). При поиске по ключевым словам пользователю предлагаются удовлетворяющие запросу белки (из базы GWIDD) и организмы, из которых они

выделены (NCBI). Выбрав подходящий белок, пользователь получает список доступных в базе взаимодействий с ним. Выбрав любое из них, пользователь перенаправляется в раздел Структур, где предлагается либо рассмотреть уже имеющееся взаимодействие (ссылка на первоисточник), либо самостоятельно повторить процесс состыковки, для чего придется загрузить на сервер PDB-модели белков или воспользоваться уже имеющимися, если они есть.

В раздел **Структура** (Structure) происходит построение моделей или (если доступны в базе) получение кристаллографических структур для последовательностей, выбранных на предыдущем этапе (Sequence) или непосредственно введенных последовательностей. Здесь можно ввести однобуквенную последовательность либо отправить текстовый файл. Сначала программа ищет совпадения в PDB и собственной базе и, если не находит, предлагает построить собственную модель. При построении модели система спрашивает разрешения добавить результаты моделирования в базу GWIDD. Согласившемуся, пользователю открываются дополнительные опции моделирования, в противном случае используются параметры по умолчанию. В дополнительных опциях можно выбрать протокол моделирования BLAST & NEST или MODELLER (требует наличия лицензии) а также максимальное значение e-value, минимальное покрытие модели, границы идентичности моделируемого белка и эталонных шаблонов-образцов, субституционную матрицу и количество моделей в выходном файле (наиболее вероятных). Результаты моделирования высылаются по электронной почте.

Также ресурс выполняет **Состыковку** (Docking) двух выбранных на этапе Structure или непосредственно введенных структур. Результаты моделирования также высылаются по электронной почте.

Чтобы просто выполнить состыковку, можно напрямую воспользоваться доступным здесь же ресурсом GRAMM, **Global Range Molecular Matching**. Для предсказания структуры комплекса данная программа использует только координаты атомов составляющих его молекул, не обращаясь к дополнительной информации о местах посадки. Программа выполняет исчерпывающий 6-мерный поиск по сходным трансляциям и ротациям молекул. В качестве молекулярных пар могут выступать два белка, белок и малая структура, две трансмембранные спирали и т. д. GRAMM можно использовать для моделей с высоким разрешением, для неточных моделей (в которых известны только основные структурные черты), в случаях сильных конформационных изменений и т. д.

Методика Global RAnge Molecular Matching (GRAMM) — это эмпирический подход к сглаживанию функции внутримолекулярной энергии изменением значения потенциалов межатомных взаимодействий. Метод находит область глобального минимума внутримолекулярной энергии для структур различной точности. Качество предсказания зависит от точности структур. Таким образом, состыковка структур с высоким разрешением с малыми конформационными изменениями даёт точное предсказание, в то время как состыковка структур с очень низким разрешением даст лишь общие черты комплекса.

По многочисленным просьбам различных лабораторий GRAMM стал общедоступным, также методика и сама программа, которая сейчас активно развивается, собираются отчеты об ошибках.

GRAMM скомпилирован на платформах SGI R10000, SGI R4000, SGI R4400, SGI R8000, Sun SPARC, IBM RS6000, DECAlpha, и ПК (Windows и Linux). Windows-версия рассчитана на все 32-разрядные разновидности MS Windows. Linux-версия скомпилирована на RedHat с библиотеками glibc2.0.

Скачать программу GRAMM v1.03 для различных платформ можно свободно (заполнив анкету) с сайта

<http://vakser.bioinformatics.ku.edu/resources/gramm/gramm1/>

Также на <http://vakser.bioinformatics.ku.edu/resources/gramm/grammx/> доступен GRAMM-X Protein Docking Web Server v.1.2.0.

Эта программа отличается от оригинального GRAMM, за исключением того, что оба пакета используют быстрое преобразование Фурье для глобального поиска наиболее устойчивой конформации.

Пользователь может отправить на этот web-сервер входные файлы и параметры, сам процесс состыковки будет запущен на компьютерном кластере Канзасского Университета.

Пользователю предлагается заполнить следующие поля:

Receptor PDB file (обязательно) – путь к загружаемому на сервер PDB-файлу, содержащему модель, которая выполняет роль рецептора.

Chain Ids of the Receptor – указание идентификаторов цепей рецептора одной или несколькими буквами. Например, LH – цепи L и H; C – цепь C; если оставить поле пустым, будут использоваться все цепи из файла рецептора.

Ligand PDB file (обязательно) путь к загружаемому на сервер PDB-файлу, содержащему модель, которая выполняет роль лиганда.

Chain Ids of the Ligand — указание идентификаторов цепей лиганда одной или несколькими буквами. Например, AB – цепи A и B.

B; X - цепь X; если оставить поле пустым, будут использоваться все цепи из файла лиганда.

Адрес электронной почты (обязательно) - Когда результаты готовы, они сохраняются во временной директории веб-сервера, а пользователю по электронной почте высылается ссылка на них.

Также есть дополнительные параметры:

Number of models to output - количество моделей (альтернативных предсказаний), которые нужно сохранить в конечном выходном файле (от 1 до 300).

Number of identical monomers if this complex is a homo-multimer - если комплекс представляет собой симметричный гомо-мультимер, то программа пытается построить его из набора идентичных субъединиц, указанных как рецептор. Если это не гомо-мультимер, поле надо оставить пустым. В настоящее время системой поддерживается только значение 2 (для гомо-димера).

2.5. Программа Phyre

На сайте Группы структурной биоинформатики отделения молекулярных биологических наук факультета естественных наук (Structural Bioinformatics Group, Division of Molecular Biosciences, Faculty of Natural Sciences) по адресу <http://www.sbg.bio.ic.ac.uk/> доступен целый ряд программ для построения 1D и 3D-моделей белков, выполнения выравниваний, доступа к базам данных и графического представления полученных результатов. Последнее обновление сервиса - 21 марта 2007 года. Особый интерес в нашем случае представляет программа PHYRE (Protein Homology/analogY Recognition Engine). Текущая версия 0.2. Данный сервис предоставляется «КАК ЕСТЬ» на страх и риск пользователя, Imperial College не предоставляет никаких гарантий и не берёт на себя никаких обязательств относительно работы сервиса и предоставленных им результатов (за их точность, полноту, качество и т. д.).

3 Белок VirE2 описание, функции

3. 1. Роль белка VirE2 в процессе трансформации растений

Агробактерии (*Agrobacterium tumefaciens*) являются постоянными обитателями корневой зоны растений. Эти бактерии проникают в ткань через поврежденные участки и размножаются в межклетниках растений. *A. tumefaciens*; содержит большую плазмиду, так называемую Ti- плазмиду (Ti – Tumor inducing, индуцирующая опухоль). Фрагмент плазмидной ДНК (Т-ДНК) интегрируется в хромосомную ДНК растительных клеток и вызывает их опухолевый рост. Круг хозяев агробактерий очень широк: они трансформируют клетки практически всех двудольных и однодольных растений. Интегрированная в состав генома растения Т-ДНК наследуется как простой доминантный признак в соответствии с законами Менделя.

Процесс образования опухолей у растений имеет теоретический интерес для исследователей, поскольку понимание механизма этого явления позволит понять некоторые общие биологические принципы: перенос ДНК через мембраны, интеграцию чужеродной ДНК в растительный геном, дифференцирование тканей, обмен сигналами между бактериями и растениями. В то же время, образование опухолей наносит значительный экономический ущерб для фруктовых деревьев, винограда и некоторых других растений (томатов, подсолнечника, табака, косточковых плодовых), что предопределяет интерес к агробактериальной трансформации практиков сельского хозяйства. По некоторым оценкам агробактерии являются третьим по значению фитопатогеном (Пирузян, 1988). Кроме того, это явление служит предметом пристального внимания, поскольку агробактериальная Т-ДНК является удобным природным вектором для переноса генов в растение. «Обезоруженная» (без генов продукции фитогормонов) Т-ДНК не образует опухоль на растении, но способна переносить различные гены, расположенные между правой и левой границами Т-ДНК и, встраивая в хромосому, изменять хозяйственно-полезные свойства растения. Использование трансгенных растений в сельском хозяйстве в последние годы быстро увеличивается, принося большую экономическую выгоду. Недавно описан перенос Т-ДНК в клетки дрожжей (Bundock et al., 1996). Это дает перспективу использования Т-ДНК для улучшения качества пива и хлебопекарных изделий. И, наконец, в феврале 2001 года группой американских исследователей представлены доказательства агробактериальной трансформации животных клеток (Kunik et al.,

2001). Последний факт открывает очень интересные перспективы использования агробактерий для биомедицинских исследований.

Несмотря на большую практическую значимость агробактериальной трансформации, её молекулярно-генетический механизм исследован не полностью.

Экспрессия генов вирулентности (*vir*) у агробактерий происходит под воздействием раневых экссудатов растений, которые рецептируются мембранным белком VirA (Stachel et al., 1986a,b). Сигнал проводится внутрь бактериальной клетки через двухкомпонентную систему белок VirA-белок VirG (Huang et al., 1990; Нооукаас et al., 1991). При участии активированных ацетосирингоном *vir*-генов продуцируется одонитевая, неполярная, линейная молекула Т-ДНК (Т-нить) (Stachel et al., 1986c, 1987). Сам *vir*-регион Тi-плазмиды, детерминирующий перенос, не переносится в хозяйскую клетку. Перенос Т-нити в растительную цитоплазму происходит с помощью VirD2 белка, где происходит формирование Т-комплекса: Т-нить покрывается несколькими сотнями агробактериальных белковых молекул VirE2 (Citovsky et al., 1989, 1994; Rossi et al., 1996; Tinland et al., 1993; Tinland, 1996). Перенос элементов Т-комплекса через мембраны бактериальной и растительной клеток осуществляется через *virB*-зависимый канал, при участии VirB1-VirB11, VirD4 белков (Fullner, 1998), но точный механизм переноса неизвестен. Комплекс Т-ДНК-VirD2 и VirE2 белок попадают в растительную клетку независимо (Gelvin, 1998) и, возможно разными каналами.

Экспорт белка VirE2 из агробактериальной клетки в растительную зависит от наличия белка VirE1 (Sundberg et al., 1996; Mysori et al., 1998). Белки VirE2 и VirE1 существуют в бактериальной клетке и экспортируются через наружную бактериальную мембрану в виде единого комплекса (Chen et al., 2000). Где, и при каких условиях происходит диссоциация VirE2/VirE1 комплекса неизвестно, но, вероятнее всего, в растительной клетке.

В контакте перед переносом Т-ДНК участвуют растительные рецепторы. Завершающие стадии процесса трансформации (встраивание Т-ДНК в растительную хромосому) происходят с участием растительных ферментов (Rossi, 1998). В растительной клетке с Т-ДНК считываются гены, ответственные за синтез фитогормонов, под воздействием которых происходит сдвиг гормонального баланса растительной клетки и формирование опухоли.

В результате индукции *vir* генов синтезируются белки вирулентности, которые формируют Т-комплекс. Вначале, после атаки эндонуклеаз

(белки VirD1 и VirD2) (Yanofsky et al., 1986; Stachel et al., 1987) происходит разрыв между 3 и 5 основаниями одной из нитей ДНК, которые используются для инициации и терминации синтеза ДНК при формировании линейной однонитевой копии ДНК, обозначаемой как Т-нить (Stachel et al., 1986с; Zambryski, 1992). Мутации в генах *virD1*, *virD2* приводят к нарушениям формирования Т-ДНК при образовании разрывов и продукции Т-нити. Высокий уровень экспрессии генов *virD1* и *virD2* приводит к повышению числа сформированных Т-комплексов и частоты трансформации (Wang et al., 1990). В виде Т-нити Т-ДНК существует только в бактериальной клетке. В настоящее время доказано, что белок VirE2 защищает Т-нить от атаки нуклеаз в растительной клетке и поддерживает линейную структуру Т-нити (Citovsky et al., 1989; Tinland, 1996; Christie, 1997). Белок VirE2 присоединяется к Т-ДНК не специфически, кооперативно и полностью покрывает всю нить.

