

Подсистема «Компьютерная протеомика»

Структура документа (оглавление)

1 Цель и задачи подсистемы «Протеомика».....	1
2 Структура подсистемы «Протеомика» и детальное руководство по ее применению.....	2
2.1 Программные компоненты подсистемы «Протеомика».....	2
2.1.1 Программная компонента BLAST: поиск гомологов в базах аннотированных последовательностей белков.....	2
2.1.2 Программная компонента PrositeScan: поиск мотивов и паттернов в первичных структурах белков.....	4
2.1.3 Программная компонента PDBSiteScan: распознавание функциональных сайтов в пространственных структурах белков.....	7
2.1.4 Программные компоненты Subloc и SubPrediction : предсказание внутриклеточной локализации белков прокариот и эукариот.....	10
2.1.5 Программная компонента PDBSiteDocking: реконструкция пространственных структур комплексов "белок-ион металла", "белок-низкомолекулярные лиганды", "белок-белок", "белок-ДНК/РНК".....	14
2.1.6 Программная компонента Veritope: предсказание В-клеточных эпитопов в аминокислотной последовательности.....	17
2.1.7 Программная компонента Cleavage: предсказание сайтов протеасомного протеолиза в аминокислотной последовательности.....	19
2.1.8 Программная компонента WebProAnalyst: предсказание мутаций, направленно меняющих биологические активности и свойства белков на основе анализа взаимосвязи между активностями белков и физико-химическими характеристиками аминокислот сайтов в выравнивании белковых последовательностей.....	21
3 Полезные ссылки.....	24

1 Цель и задачи подсистемы «Протеомика»

Подсистема «компьютерная протеомика» предназначена для решения задач комплексной структурно-функциональной аннотации белков и задач компьютерного дизайна модифицированных вариантов белков с использованием технологии конвейерного анализа. Программные компоненты подсистемы «Протеомика» ориентированы на структурно-функциональную аннотацию протеом бактерий, растений и других организмов, что позволит осуществлять реконструкцию сетей молекулярно-генетических взаимодействий, интерпретировать нуклеотидный полиморфизм на уровне структуры и функции белков, а также предсказывать мутации в белках, направленно изменяющие величину биологической активности белков, что имеет большое значение для дизайна белков с улучшенными медико-биологическими свойствами. Таким образом, задачи, решаемые с использованием подсистемы, являются важной составляющей в области молекулярной системной биологии, медицины, биотехнологии и биоинженерии.

Подсистема «Протеомика» ориентирована на решение следующих задач:

- 1) поиск гомологов в базах аннотированных последовательностей белков;
- 2) поиск мотивов и паттернов в первичных структурах белков;
- 3) предсказание внутриклеточной локализации белков;
- 4) распознавание функциональных сайтов в пространственных структурах белков, включая сайты каталитических центров ферментов, сайты посттрансляционных модификаций белков, сайты связывания ионов металлов, сайты связывания органических и неорганических лигандов, сайты связывания ДНК и РНК;

5) реконструкция пространственных структур молекулярных комплексов, включая: реконструкцию комплексов "белок-ион металла", "белок-низкомолекулярные лиганды", "белок-белок", "белок-ДНК/РНК";

6) предсказания иммунологических характеристик белков, включая предсказание сайтов конститутивного и иммунопротеасомного протеолиза, предсказание В-клеточных эпитопов;

7) предсказания мутаций, направленно меняющих биологические активности и свойства белков на основе анализа взаимосвязи между активностями белков и физико-химическими характеристиками аминокислот сайтов в выравнивании белковых последовательностей.

2 Структура подсистемы «Протеомика» и детальное руководство по ее применению

В подсистему «Компьютерная протеомика» входят следующие компоненты:

1) компонента, осуществляющая поиск гомологов в базах аннотированных последовательностей белков;

2) компонента, осуществляющая поиск мотивов и паттернов в первичных структурах белков;

3) компонента, осуществляющая распознавание функциональных сайтов в пространственных структурах белков;

4) компонента, осуществляющая предсказание внутриклеточной локализации белков;

5) компонента, осуществляющая реконструкцию пространственных структур молекулярных комплексов белок-лиганд;

6) компонента, осуществляющая предсказание иммунологических характеристик белков;

7) компонента, осуществляющая предсказание мутаций, направленно меняющих биологические активности и свойства белков.