Для покрытия одной цепи Т-ДНК требуется около 600 молекул VirE2 (Rossi et al., 1998). Количество белка VirE2 в клетке составляет до 0,05% от всего клеточного белка, или более 1000 молекул. Имеются данные, позволяющие предполагать, что комплекс Т-ДНК-VirD2 и белок VirE2 переносятся независимо в растительную клетку, и экспорт VirE2 зависит от VirE1 (Sundberg et al., 1996; Mysori et al., 1998). Белки VirE2 и VirE1 существуют в бактериальной клетке и экспортируются через наружную бактериальную мембрану в виде единого комплекса (Chen et al., 2000). Где, и при каких условиях происходит диссоциация VirE2/VirE1 комплекса неизвестно, но, вероятнее всего, в растительной клетке. Несмотря на высокую способность белка VirE2 связываться с Т-ДНК, в клетке бактерии белок VirE2 не покрывает Т-нить. Присоединение происходит только внутри растительной клетки. Вышеуказанный факт был подтвержден экспериментами с табаком. Трансгенные растения табака, экспрессирующие VirE2 белок, были чувствительны к заражению мутантом агробактерий по *virE2* гену (Citovsky et al., 1992).

Белок VirE2, помимо защиты однонитевой Т-ДНК, выполняет, по-видимому, какие-то другие функции, поскольку белок ResA, который может связываться с однонитевой ДНК, не может полностью заменить VirE2-белок. Эта функция необходима при переносе Т-ДНК через мембраны или для транспорта Т-комплекса в цитоплазме, поскольку ResA-белок может заменить VirE2-белок на этапе транспорта через ядерную пору в растительной клетке. VirE2-белок обнаруживается в цитоплазме индуцированных ацетосирингоном агробактерий (Christie, 1997). Небольшая, но значительная фракция VirE2-белка

обнаруживается во внутренней мембране, и незначительная часть обнаруживается во внешней мембране и периплазматическом пространстве (Christie, 1997).

В 2001 году Б. Хон с коллегами была обнаружена способность белка VirE2 взаимодействовать с липидными мембранами и формировать поры в них (Dumas et al., 2001). В этой работе с помощью биофизических методов авторы продемонстрировали образование канала в липидной мембране после взаимодействия VirE2-белка с мембраной.

Экспериментальное исследование способности белка VirE2 взаимодействовать с искусственной бислойной мембраной и с одноцепочечной Т-ДНК также было проведено группой М. И. Чумакова в 2004 году (Богданов и др., 2004). При изучении электропроводности плоских мембран после внесения препаратов белка VirE2 зарегистрировано скачкообразное увеличение проводимости мембран, которое может свидетельствовать о формировании одиночных долгоживущих (1,5-7 с) пор. Длительность раскрытия пор теоретически позволяет проходить за это время фрагментам ДНК, размер которых сравним с природной Т-ДНК агробактерий.

Скачки (или ступени) проводимости происходили лишь при достаточно малых (10-50 мВ) напряжениях и имели вид ступенек с резким фронтом, аналогично электрической активности липидных пор.

3.2. Механизмы формирования белковых пор в липидных мембранах

Проблемы формирования пор интересуют не только биологов. Огромное значение имеет проектирование искусственных пор, которые могли бы использоваться в терапии для уничтожения злокачественных клеток-мишеней при непосредственном или селективном проникновении в них цитотоксических препаратов (Ranchal et al., 1996) В этом направлении уже есть положительные результаты. Например, для *in vitro*-лечения опухолевых клеток в мышцах использовался проектированный пороформирующий белок SGP (Ellerby et al., 2003).

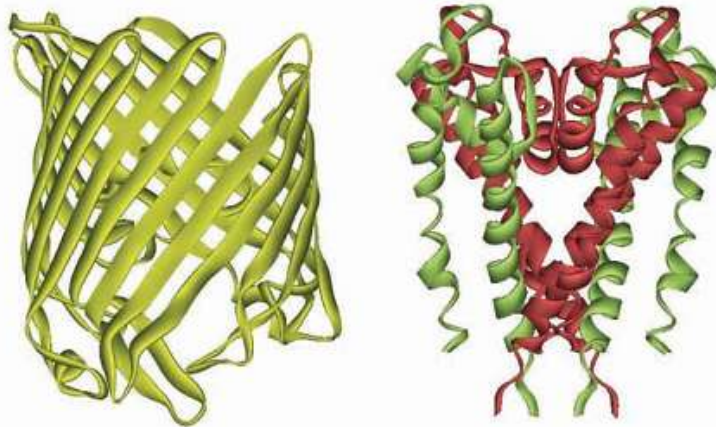


Рисунок 1.2.1. Два основных типа структур мембранных каналов и пор. Слева – β -складочная, тримерный порин OmpF(1BT9); справа – связка α -спиралей, калиевый канал KcsA(1BL8) из четырёх субъединиц, две из которых окрашены в зелёный, две – в красный (Bayeley, Jayasinghe, 2004)

На формирование водородных связей основными амидами внутри гидрофобного ядра липидного бислоя затрачивается большое количество энергии, поэтому большинство искусственных пороформирующих белков состоят из β -складок или связанных α -спиралей. Типичные структуры представлены на рисунке 1.2.1. (Bayeley, Jayasinghe, 2004)

Точный механизм образования пор белком VirE2 пока неизвестен, но не исключено, что он определяется одним из приведённых выше типом структур.

3.3.1 Примеры использования баз и программ для анализа порообразующих белков

3.3.1.1 Базы белков NCBI и PDB

Аминокислотные последовательности белков VirE2, VirE1 и VIP1 были взяты в базе центра National Center for Biotechnology Information (NCBI) по адресу <http://www.ncbi.nlm.nih.gov/> с идентификаторами NP_059819, NP_059818 и NP_564486 соответственно.

Для проверки корректности построенных моделей белков и их комплексов известные структуры белка фоторецептора *Bacteriorhodopsin* и комплекса инсулинового фактора роста IGF-I—IGFBP-5 и составляющих его субъединиц были взяты в базе белковых структур RCSB Protein Data Bank (PDB) по адресу <http://www.pdb.org>

3.3.2. Программа PHYRE

Трёхмерные модели рассматриваемых белков VirE2, VirE1 и VIP1 были построены при помощи ресурса PHYRE (доступного по адресу <http://www.sbg.bio.ic.ac.uk/~phyre/>), алгоритм которого основан на поиске совпадений при сравнении «профиль-профиль» (для профилей исследуемой и известной структур, полученных при помощи PSI-BLAST) совместно с поиском совпадений в предсказанной вторичной структуре.

3.3.3. Программы Deep View Swiss-PdbViewer

Состыковка структур белков VirE2 с VirE1 и VIP1 проводилась в программе для их визуализации Deep View Swiss-PdbViewer (самая свежая версия программы доступна в сети Интернет по адресу <http://www.expasy.ch/spdbv/text/download.htm>) при использовании инструмента Magic Fit.

4. Использование баз и программ для анализа порообразующих белков

4.1. Получение модели белка VirE2 по его аминокислотной последовательности

Белок VirE2 не имеет близких гомологов. По этой причине большинство ресурсов предсказания структуры на ввод последовательности белка VirE2 возвращают сообщения вида

Your sequence has no similar sequence of known structure¹ либо, как, например, SwissProt, выдают сообщение о том, что введённая последовательность сходна с последовательностью белка VirE2, но подробных данных о нём нет.

Некоторые программы, такие как Phyre, при отсутствии возможности построения модели всей белковой молекулы, возвращают неполную модель (лишь для некоторой части молекулы белка). При этом часто в официальной документации об этом не говорится, поэтому к полученной модели нужно относиться с осторожностью. Конкретного рецепта по решению данной проблемы нет, поэтому исследователю остаётся только руководствоваться своим здравым смыслом и физическим и биологическим чутьём.

Однако алгоритмы некоторых программ позволяют строить молекулярные модели без использования баз типичных белковых конфигураций. Подобно большинству программ, используемых химиками, таких как HyperChem или ChemSite, в их основе лежит принцип решения уравнений квантовой теории и нахождение состояния молекулы с наименьшей энергией. Полученная модель не всегда адекватна, т. к. может не учитывать особых условий среды, окружающей изучаемый белок. Тем не менее, и такая модель представляет ценность, потому что позволяет хотя бы грубо предположить структуру и свойства исследуемого белка. Контролем качества полученной модели может служить сравнение предсказанных свойств с экспериментально обнаруженными.

4.2. Методика построения и анализа модели белка VirE2

В данной работе была использована модель, построенная по последовательности белка VirE2 (взятой из базы NCBI по адресу <http://www.ncbi.nlm.nih.gov/>) при помощи программы PHYRE 0.2

Модель, построенная Phyre, основана на выравнивании исследуемой последовательности с известными структурами. Далее

¹ Сообщение от программы ESyPred3D с ресурса urbm.fundp.ac.be

координаты атомов из известной структуры копируются и переразмечаются в соответствии с исследуемой последовательностью. При этом может быть обнаружена отдалённая гомология, которая не обнаруживается традиционными методами. Потому что алгоритм RHYRE построен на поиске совпадений при сравнении «профиль-профиль» (для профилей исследуемой и известной структур, полученных при помощи PSI-BLAST) совместно с поиском совпадений в предсказанной вторичной структуре (см. <http://www.sbg.bio.ic.ac.uk/.../help.htm>)

Программа RHYRE возвращает результаты моделирования в виде текста на адрес электронной почты пользователя. Для дальнейшей работы с моделью необходимо сохранить присланный текст в простой текстовый документ (неформатированный текст в кодировке ASCII) и переименовать его, присвоив ему расширение *.PDB или *.ENT.

Одним из достоинств программы RHYRE является то, что модель строится даже в том случае, если смоделировать всю молекулу не представляется возможным, поэтому нужно внимательно читать комментарии (строки, начинающиеся с ключевого слова REMARK) в файле с полученной моделью. В нашем случае из 556-аминокислотной последовательности белка VirE2 был смоделирован лишь отрезок со 172-ой по 529-ую аминокислоты, т. е. чуть больше 64% цепочки, взятые почти по центру. Для визуализации полученной модели использована

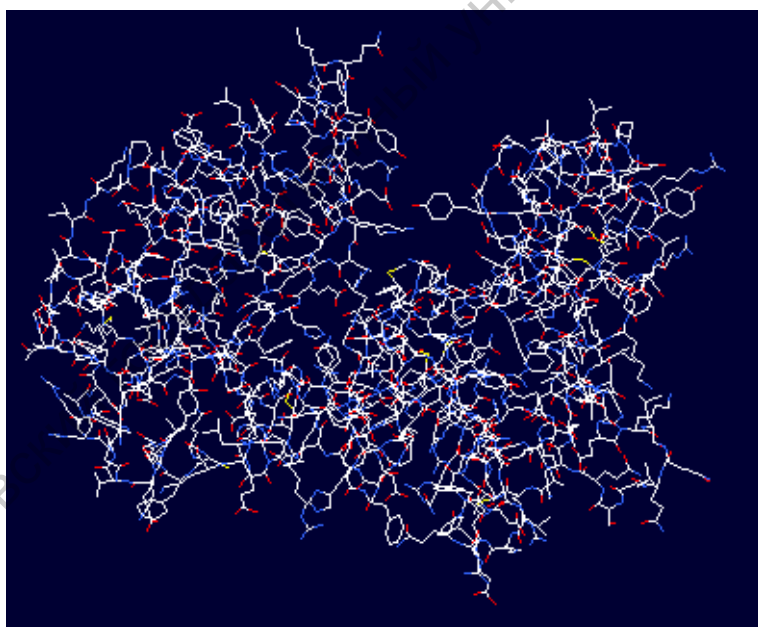


Рис. 4.2.1. Общий вид модели молекулы белка VirE2, построенной при помощи программы RHYRE, визуализированной в Deep View Swiss-PdbViewer при стандартном окрашивании

программа Deep View Swiss-PdbViewer (см. раздел 2.1. Swiss-Prot)

Для начала анализа нужно открыть полученный по электронной почте с ресурса RHYRE файл, содержащий трёхмерную модель белка (**File > Open PDB File...**).

Вид спереди по умолчанию (при окрашивании **Color > By CPK**) представлен на рисунке 4.2.1. Такой вид позволяет получить

общее представление о внешнем виде молекулы белка, но крайне

неудобен, поскольку не даёт информации ни о структуре молекулы, ни о свойствах её отдельных элементов

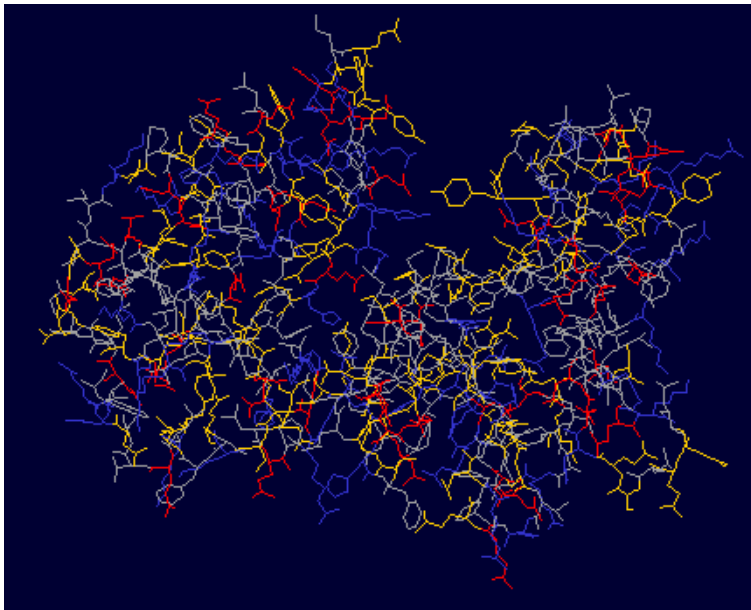


Рис. 4.2.2. Общий вид модели молекулы белка VirE2 при окрашивании аминокислотных остатков по типу

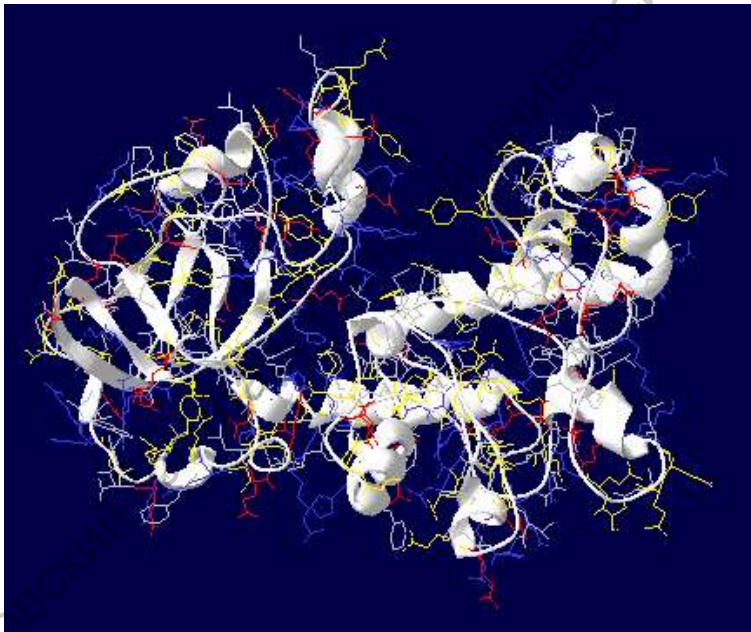


Рис. 4.2.3. Общий вид модели молекулы белка VirE2 после включения отображения её ленточной структуры

Для получения представления о свойствах отдельных элементов молекулы нужно применить окрашивание аминокислот по типу: **Color > by Type**. В результате остатки основных аминокислот окрашиваются синим, кислые – красным, полярные – жёлтым, а гидрофобные – серым (см. рис. 4.2.2.)

На практике очень редко рассматривают подробное рассмотрение атомов в белковой молекуле. Для получения сведений о структуре и функциях белка, как правило, используют так называемое ленточное представление. Чтобы рассмотреть ленточную структуру молекулы в панели **Control Panel** нужно поставить флажки **v** напротив каждой аминокислоты в столбце **ribn**.

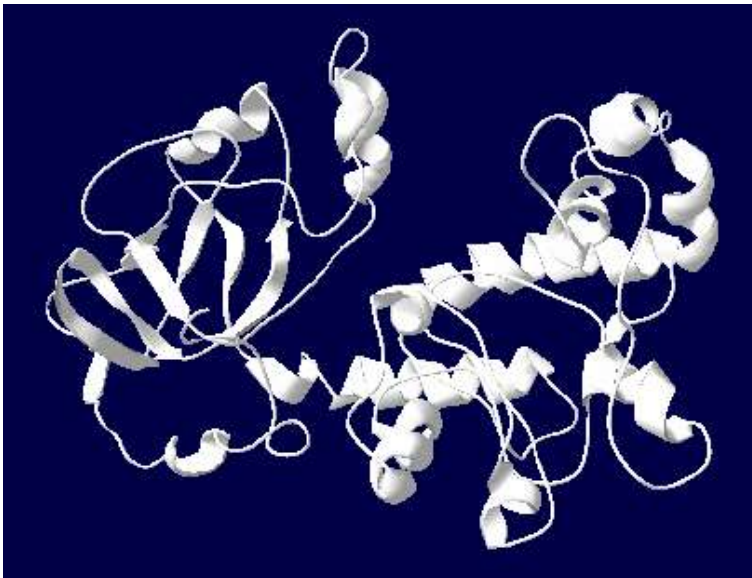


Рис. 4.2.4. Ленточный вид модели молекулы белка VirE2, построенной при помощи программы RHYRE, визуализированной в Deep View Swiss-PdbViewer

При рассмотрении структуры белковой молекулы подробности её атомного строения скорее мешают, чем помогают – нужно снять флажки **show** напротив всех аминокислот в панели управления. Результат этого действия представлен на рисунке 4.2.4

Из рисунка 4.2.4 видно, что структуры типа «складка» сгруппированы в одной части (при данном на рисунке угле обзора – слева). В то же

время основная часть спиралей – в противоположной.

При полимеризации присутствующие в исследуемой молекуле складочные или спиральные структуры могут образовывать поры (Bayeley, Jayasinghe, 2004).

Получив представление о структуре белковой молекулы, может возникнуть потребность в рассмотрении его молекулярной поверхности. Для выполнения соответствующего расчёта необходимо воспользоваться инструментарием программы Deep View: **Tools > Compute Molecular Surface**. Полученный результат представлен на рисунке 4.2.5.

Большинство современных ЭВМ оснащено видеоадаптером, поддерживающим работу с трёхмерной графикой, поэтому для улучшения визуального восприятия имеет смысл выполнить обработку, повышающую качество изображения – рендеринг (**Display > Render in solid 3D**). Рисунки 4.2.5 и 4.2.6 отличаются лишь тем, что на первом использована псевдотрёхмерная графика, а во втором – настоящая, с более естественным отображением теней и бликов.

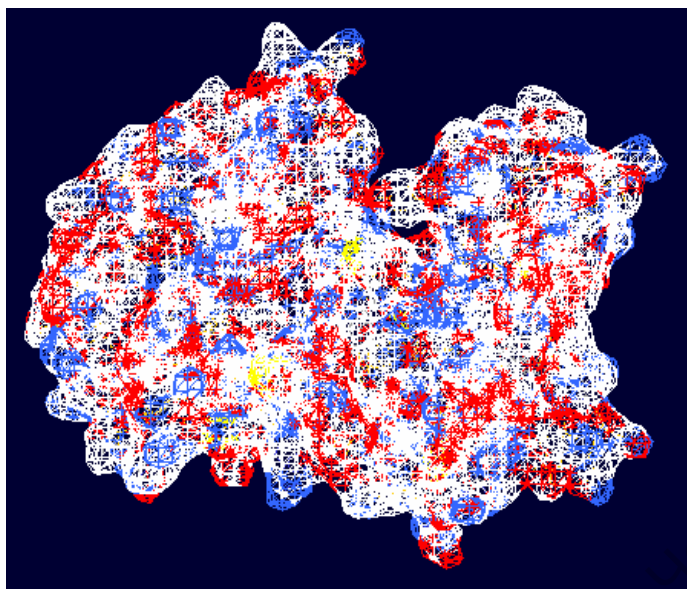


Рис. 4.2.5. Внешний вид поверхности молекулы белка VirE2, смоделированной при помощи программы RHYRE, визуализированной в Deep View Swiss-PdbViewer

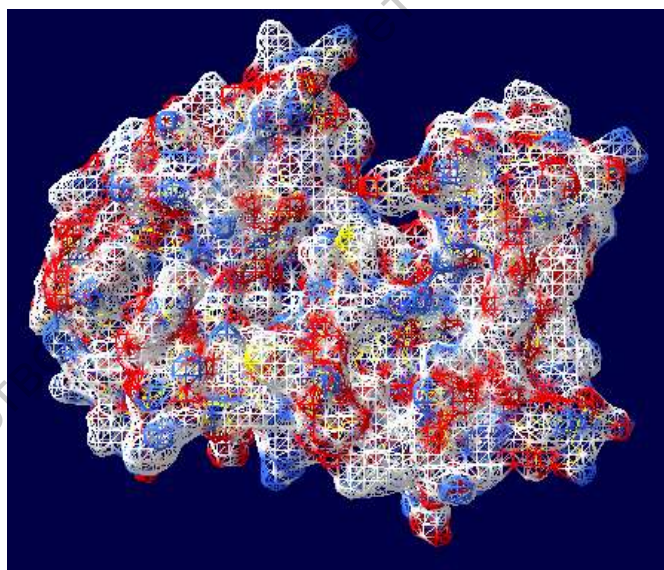


Рис. 4.2.6. Молекулярная поверхность белка VirE2, после применения рендеринга трёхмерного изображения

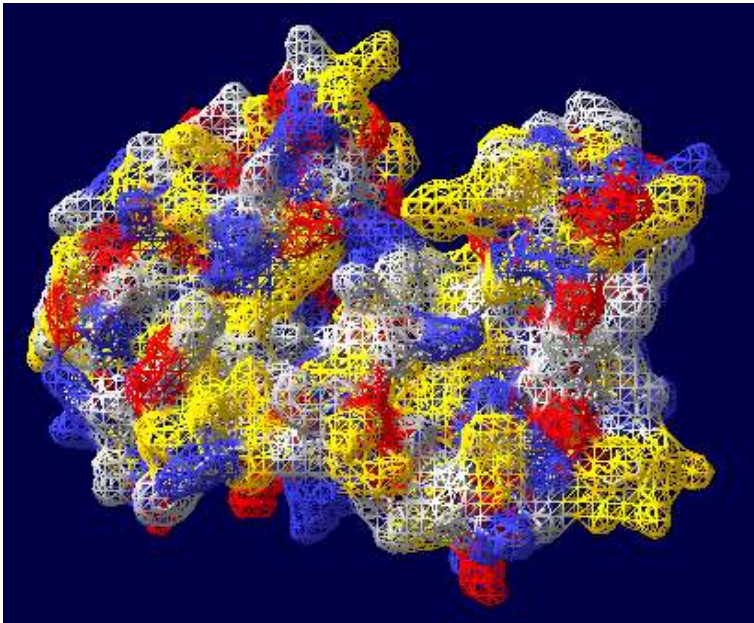


Рис. 4.2.7. Молекулярная поверхность белка VirE2 при окрашивании аминокислотных остатков по типу

окрашена по типу аминокислотных остатков. Таким образом, рассмотрев поверхность молекулы, можно прийти к выводу, что она не имеет четко выраженной структуры распределения типичных элементов – полярные и неполярные участки перемежаются между собой, как и кислотные с

основными.

4.3. Моделирование комплекса белков VirE1 и VirE2

Для того чтобы получить представление о структуре комплекса VirE1/VirE2, сначала была получена модель – белка VirE1

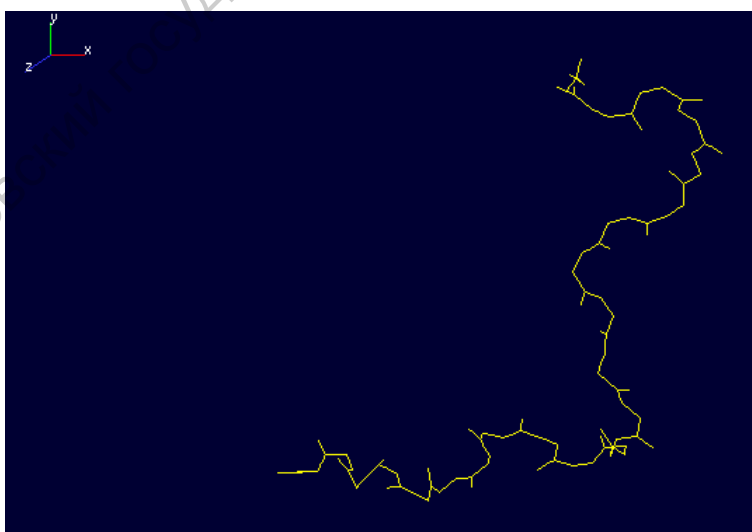


Рис. 4.3.1. Основная скелетная цепь модели молекулы белка VirE1, построенной при помощи программы PHYRE, визуализированной в Deep View Swiss-PdbViewer

По умолчанию поверхность молекулы окрашивается по СРК или распределению потенциала. Для того, чтобы окрасить её иным образом, нужно в **Control Panel** в режиме пятого столбца (кликнув по треугольнику ) выбрать пункт **molecular**, а в последнем столбце  выбрать **surface**. Теперь, меняя параметры меню **Color**, можно выбрать цветовую схему поверхности.