2.1 Программные компоненты подсистемы «Протеомика»

Ниже будут рассмотрены компоненты подсистемы «Протеомика» и приведено детальное (пошаговое) описание использования этих компонентов.

2.1.1 Программная компонента BLAST: поиск гомологов в базах аннотированных последовательностей белков

Для любых двух белков можно определить степень сходства их аминокислотных последовательностей – уровень гомологии. Два белка, имеющих уровень гомологии, превышающий пороговое значение, являются гомологами. Белки-гомологи обычно имеют сходное происхождение и большей частью принадлежат к одному семейству. Часто белки, выполняющие одну и ту же функцию, являются гомологами, обратное же утверждение справедливо лишь в некоторой степени.

Поиск гомологов заданного белка позволяет охарактеризовать его, опираясь на существующие знания об уже изученных белках. Таким образом, поиск гомологов может применяться, в частности, для определения семейства, к которому принадлежит вновь открытый белок, что дает сведения об его возможных функциях в клетке. Так же определение набора гомологов заданного белка делает возможным исследование молекулярной эволюции заданного белка и установление направления изменения его биологических активностей в эволюционной истории.

Поиск гомологов для заданной пользователем последовательности осуществляется в базе данных SWISS-Prot. На вход программного модуля, реализующего поиск гомологов

подается последовательность и указывается тип организма, грам-положительная или грам-отрицательная бактерия. Результат BLAST и PsiBLAST содержит идентификаторы (ID) SWISS-Prot записей для найденных гомологий. С использованием этих ID для каждой из гомологичных последовательностей извлекаются их карточки из базы данных SWISS-Prot. Из карточки экстрагируются тип организма и внутриклеточная локализация и отбираются только те карточки, указанный организм в которых соответствует организму, введенному пользователем. Таким образом формируется список потенциальных локализаций для анализируемой последовательности.

Программа предназначена для нахождения белков, аминокислотная последовательность которых гомологична последовательности заданного белка [1].

2.1.1.1 Работа с программной компонентой BLAST

Для подготовки к выполнению операции, соответствующей данной программной компоненте, надо войти в подсистему «Компьютерная протеомика» и с помощью мыши кликнуть надпись «Поиск гомологов в базах аннотированных последовательностей белков» в списке операций этой подсистемы. В результате появится HTML-страница с интерфейсом этой операции, который показан на Рис.1.

Выполнение операции начинается с ввода аминокислотной исследуемого белка. Для этой цели используется раздел “Введите последовательность белка в FASTA формате” HTML-интерфейса операции, как это показано на Рис.2. Последовательность может быть набрана в поле экрана для ввода последовательности или скопирована с использованием стандартных операций Интернет-браузера.

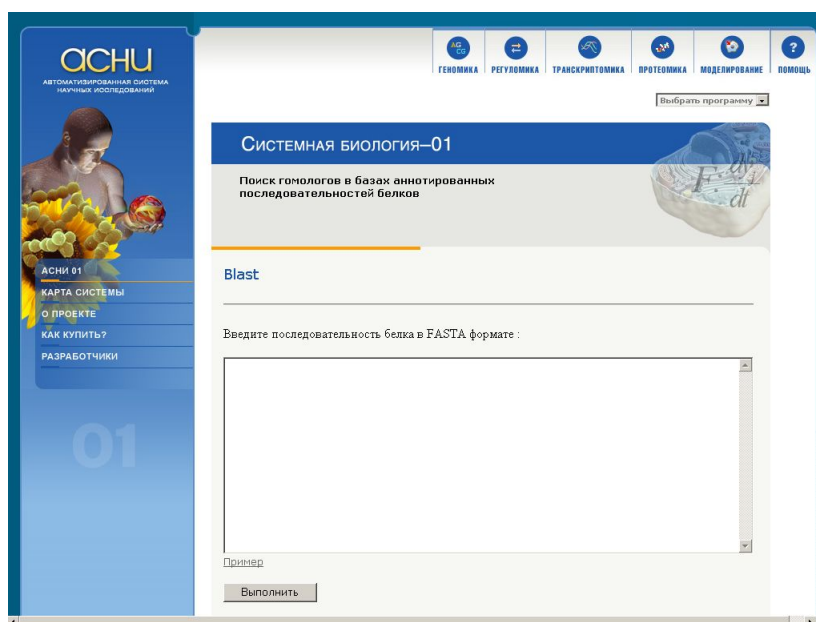


Рисунок 1. HTML-интерфейс операции, подготовленный к началу ее выполнения.