На рисунке 4.2.7 поверхность

(аминокислотная последовательность взята из базы NCBI по адресу www.ncbi.nlm.nih.gov/).

Из 83 аминокислот полипептида в модель, построенную при помощи PHYRE, вошли только 24: с 56-ой по 79-ую. Тем не менее, потеряв некоторую информацию о

свойствах, можно попытаться построить модель комплекса и попытаться проанализи-

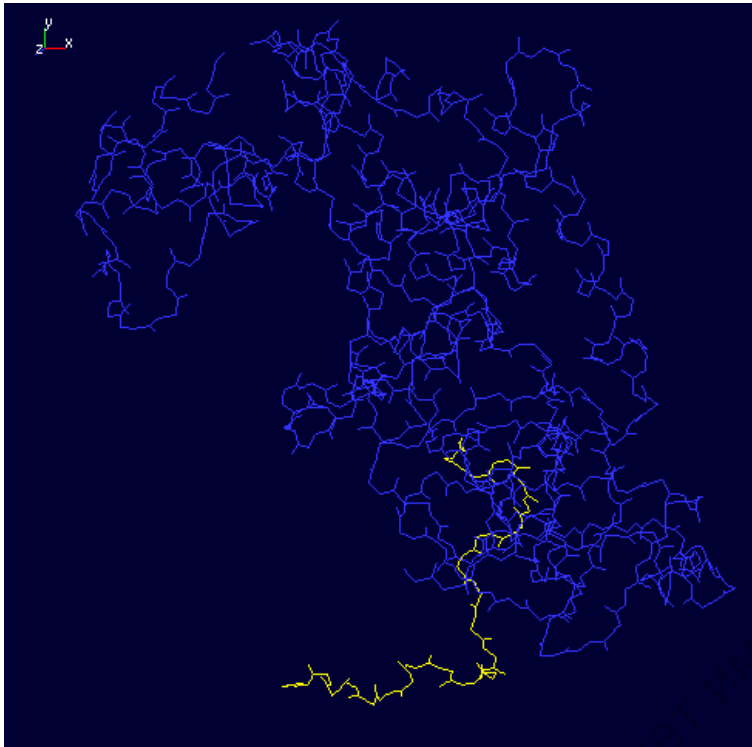


Рис. 4.3.2. Суперпозиция скелетных цепочек моделей молекул белков VirE1 (жёлтая), и VirE2 (синяя)

ровать его на качественном уровне.

Сначала в качестве первого слоя нужно открыть файл, содержащий модель белка VirE1 (**File > Open PDB File...**). Щелчком правой кнопки мыши по столбцу **sidechain** скрываются все боковые цепочки, оставив только основную. Далее применив окрашивание по слою (**Color > by Layer**) получается жёлтая цепочка, изображённая на рисунке 4.3.1.

Точно таким же образом нужно добавить белок VirE2 и состыковать оба белка (**Fit > Magic fit**) с параметрами по умолчанию. Результат произведённых действий приведён на рисунке 4.3.2.

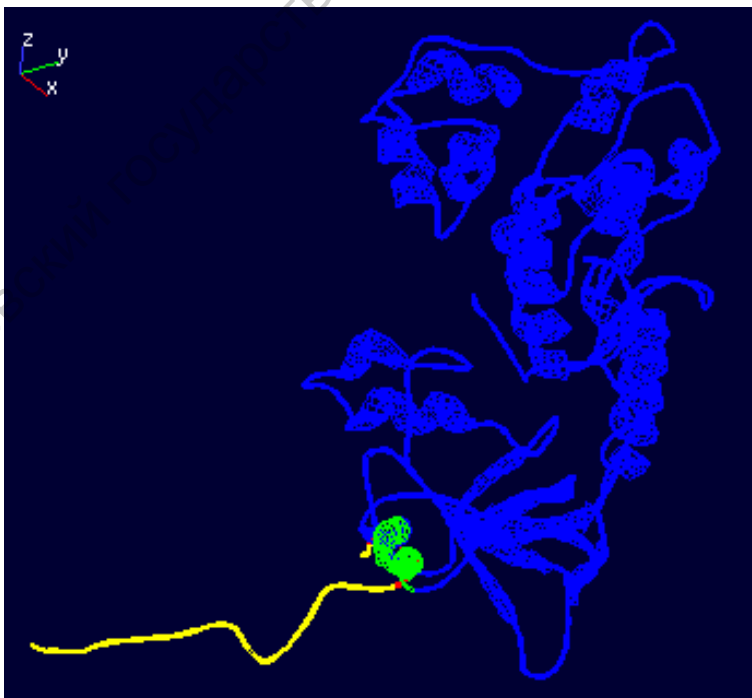


Рис. 4.3.3. Модель суперпозиции молекул белков VirE1 и VirE2 в ленточном представлении

Для удобства рассмотрения оба белка лучше перевести в ленточный вид. Для этого на панели **Layers Infos** нужно выбрать активный слой. Далее нужно обратиться к панели управления (**Control panel**): щелчок правой кнопкой мыши по столбцу **ribbon** включает отображение в виде ленты, по столбцу **show** выключает

отображение молекулярной цепочки, а щелчок левой клавишей мыши по самому правому столбцу вызывает меню, выбрав в котором **ribbon**, включается режим манипуляций над «лентой». С другим

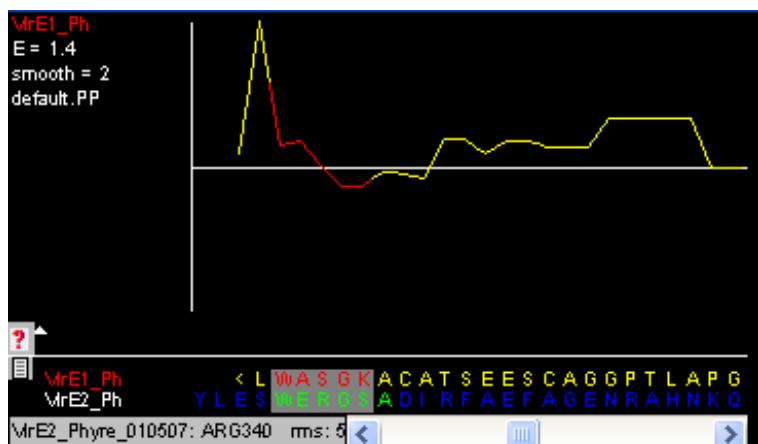


Рис. 4.3.4. Окно **Alignment** программы Deep View Swiss-PdbViewer, дающее графическое представление выравнивание последовательностей белков VirE1 и VirE2

слоем нужно повторить те же действия. Далее снова нужно применить окрашивание по слою и вручную подкрасить пересекающиеся области молекул контрастными цветами: ту часть VirE1, которая оказалась внутри – красным, а окутавшую её спираль VirE2 – зелёным (см. рис. 4.3.3).

Далее в меню **Window** нужно включить отображение окна выравнивания **Alignment**. В окне отображается выравнивание необработанных («сырых») последовательностей, TRP57 не соответствует GLU173. Необходимо сделать структурное выравнивание. Нужно выделить в строках выравнивания заранее подсвеченные подстроки (области пересечения) и выбрать в главной панели программы **Fit > Generate Structural Alignment**. Получается выравнивание «белок-белок», E=1.4, графическое его представление дано на рисунке 4.3.4., а текстовое (внешний вид оптимизирован исследователем) ниже:

VirE1_Ph	56	LWAS	GKACATSEES	CAGGPTLAPG	-
VirE2_Ph	301	DSKYLESWER	GSADIRFAEF	AGENRAHNKQ	FPAATVNMGR
		*	*	*	.

Далее возникает необходимость оценки сходства состыкованных участков. Критерием служит отклонение RMS, вычисляемое как

$$RMS = \sqrt{\frac{E(dist)^2}{nbAtoms}}$$

где *dist* – расстояние между парой связанных атомов, выраженное в ангстремах. Вычислив RMS (**Fit > Calculate RMS...**), с параметром **Backbone only** (учитывая только углеродный скелет белков и их C- и N-концы) получается, что во взаимодействии участвуют 20 атомов (nb atoms), а само **RMS deviation** равно 2.28 E.

Применив теперь окрашивание по RMS (**Color > by RMS**),

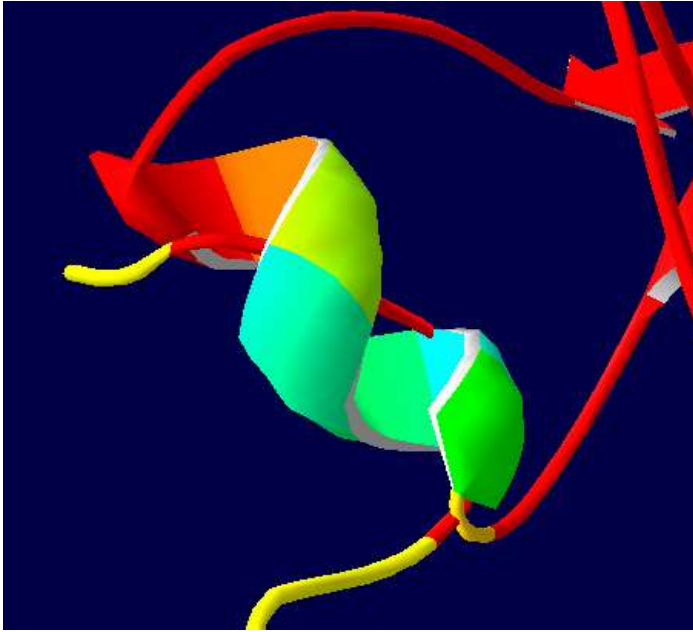


Рис. 4.3.5. Область взаимодействия молекул белков VirE1 (внутри) и VirE2 (окутывающая снаружи спираль)

получаются тёмно-синими те аминокислоты, которые наиболее точно соответствуют опорной структуре (белку, загруженному в качестве первого слоя – VirE1). По мере роста величины отклонения цвет меняется от синего к жёлтому и для структур, в которых нет ничего общего – красный. В результате области несущественного сходства ясно видны, т. к. окрашены красным.

Если ЭВМ исследователя оснащена графическим адаптером, поддерживающим трёхмерную графику, то, для

улучшения восприятия картинки можно применить рендеринг (**Display > Use OpenGL Rendering**) и, воспользовавшись инструментом увеличения изображения, получить область взаимодействия белковых молекул, изображённую на рисунке 4.3.5. Вид всей системы приведён на рисунке 4.3.6.

При проведении расчётов на маломощной машине рендеринг лучше отключать. Для улучшения качества совмещения белковых молекул можно также использовать опцию улучшения соответствия (**Fit > Improve Fit...**) и итерационное соответствие (**Fit > Iterative Magic Fit**), но в данном случае область касания молекул мала и любое из этих действий вызывает сообщение

Protein does not fit well; aborting.

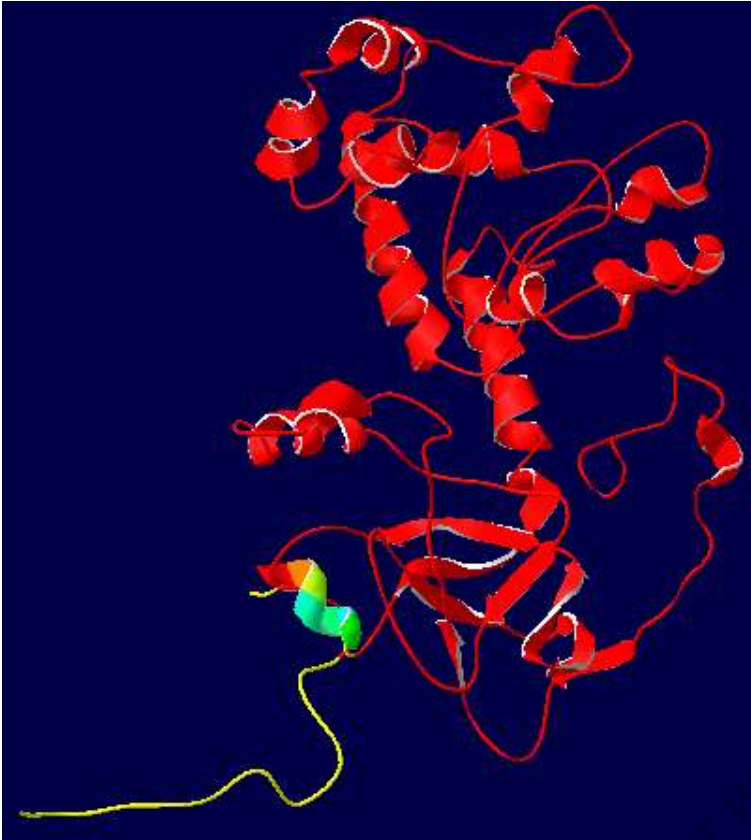


Рис. 4.3.6. Модель взаимодействия молекул белков VirE1 (опорный) и VirE2 – окрашенный в соответствии с величиной отклонения RMS: полному сходству соответствует синий цвет, который по мере роста величины отклонения меняется к жёлтому и для структур, в которых нет ничего общего – красный.

гидрофобные – серым. После применения рендеринга (**Display > Use OpenGL Rendering**), получится поверхность, изображённая на рисунке 4.3.7.

Из рисунка 4.3.7 видно, что поверхность основной части комплекса представлена полярными, кислыми и основными остатками – это может быть его гидрофильной частью. Также комплекс имеет небольшой, но ярко выраженный гидрофобный отросток.

Итак, получив модель комплекса двух белков, можно перейти к моделированию молекулярной поверхности. Для этого в меню **Tools** нужно выбрать пункт **Compute Molecular surface** применительно к каждому слою. Далее в самом правом столбце панели управления нужно выбрать **surface**, чтобы дальнейшие манипуляции проводить с молекулярной поверхностью. После

применения к каждому слою окрашивания по типу аминокислот (**Color > by Type**) основные аминокислоты окрашиваются синим, кислые – красным, полярные – жёлтым, а

Саратовский государственный университет

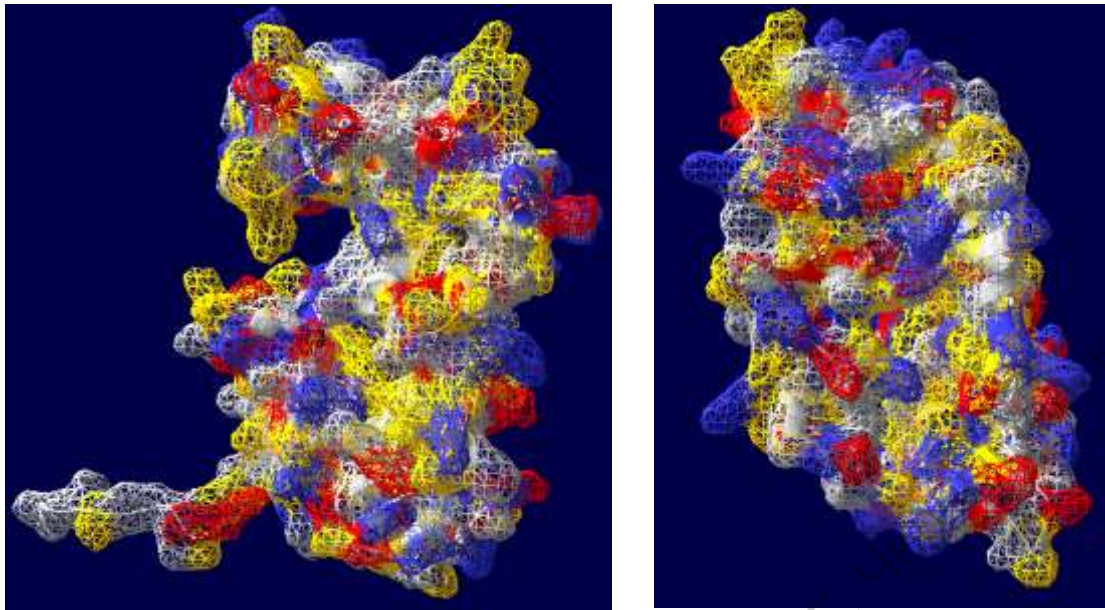


Рис. 4.3.7. Рассчитанная при помощи программы Deep View молекулярная поверхность комплекса белков VirE1 и VirE2 в плоскости отрезка (слева) и повернутая относительно вертикальной оси на 90° по часовой стрелке (справа). Аминокислотные остатки окрашены по типу.

4.4. Моделирование комплекса белков VirE2 и VIP1

Известно, что белок VirE2 взаимодействует с белком VIP1 (VirE2-interacting protein 1) растения *Arabidopsis thaliana* (Tzfira et al., 2001), первичную последовательность которого нетрудно найти в базе NCBI в локусе NP_564486. Структуру белка VIP1 можно найти в базе GWIDD под идентификатором **gwd21949**. Правда, получена она моделированием, а не рентгеноструктурным анализом, и является

неполной: из 341-аминокислотной последовательности смоделирован фрагмент со 196 по 248 (менее 16%), представляющий собой несколько изогнутую α -спираль (изображена белым цветом на рисунке 4.4.1).



Рис. 4.4.1. Ленточный вид модели молекулы белка VIP1 из базы GWIDD (ID gwd21949), визуализированной в Deep View Swiss-PdbViewer

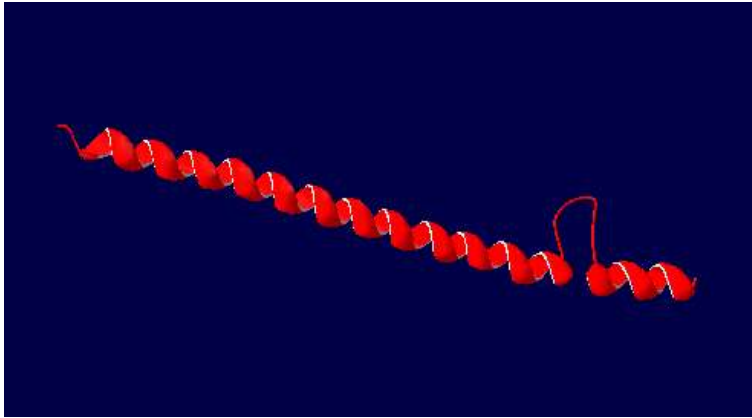


Рис. 4.4.2. Ленточный вид модели молекулы белка VIP1, построенной при помощи программы PHYRE, визуализированной в Deep View Swiss-PdbViewer

Фрагмент, который удалось смоделировать при помощи PHYRE не намного длиннее и покрывает со 196 по 261 остатки (менее 20%), но содержит не одну α -спираль, а две, причем неизогнутые (красная).

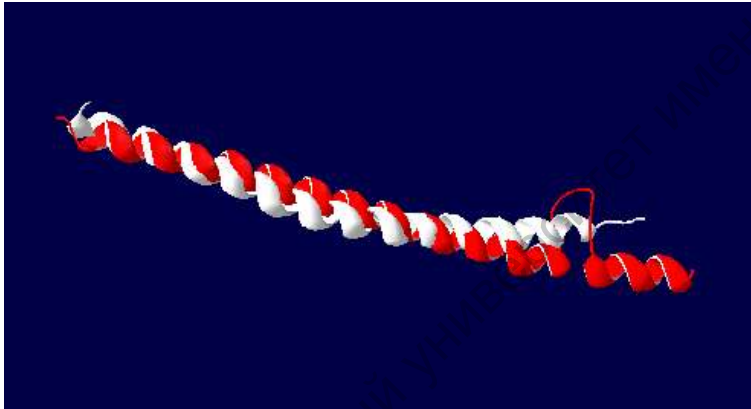


Рис. 4.4.3. Сравнение двух моделей молекулы белка VIP1: полученной из базы GWIDD (белая) и смоделированной при помощи PHYRE (красная)

В целом, полученные спирали похожи. В этом можно легко убедиться, совместив их при помощи функции **Magic Fit**.

В дальнейшем в качестве белка VIP1 будем рассматривать модель gwd21949, имеющую своеобразный изгиб.

Выполнить состыковку молекул VirE2 и VIP1 при помощи GWIDD и GRAMM-X не удалось: программы возвратили ошибки на этапах анализа исходных моделей и нахождения глобального минимума энергии.



Рис. 4.4.4. Модель суперпозиции молекул белков VirE2 и VIP1 в ленточном представлении.

Остается только вариант совмещения молекул внутри программы для просмотра моделей.

Сопоставив при помощи Magic Fit модели молекул VIP1 и VirE2

(отмечена на рисунке 4.4.4 жёлтым цветом), получим сплетение.

После совмещения белковых молекул их конформация не изменилась. Геометрию молекул нужно исправить, переместив атомы в соответствии с новыми образуемыми ими внутренними связями. Для этого воспользуемся инструментом минимизации энергии **Tools > Energy minimization**. Данный процесс выполняется при помощи встроенной в пакет Swiss-PdbViewer программы **GROMOS**, подробную информацию о которой можно найти на сайте <http://igc.ethz.ch/gromos/gromos.html> и у W.F. van Gunsteren et al. (1996) Тем не менее, Energy minimization подходит только для нахождения локальных связей, «потеснившихся» аминокислотных остатков, но не преодолевает высокие энергетические барьеры и останавливается на *локальном минимуме*.

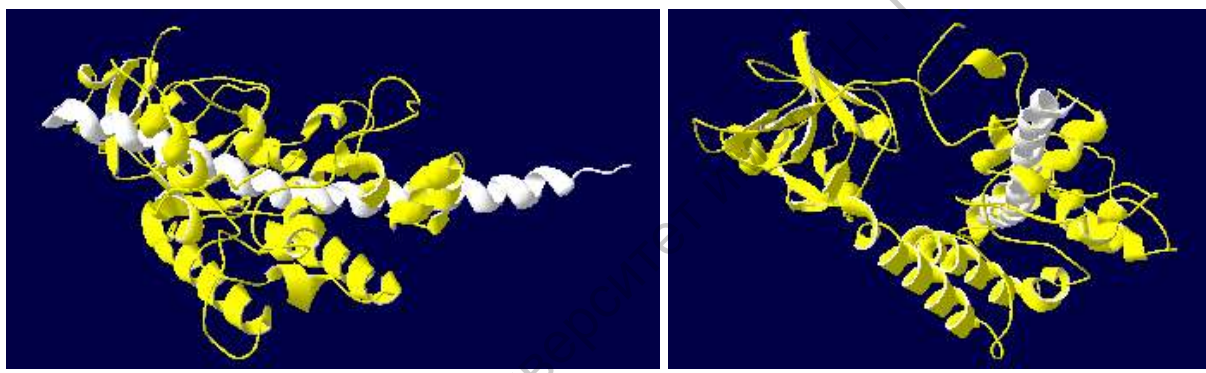


Рис. 4.4.5. Модель суперпозиции молекул белков VirE2 и VIP1 (вид с разных сторон) в ленточном представлении после выполнения операции минимизации энергии.

В документации к Swiss-PdbViewer четко оговорены ограничения, что при работе с силовыми полями для каждого остатку нужно учесть топологию, содержащую частичный заряд каждого атома, все связи, углы, торсии (спиральные закручивания) и так далее. Так как существуют тысячи возможных структур, предоставить топологии для каждой из них невозможно.

Далее можно оценить взаимодействие состыкованных молекул, например, оценить существующие между ними водородные связи. Провести такую оценку Deep View Swiss-PdbViewer позволяет только внутри одного слоя. Для дальнейшей работы нужно объединить «слои» молекул VirE2 и VIP1. Для этого необходимо выделить все аминокислотные остатки обоих исходных слоев и в меню **Edit** выбрать пункт **Create Merged Layer from Selection**. Далее нужно воспользоваться инструментом **Tools > Compute H-bonds**.

Водородные связи можно оценить даже в том случае, если в файле не присутствуют атомы водорода. Водородные связи обнаруживаются

из условия, что атом водорода находится на расстоянии от 1,2 до 2,76Å от «подходящего» донорного атома. Если в PDB-файле не содержатся координаты атомов водорода, водородные связи отмечаются, если расстояние между донором и акцептором лежит в пределах 2,35 до 3,2Å. Эти настройки можно изменить в настройках **H-bonds detection threshold**.

Чтобы увидеть рассчитанные водородные связи (зеленым пунктиром), нужно выбрать **Display > Show H-bonds**. Водородные связи не отображаются в ленточном представлении белков, поэтому нужно отключить Ribbon и включить отображение молекулярного остова.

При этом будут отображаться все водородные связи, нас же интересуют только связи между субъединицами VIP1 и VirE2. Чтобы выделить те водородные связи, которые связывают белую (VIP1) и желтую (VirE2) части, необходимо воспользоваться инструментом

увеличения картинки, поворачивая её.