После ввода данных как было описано выше необходимо нажать на кнопку “Выполнить” (Рис. 2), в результате чего появляется HTML-страница с результатом поиска гомологов в базе NR NCBI BLAST для белковых последовательностей, выполненного программой NCBI BLAST, как это показано ниже на Рис.3.

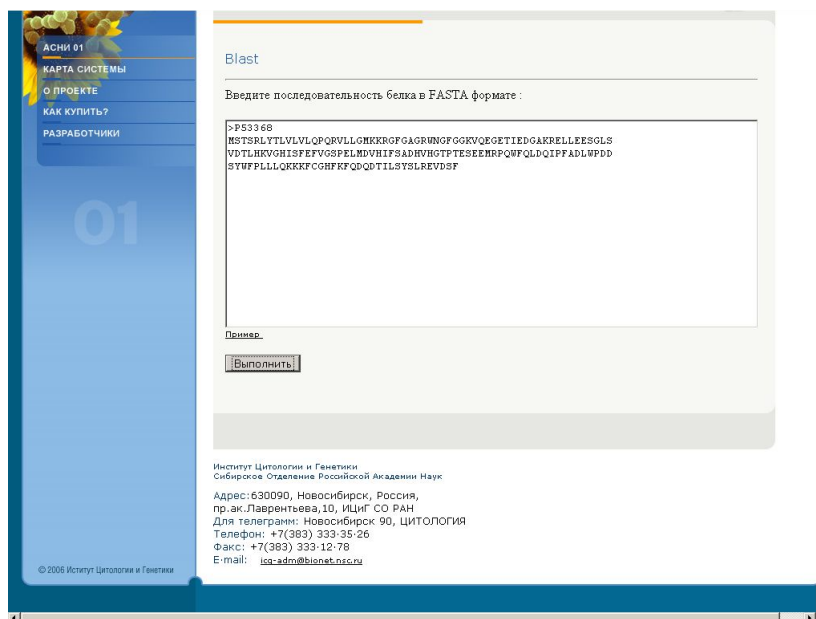


Рисунок 2. Ввод аминокислотной последовательности белка.

С помощью стандартных средств Интернет браузера пользователь может сохранить это предсказание в виде текстового или html-файла, например, с целью накопления таких предсказаний для разных белковых семейств и их последующего сравнительного анализа (Рис. 3).

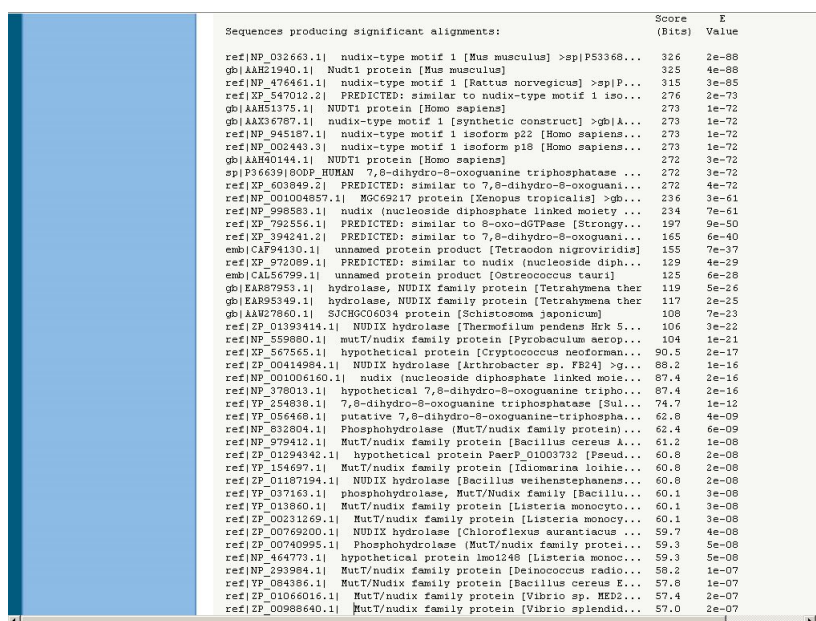


Рисунок 3 Пример HTML-страницы, содержащей результат поиска гомологов для введенной последовательности.