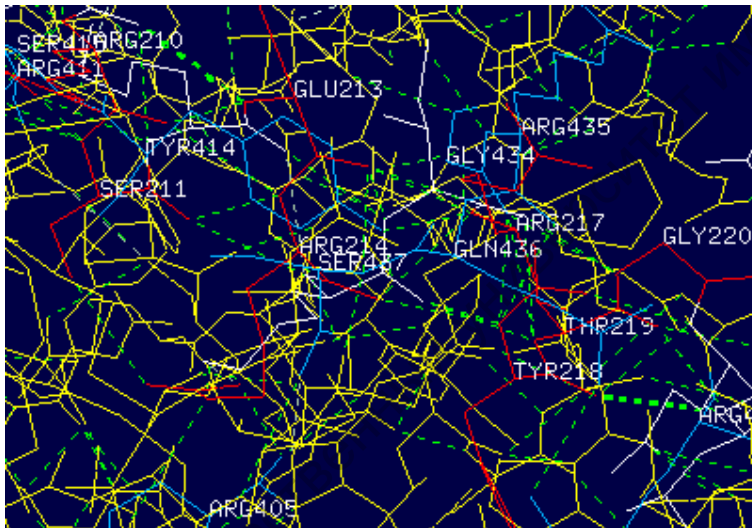


Рис. 4.4.6. Фрагмент модели суперпозиции молекул белков VirE2 и VIP1 в виде молекулярного остова с боковыми цепочками и отображением (зелёным пунктиром) всех водородных связей.

На рисунке 4.4.6 приведен пример такого рассмотрения. Активные элементы белой цепочки для контраста окрашены красным цветом, желтой цепочки – синим. Водородные связи отображены зеленым пунктиром. Определить нужные водородные связи в этой картине сложно, но можно. Искомые водородные связи (между красными

и синими элементами) выделены жирным пунктиром.

В результате внимательного рассмотрения модели были выявлены и представлены в таблице 4.4.1 аминокислоты (в формате трёхбуквенного обозначения с порядковым номером в первичной последовательности), за счет которых, возможно, происходит взаимодействие субъединиц комплекса белков VirE2 и VIP1.

Таблица 4.4.1. Группы аминокислот, образующие водородные связи при взаимодействии белков VirE1 и VirE2

VirE2	VIP1
-------	------

ARG185, GLU186	ILE201, LEU202, ARG205, GLN206
ARG405	ARG214
ARG411	ARG210, SER211
TYR414, SER415	ALA208, ARG217
GLU418	GLN206
VAL427, GLY434, ARG435, GLN436, SER437, ARG462	GLU213, ARG217, TYR218, THR219, GLY220
LEU504, GLN508, ALA512, GLN514, LEU515, GLU518, PRO519	ARG224, LEU229, GLN230, ASN231, THR235

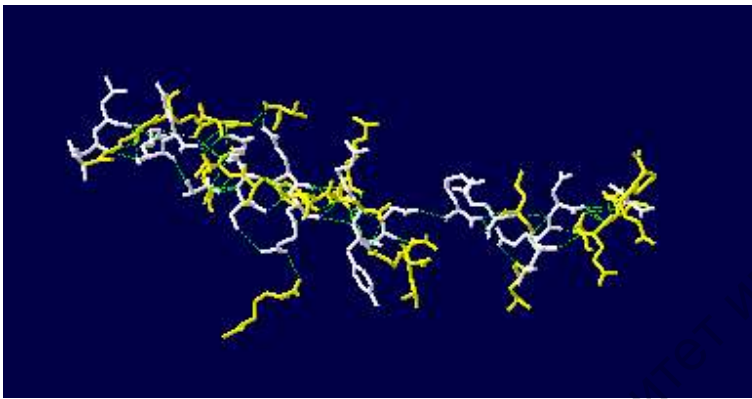


Рис. 4.4.7. Фрагменты модели суперпозиции молекул белков VirE2 и VIP1, отвечающие за их взаимодействие (только взаимодействующие аминокислоты)

Сами взаимодействующие аминокислоты (в виде остова с боковыми цепочками) и их взаимное расположение показаны на рис. 4.4.7, а вся модель комплекса VirE2-VIP1 с указанием водородных связей между субъединицами изображена на рис. 4.4.8.

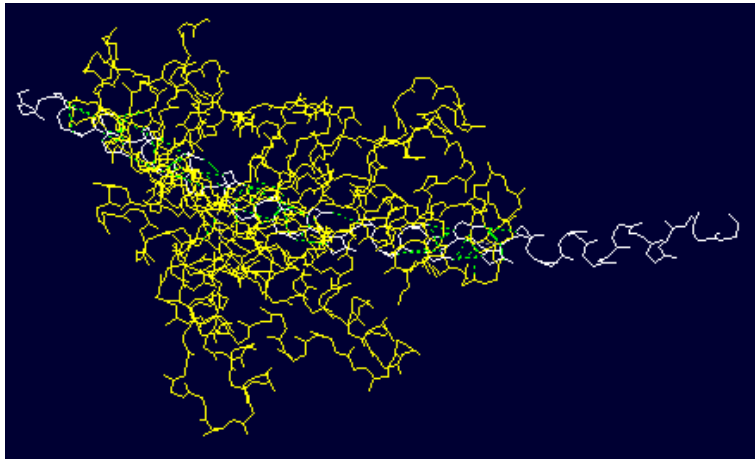


Рис. 4.4.8. Модель суперпозиции молекул белков VirE2 и VIP1 в виде молекулярного остова с отображением (зелёным пунктиром) водородных связей между субъединицами.

4.5. Оценка возможного взаимодействия VirE2 с Т-ДНК

Несмотря на высокую способность белка VirE2 связываться с Т-ДНК, в клетке бактерии белок VirE2 не покрывает Т-нить. Присоединение происходит только внутри растительной клетки, что было подтверждено экспериментами с табаком. Трансгенные растения табака, экспрессирующие VirE2 белок, были чувствительны к заражению мутантом агробактерий по VirE2-гену (Citovsky et al., 1992). Белки VirE2 и VirE1 существуют в бактериальной клетке и экспортируются через наружную бактериальную мембрану в виде единого комплекса (Chen et al., 2000). Где, и при каких условиях происходит диссоциация комплекса VirE1-VirE2 неизвестно, но, вероятнее всего, в растительной клетке. Тот факт, что связывание VirE2 Т-ДНК происходит после диссоциации комплекса VirE1-VirE2, позволяет предположить, что область взаимодействия VirE2 с VirE1 и Т-ДНК одна и та же.

Для оценки сайта связывания VirE2 с Т-ДНК рассматривалось взаимодействие VirE2 с белком VirE1.

Область взаимодействия белков VirE1 и VirE2 изображена на рис. 4.5.1 в том же ракурсе, что и на рис. 4.3.5, но с указанием взаимодействующих аминокислотных остатков и водородных связей между ними.

При детальном рассмотрении модели комплекса белков VirE1 и VirE2 в области их взаимодействия можно прийти к выводу, что этот

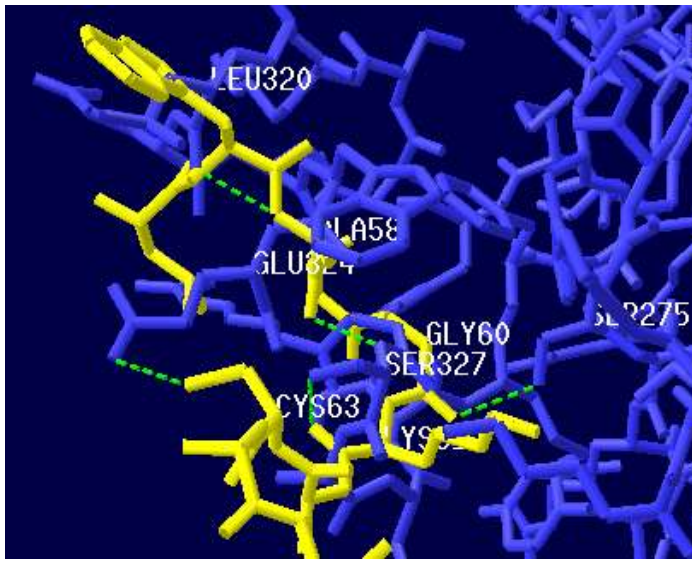


Рис. 4.5.1. Область взаимодействия белков VirE1 (желтый) и VirE2 (синий) в том же ракурсе, что и на рис. 3.3.5, но с указанием взаимодействующих аминокислотных

комплекс удерживается водородными связями между четырьмя парами аминокислотных остатков, которые приведены в таблице 4.5.1 (в формате трёхбуквенного обозначения с порядковым номером в первичной последовательности).

Таблица 4.5.1. Аминокислоты, образующие водородные связи при взаимодействии белков VirE1 и VirE2

VirE1	VirE2
ALA58	LEU320
GLY60	SER275
LYS61	SER327
CYS63	GLU324

4.6. Проверка корректности построенных моделей белков

Для проверки корректности моделей белков, построенных при помощи PHYRE, была построена тестовая модель – для белка *Bacteriorhodopsin*, структура которого известна. Для этого в базе белков RCSB Protein Data Bank по адресу <http://www.pdb.org> была взята последовательность и полученная рентгеноструктурным методом модель белка фоторецептора *Bacteriorhodopsin*, имеющего в базе идентификационный номер P02945 (сама модель в библиотеке структур имеет ID 1AP9). Загруженная в Deep View/Swiss-PdbViewer, модель имеет вид, представленный на рисунке 4.6.1.

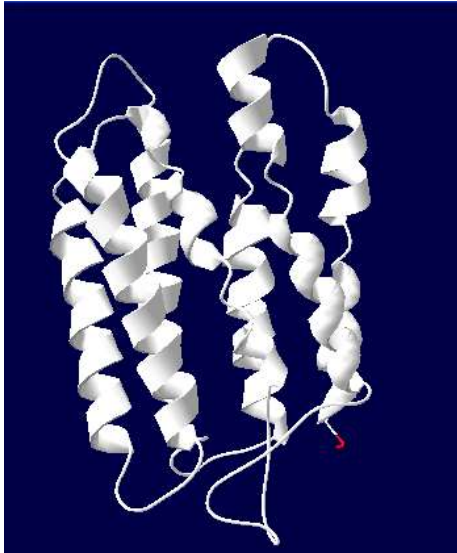


Рис. 4.6.1. Модель белка *Bacteriorhodopsin*, полученная методом рентгеноструктурного анализа

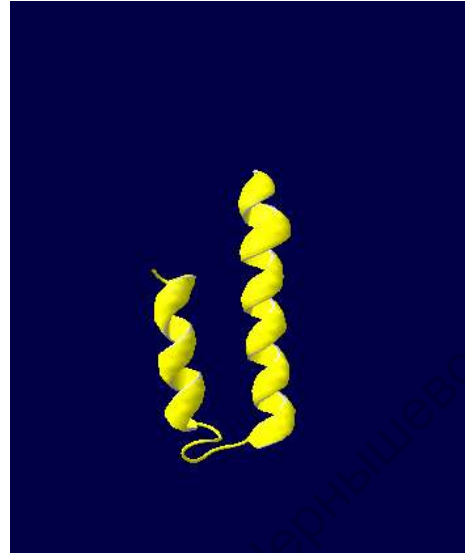


Рис. 4.6.2. Молекула белка *Bacteriorhodopsin*, смоделированная при помощи PHYRE

По полученной последовательности по описанной выше схеме в программе PHYRE была получена трёхмерная модель того же белка, изображённая на рисунке 4.6.2. При этом если в модель, полученную рентгеноструктурным методом, вошли 220 из 249 аминокислотных остатков (что составляет 88% последовательности), то программа PHYRE смогла смоделировать лишь отрезок со 180-ой по 221-ую аминокислоты (то есть лишь около 17% всей последовательности или около 19% полученной рентгеноструктурным методом). Программе PHYRE удалось смоделировать лишь 2 спирали из 11. Этого, конечно, мало для того, чтобы получить представление обо всей молекуле белка, но нужно оценить ещё одну характеристику построенной модели - её **точность**. Для этого нужно совместить оцениваемую модель с моделью, полученной рентгеноструктурным методом. Результат простой загрузки обеих моделей без использования их совмещения (**Fit**) представлен на рисунке 4.6.3.

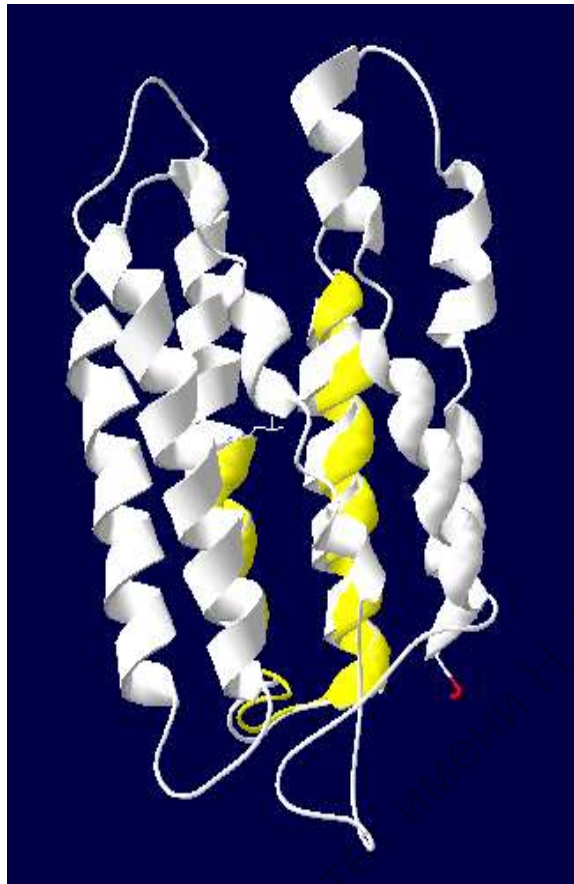


Рис. 4.6.3. Сравнение двух моделей молекулы белка *Bacteriorhodopsin*: полученной методом рентгеноструктурного анализа (белая) и смоделированной при помощи PHURE (жёлтая)

Как видно из рисунка 4.6.3. сходство очевидное, применение инструмента **Magic Fit** практически не меняет картины (спирали просто ещё плотнее прижимаются друг к другу простым параллельным переносом). Таким образом, можно сделать заключение о том, что получаемая модель хотя и не полная, но точная.

4.7. Проверка корректности построенных моделей белковых комплексов

Для того чтобы проверить, насколько точно полученные модели взаимодействия белков соответствуют реально существующим в природе, в базе белков RCSB Protein Data Bank по адресу <http://www.pdb.org> были взяты структуры белков IGF-I (ID 2DSP, получен рентгеноструктурным методом с разрешением 2,5 Å, цепь I, остатки со 2 по 68) и IGFBP-5 (ID 1BOE получен методом NMR, представлены 4 схожих модели, остатки с 13 по 58). Структура образуемого ими комплекса известна и имеет в той же базе

идентификатор 1H59 (от соматомедина IGF-I 49-118 остатки в цепи А, от белка IGFBP-5 13-58 остатки в цепи В).

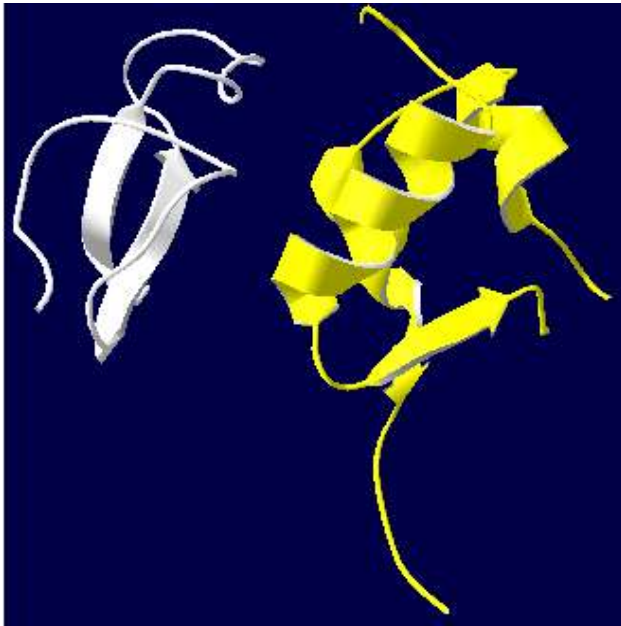


Рис. 4.7.1. Суперпозиция моделей белков IGF-I (жёлтый) и IGFBP-5 (белый).

Сначала с использованием описанных выше методик была построена изображённая на рис. 4.7.1 модель суперпозиции белков IGF-I (жёлтая субъединица) и IGFBP-5 (белая). Далее в Deep View/Swiss-Pdb Viewer была загружена модель реального комплекса IGF-I - IGFBP-5, приведённая на рис. 4.7.2 (цвета те же). Сопоставив рисунки 4.7.1 и 4.7.2, приходим к выводу, что построенная суперпозиция в общих чертах обнаруживает достаточное сходство с

известным комплексом.

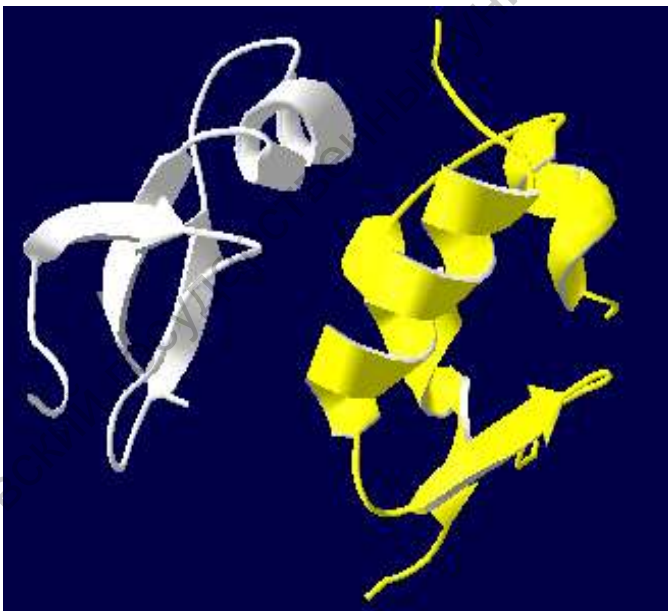


Рис. 4.7.2. Модель комплекса белков IGF-I (жёлтый) и IGFBP-5 (белый) полученная методом рентгеноструктурного анализа

4.8. Компьютерное и экспериментальное исследование свойств белка VirE2 формировать поры в биологических мембранах

Для проверки гипотезы о возможности образования канала при внесении VirE2 проведены эксперименты на плоском бислое (Dumas et al., 2001). После образования липидного бислоя добавляли очищенный белок VirE2, после чего по обе стороны мембраны накладывали разность потенциалов ~ 100 мВ. Наблюдалось постепенное увеличение проводимости, показывающее, что добавление VirE2 вызывало формирование различных каналов в мембране. Изучая 56 скачков проводимости, наблюдаемых в 7-ми независимых экспериментах, была найдена проводимость одного канала $1,5 \pm 0,15$ нСм. Различные эксперименты (Morgan et al., 1990; Shabala, Newman, 1995; Pouliquin et al., 1997) продемонстрировали, что большинство мембранных каналов открыто не всегда, а могут закрываться при приложении к мембране потенциала выше определённого граничного значения. Это явление, известное как *voltage gating* (потенциальная калитка), также наблюдалось для VirE2-каналов. В 100 проведённых экспериментах все каналы закрывались при трансмембранном потенциале 120 ± 20 мВ. Так как в растительной клетке этот потенциал изменяется в диапазоне 100-250 мВ (Newman, 1995; Pouliquin et al., 1997), то VirE2-каналы, если они существуют, вероятно, находятся в закрытом состоянии *in vivo*. Регуляция открывания канала на молекулярном уровне представляется возможной для растительной клетки.

Также имеются экспериментальные данные об ионной избирательности VirE2-канала. Эксперименты, проведённые в условиях асимметричной концентрации (0,1 М и 1 М КСl на цис- и транс-сторонах мембраны, соответственно) показали избирательность анионов в каналах. Относительная проницаемость (P_{Cl^-}/P_{K^+}), вычисляемая в соответствии с уравнением Голдмана-Хоткина-Катца (Hille, 2000), показала, что на каждый катион, проходящий через канал, приходится 9 анионов.

Так как диаметр такого анион-селективного канала велик, то он может осуществлять механизм переноса, то есть канал, возможно, взаимодействует с однострочной ДНК. Если на одну сторону бислоя добавить олигонуклеотидную 19-mer ДНК, то она может частично перекрывать канал, в зависимости от концентрации. Поток ионов через мембрану снижается приблизительно на 20%. Эксперименты, с использованием других олигонуклеотидов давали такие же результаты. Способность поры VirE2 взаимодействовать с однострочной ДНК чётко отличает этот белок от других каналов.

Интересно также отметить, что двухцепочечные молекулы ДНК оказались не способными проходить через VirE2-канал (Dumas et al., 2001).

Благодарности: Работа частично поддержана грантом 11-04-01331а (РФФИ), госконтрактами №№ 8592, 8728 (Министерства образования и науки РФ).

Я признателен Мазилу С.И. и Гусеву Ю.С. за участие в работе над белком VirE2, с использованием программ, указанных в данном пособии, что вылилось в написание и защите ими дипломов, кандидатской диссертации Мазилова С.И. и статей, указанных ниже.

Статьи по моделированию белка VirE2 и его комплексов:

1. Чумаков М. И., Мазилев С. И., Гусев Ю. С., Волохина И. В. Исследование способности агробактериального белка VirE2 к образованию пор в мембранах // Биомембраны 2010. т.27. №5. С.449-454.
2. Волохина И.В., Гусев Ю.С., Мазилев С. И., М.И. Чумаков Надмолекулярные комплексы белка вирулентности VirE2 *Agrobacterium tumefaciens* // Биохимия 2011. т. 76, вып.11. С.1576-1582.
3. Volokhina I.V., Gusev Yu. S., Mazilov S.I. Chumakov M.I. VirE2-dependent pores for ss-DNA transfer across artificial and cell membranes. // J.Bioinform.&Comput.Biol. 2012, Vol. 10, No. 2. 1241009. DOI: 10.1142/S0219720012410090

Литература

1. **Богданов В. И., Бурматов А. В., Ермошина О. В., Широков А. А., Великов В. А., Волохина И. В., Чумаков М. И.** Компьютерный и экспериментальный анализ способности белка VirE2 к образованию пор в липидных мембранах. In: Электронная конференция ИВТН-2004: "Информационно-вычислительные технологии в решении фундаментальных научных проблем и прикладных задач химии, биологии, фармацевтики и медицины" Москва, июль, 2004 (www.ivtn.com).
2. **Лахно В. Д.** Биоинформатика и высокопроизводительные вычисления // Вестник РФФИ, 2000, № 3 (21), С. 38-45.
3. **Пирузян Э. С.** Основы генетической инженерии растений, М.: Наука -1988, 303 с.
4. **Финкельштейн А. В., Рыкунов Д. С., Лобанов М. Ю., Бадретдинов Ф. Я., Рева Б. А., Скольник Дж.** Как и когда отдаленные гомологи могут превозмочь погрешности энергетических оценок и сделать возможным предсказание пространственных структур белков// Биофизика, 1999, № 44, С. 980-992
5. **Чумаков М. И.** Механизм агробактериальной трансформации растений, Саратов: Слово, - 2001, 256 с.
6. **Bayeley H and Jayasinghe L.** Functional engineered channels and pores (Review) // Molecular Membrane Biology, 2004, V. 21, P. 209-220
7. **Bundock P. and Hooкас P. J. J.** Integration of *Agrobacterium tumefaciens* T-DNA in *Saccharomices cerevisiae* genome by illegitimate recombination // Proc.Natl.Acad.Sci.USA, 1996, V. 93, P. 15272-15275.
8. **Christie P. J.** *Agrobacterium tumefaciens* T-complex transport apparatus: a paradigm for a new family of multifunctional transporters in *eubacteria* // J. Bacteriol., 1997, V. 179, P. 3085-3094.
9. **Chen L., Li C. M., Nester E. W.** Transferred DNA (T-DNA)-associated proteins of *Agrobacterium tumefaciens* are exported independently of *virB* // Proc.Natl. Acad.Sci.USA.2000, V. 97, P. 7545-7550
10. **Citovsky V., Warnick D., Zambryski P.** Nuclear import of *Agrobacterium* VirD2 and VirE2 proteins in maize and tobacco // Proc. Natl. Acad. Sci. USA. 1994, V. 91, P. 3210-3214.