2.1.2 Программная компонента PrositeScan: поиск мотивов и паттернов в первичных структурах белков

Группа аминокислотных последовательностей белков, сходных между собой, но не идентичных, может быть представлена в виде одной общей символической последовательности – паттерна. Для представления группы последовательностей в виде паттерна каждый аминокислотный остаток одной последовательности должен быть, по возможности, поставлен в соответствие какому либо остатку каждой из других

последовательностей. Обычно это достигается посредством выравнивания заданных последовательностей. Алфавит, из символов которого составляются паттерны, включает в себя символы обозначений 20-ти аминокислотных, входящих в белки, а также ряд специальных символов, описывающих вариацию аминокислотных остатков в заданном участке описываемых последовательностей.

Среди похожих белков степень вариации аминокислот, составляющих сайт с определенной функцией, изменяется от одной аминокислотной позиции к другой. Некоторые аминокислоты должны оставаться одними и теми же во всех сходных сайтах, другие могут варьировать в узких пределах, третьи могут быть заменены практически без вреда для функции сайта. Эту структуру функционального сайта можно описывать с помощью паттернов первичных последовательностей белков [2].

Таким образом, программа поиска мотивов и паттернов в первичных структурах белков может применяться в биоинформатике для функциональной аннотации аминокислотных последовательностей белков и изучения структуры функциональных сайтов в белках.

2.1.2.1 Работа с программной компонентой PrositeScan

Для подготовки к выполнению операции надо войти в подсистему «Компьютерная протеомика» и с помощью мыши кликнуть надпись «Поиск мотивов и паттернов в первичных структурах белков» в списке операций этой подсистемы. В результате появится HTML-страница с интерфейсом этой операции, который показан на Рис.4.

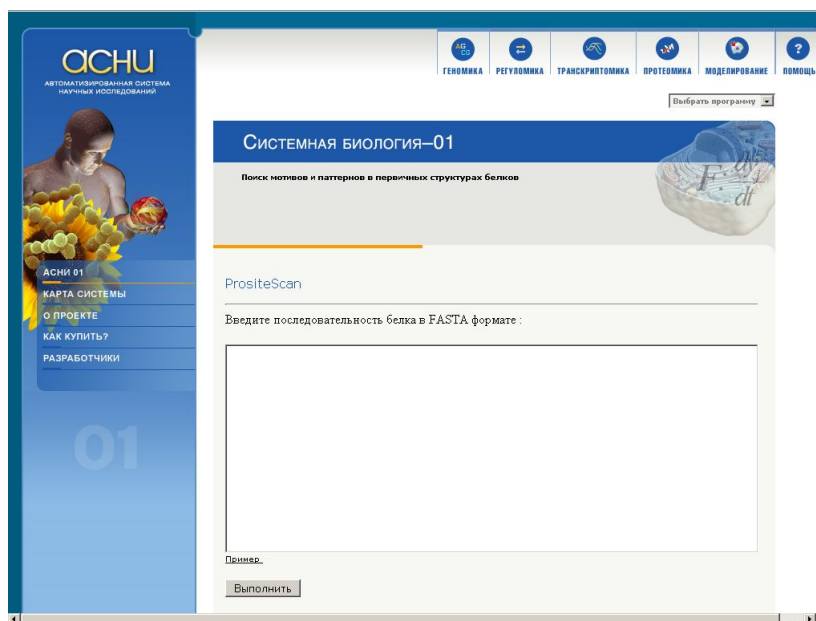


Рисунок 4. HTML-интерфейс операции, подготовленный к началу ее выполнения.