11. **Citovsky V. C., Wong M. L., Zambryski P.** Cooperative interaction of *Agrobacterium* VirE2 protein with single-stranded DNA: implications for the T-DNA transfer process // Proc. Natl. Acad. Sci. USA, 1989, V. 86, P. 1193-1197
12. **Citovsky V., Zupan J., Warnick D., Zambryski P.** Nuclear localization of *Agrobacterium* VirE2 protein in plant cells // Science, 1992, V. 256, P. 1802-1805
13. **Cuff J. A., Clamp M., Siddiqui A., Finlay M., and Barton G. J.** JPred: a consensus secondary structure prediction server // Bioinformatics, 1998, V. 14, P. 892-983
14. **Dumas F., Duckely M., Pelczar P., Van Gelder P. and Hohn B.** An *Agrobacterium* VirE2 channel for transferred-DNA transport into plant cells // Proc. Natl. Acad. Sci. USA, 2001, V. 98, P. 485-490
15. **Ellerby H. M., Lee S., Ellerby L. M., Chen S., Kiyota T., del Rio G., Sugihara G., Sun Y., Bredesen D. E., Arap W. and Pasqualini R.** An artificially designed pore-forming protein with anti-tumor effects. // J. Biol. Chem., 2003, V. 278, P. 35311-35316
16. **Fischer D., Rice D., Bowie J. U., Eisenberg D.** Assigning Amino Acid Sequences To 3-Dimensional Protein Folds, //Faseb Journal, 1996, V. 10, P. 126-136.
17. **Fullner K. J.** Role of *Agrobacterium* virB genes in transfer of T complexes // J. Bacteriol. 1998, V. 180, № 2, P. 430-434.
18. **Garnier J., Osguthorpe D. J. and Robson B.** Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins // J. Mol. Biol., 1978, V. 120, P. 97-120
19. **Gelvin S.B.** *Agrobacterium* VirE2 protein can form a complex with T strand in the plant cytoplasm. // J. Bacteriol., 1998, V. 181, P. 4300-4302.
20. **Gibrat J.-F., Garnier J. and Robson B.** Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. // J. Mol. Biol., 1987, V. 198, P. 425-443
21. **Hille B.** Ionic Channels of Excitable Membranes // Proc. Natl. Acad. Sci. USA, 2000, V. 96, P. 235-241.
22. **Hooykaas P. J. J., Melchers L. S., Regensburg-Tuink A. J. G., den Dulk-Ras H., Rodenburg C. W., Turk S.** Signal transduction Via A and VirG in *Agrobacterium*. Advances in Molecular Genetics of Plant-Microbe Interactions // H.Hennecke and D. P. S. Verma, Kluwer Academic Publishers. 1991, P. 11-15.
23. **Huang Y.** VirA, coregulator of Ti-specified virulence genes, is phosphorylated *in vitro* // J. Bacteriol. 1990, V. 172, P. 1142-1144.

24. **Jones D. T.** Protein secondary structure prediction based on position-specific scoring matrices. // *J. Mol. Biol.*, 1999, V. 292, P. 195-202
25. **Kabsch W.** and **Sander C.** Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features // *Biopolymers*, 1983, V. 22. P. 2577-2637.
26. **King R. D.** and **Sternberg M. J. E.** Machine learning approach for the prediction of protein secondary structure // *J. Mol. Biol.*, 1990, V. 216, P. 441-457
27. **King R. D.** and **Sternberg M. J. E.** Identification and application of the concepts important for accurate and reliable protein secondary structure prediction.// *Protein Sci.*, 1996, V. 5, P. 2298-2310
28. **Krogh A., Brown M., Mian I. S., Sjolander K., Haussler D.** Hidden Markov models in computational biology. Applications to protein modeling // *J. Mol. Biol.*, 1994, V. 235, P. 1501-1531
29. **Kunik T., Tzfira T., Kapulnik Y., Gafni Y., Dingwall C., Citovsky V.** Genetic transformation of HeLa cells by *Agrobacterium* // *Proc. Natl. Acad. Sc.USA.* 2001, V. 98, P. 1871-1876.
30. **Levin J. M. and Garnier J.** Improvements in a secondary structure prediction method based on a search for local sequence homologies and its use as a model building tool. // *Biochim. Biophys. Acta*, 1988, V. 955, P. 283-295
31. **Morgan H., Lonsdale J. T. and Adler G.** Regulation of Ti plasmid virulence genes by a chromosomal locus of *Agrobacterium tumefaciens* // *J. Bacteriol.*, 1990, V. 164, № 2, P. 774-781.
32. **Mysori R. S., Bassuner D., Deng X. B., Darfinian N. S., Motchoulski A., Ream W.** Role of *Agrobacterium tumefaciens* VirD2 protein in T-DNA transfer and integration // *MPMI*, 1998, V. 11, № 7, P. 662-683.
33. **Ouali M. and King R. D.** Cascaded multiple classifiers for secondary structure prediction// *Protein Sci.*, 2000, V. 9, P. 1162-1176.
34. **Panchal R. G., Cusack E., Cheley S. and Bayley H.,** Tumor protease-activated, pore-forming toxins from a combinatorial library. // *Nat. Biotechnol.* 1996, V. 14, P. 852-856.
35. **Pouliquin P., Grouzis J. and Gibrat R.** Functional domain of *Agrobacterium tumefaciens* single-stranded DNA-binding protein VirE2 // *J. Bacteriol.* 1997, V. 179, P. 1165-1173.
36. **Rossi L., Hohn B., Tinland B.** Integration of complete transferred DNA units is dependent on the activity of virulence E2 protein of *Agrobacterium tumefaciens* // *Proc. Natl. Acad. Sci. USA*, 1996, V. 93, P. 126-130.

37. **Rossi L., Tinland B., Hohn B.** Role of virulence proteins of *Agrobacterium* in the plant// **In:** The Rhizobiaceae. Molecular biology of model plant-associated bacteria. 1998. P. 303-320 /Eds. H. P. Spaink, A. Kondorosi, P. J. J. Hooykaas. Kluwer Acad. Publish. Dordrecht, Boston, London
38. **Rost B. and Sander C.** Prediction of protein secondary structure at better than 70% accuracy. // J. Mol. Biol., 1993, V. 232, P. 584-599
39. **Salamov A. A. and Solovyev V. V.** Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments. // J. Mol. Biol. 1995, V. 247, P. 11-15
40. **Shabala S. N. and Newman I. A.** Intercellular *Agrobacterium* can transfer DNA to the cell nucleus of the host plant // Proc.Natl. Acad. Sci. USA, 1995, V. 92, P. 230-234.
41. **Stachel S. E., Nester E. W., Zambryski P. C.** A plant cell factor induces *Agrobacterium tumefaciens* vir gene expression. // Proc. Natl. Acad. Sci. USA. 1986a, V. 83, P. 379-383.
42. **Stachel S. E., Nester E.W.** The genetic and transcriptional organization of the vir region of the A6 Ti plasmid of *Agrobacterium tumefaciens* // EMBO J., 1986b, V. 5, P. 1445-1454
43. **Stachel S. E., Timmerman B., Zambryski P. C.** Generation of single-stranded T-DNA molecules during the initial stages of T-DNA transfer from *Agrobacterium tumefaciens* to plant cells // Nature, 1986c, V. 322, P. 706-711.
44. **Stachel S. E., Timmerman B., Zambryski P. C.** Activation of *Agrobacterium tumefaciens* vir gene expression generates multiple single-stranded T-strand molecules from the pTiA6 T-region: requirement for 5' virD product // EMBO J, 1987, V. 6, P. 857-863.
45. **Sundberg C. D., Meek L., Carroll K., Das A., Ream W.** VirE1 protein mediated export of the single-stranded DNA binding protein VirE2 from *Agrobacterium tumefaciens* into plant cells // J. Bacteriol, 1996, V. 178, No4, P. 1207-1212.
46. **Tinland B. and Hohn B.** Plant genome transformation by *Agrobacterium tumefaciens* is mediated by a nucleoprotein complex. /**In:** Virus strategies. Molecular biology and Pathogenesis. Eds W. Doerfler and P. Bohn. 1993, P. 351-359.
47. **Tinland B.** The integration of T-DNA into plant genomes// Trends of Plant Science 1996, V. 1, P. 178-184.
48. **Tzfira T., Vaidya M., and Citovsky V.** VIP1, an *Arabidopsis* protein that interacts with *Agrobacterium* VirE2, is involved in Vir2 nuclear import and *Agrobacterium* infectivity // EMBO J, 2001, V. 20, P. 3596-3607

49. **Van Gunsteren W. F.** et al. in *Biomolecular simulation: the GROMOS96 manual and user guide*. Vdf Hochschulverlag ETHZ, 1996
50. **Wang K., Herrera-Estrella A., Van Montagu M.** Overexpression of *virD1* and *virD2* genes in *Agrobacterium tumefaciens* enhances T-complex formation and plant transformation // *J. Bacteriol*, 1990, V. 172, № 8, P. 4432-4440.
51. **Yanofsky M. F., Porter S. G., Young C., Albrigh L. M., Gordon M., Nester E. W.** The *virD* operon of *Agrobacterium tumefaciens* encodes a site-specific endonuclease // *Cell*, 1986, V. 47, P. 471-477.
52. **Yi T.-M. and Lander E. S.** Protein secondary structure prediction using nearest-neighbor methods.// *J. Mol. Biol.*, 1993, V. 232, P. 1117-1129
53. **Zambryski P.** Chronicles from the *Agrobacterium*-plant cell DNA transfer story // *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 1992, V. 43, P. 465-490.

Использованные веб-сайты

1. <http://www.expasy.org/sprot/> - SwissProt, зеркала:
<http://www.ebi.ac.uk/swissprot/>, <http://ca.expasy.org/sprot/>,
<http://au.expasy.org/sprot/>, <http://br.expasy.org/sprot/>,
<http://cn.expasy.org/sprot/> и <http://kr.expasy.org/sprot/>.
2. <http://swissmodel.expasy.org/> - SwissModel
3. <http://www.expasy.ch/spdbv/> - Deep View Swiss-PdbViewer,
зеркало <http://www.genebee.msu.ru/spdbv/>
4. <http://www.expasy.ch/spdbv/text/gal.htm> - галерея изображений,
полученных с помощью Swiss-PdbViewer, альтернативный
ресурс <http://chemb16.leidenuniv.nl/~armand/>
5. <http://www.cse.ucsc.edu/> - официальный сайт Инженерной
школы Джека Бэскина (Университет Калифорнии, США),
основные биоинформационные сервисы которого
расположены по адресу
<http://www.cse.ucsc.edu/research/compbio/>
6. <http://cubic.bioc.columbia.edu/> - официальный веб-сайт Центра
биоинформатики кафедры биохимии и молекулярной
биологии Колумбийского университета, предоставляющий
PredictProtein:
http://cubic.bioc.columbia.edu/predictprotein/submit_def.html
7. <http://www.predictprotein.org> - выделенный сайт PredictProtein
Service - одного из крупнейших ресурсов анализа
последовательностей и предсказания функций и структуры
белков
8. <http://gwidd.bioinformatics.ku.edu/home> - база **Genome Wide
Docking Database** - ресурс полногеномного моделирования
белок-белковых взаимодействий; выполнить докинг двух
белков можно напрямую при помощи GRAMM-X Protein
Docking Web Server v.1.2.0 по прямой ссылке
<http://vakser.bioinformatics.ku.edu/resources/gramm/grammx/>
9. <http://www.sbg.bio.ic.ac.uk/> - официальный сайт Группы
структурной биоинформатики отделения молекулярных
биологических наук факультета естественных наук
(Великобритания)
10. <http://www.ncbi.nlm.nih.gov/> - База данных NCBI
11. <http://www.pdb.org> - база белков RCSB Protein Data Bank

Приложения

Приложение 1. Обозначения аминокислот

Название аминокислоты		Обозначение	
аланин	Alanine	A	Ala
цистеин	Cysteine	C	Cys
аспарагиновая кислота	aspartic (aspar D ic)	D	Asp
глутаминовая кислота	glutamic (glu E tamic)	E	Glu
фенилаланин	phenylalanine (F enylalanine)	F	Phe
глицин	Glycine	G	Gly
гистидин	H istidine	H	His
изолейцин	Isoleucine	I	Ile
лизин	lysine (буква перед L)	K	Lys
лейцин	Leucine	L	Leu
метионин	M ethionine	M	Met
Аспарагин	asparagi N e	N	Asn
пролин	P roline	P	Pro
глутамин	glutamine (Q -tamine)	Q	Gln
аргинин	a R ginine	R	Arg
серин	S erine	S	Ser
треонин	T hreonine	T	Thr
валин	V aline	V	Val
триптофан	tryptophan (tW yptophan)	W	Trp
тирозин	t Y rosine	Y	Tyr