Выполнение операции начинается с ввода аминокислотной исследуемого белка. Для этой цели используется раздел “Введите последовательность белка в FASTA формате” HTML-интерфейса операции, как это показано на рисунке ниже. Последовательность может быть набрана в поле экрана для ввода последовательности или скопирована с использованием стандартных операций текстового редактора Интернет-браузера (Рис 5).

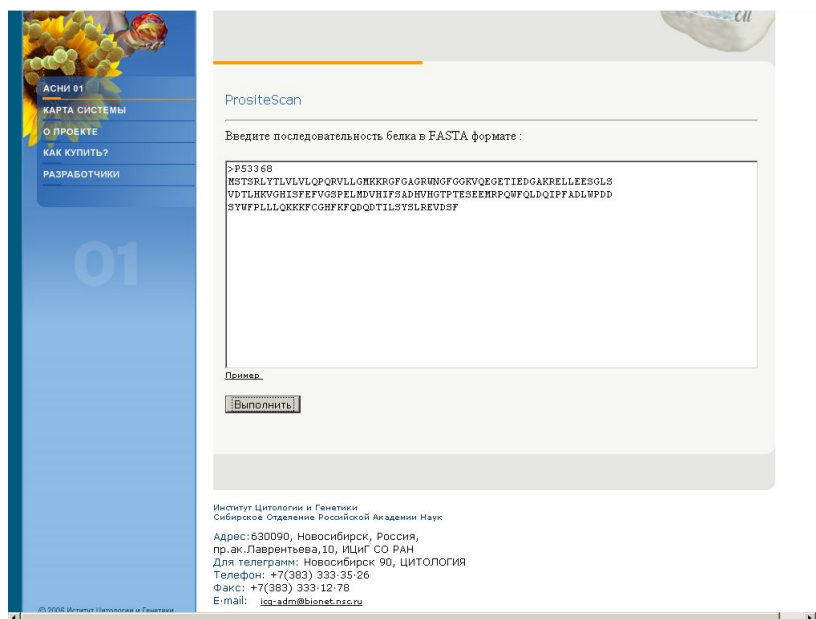


Рисунок 5 Ввод аминокислотной последовательности белка.

После ввода данных как было описано выше необходимо нажать кнопку “Выполнить”, в результате чего появляется HTML-страница с результатом поиска мотивов и паттернов в первичной структуре белка, как это показано ниже на рис.6.



Рисунок 6 Пример HTML-страницы, содержащей результат поиска мотивов и паттернов для введенной последовательности.

С помощью стандартных средств интернет-браузера пользователь может сохранить это предсказание в виде текстового или html-файла, например, с целью накопления таких предсказаний для разных белковых семейств и их последующего сравнительного анализа.

Для завершения работы с текущей (введенной) последовательности белка и перехода к обработке другого белка следует воспользоваться стандартными средствами интернет браузера для возвращения на предыдущую страницу, в результате чего осуществляется подготовка к повторному выполнению этой операции.

2.1.3 Программная компонента PDBSiteScan: распознавание функциональных сайтов в пространственных структурах белков.

Отдельные участки пространственной структуры белка необходимы для выполнения им конкретных молекулярных функций, составляющих его биологическую активность. Такие участки носят название функциональных сайтов. Все функциональные сайты разделяются на три группы – активные центры ферментов, сайты связывания лигандов или биополимеров, сайты посттрансляционной модификации белков. Биологическая активность белка во многом определяется набором его функциональных сайтов.

Проблема распознавания функциональных сайтов занимает одно из центральных мест в протеомике. Можно выделить несколько основных подходов к решению этой задачи. Первый из них основан на анализе первичных структур. На нем основаны два метода: поиск консервативных участков во множественном выравнивании белковых семейств [3, 4] и метод паттернов и весовых матриц (PROSITE) [5]. Однако эти методы эффективно применяются только для сайтов, включающих относительно протяженные консервативные участки.

Другой подход использует данные по третичной структуре белков и включает следующие методы: молекулярный докинг [6], распознавание каталитических центров, на основе различий между моделированными кривыми титрования (THEMATICS) [7].

В последнее время активно развиваются методы, построенные на использовании пространственных паттернов функциональных сайтов, содержащих положение в пространстве функционально важных групп атомов. Наиболее известны методы поиска структурных паттернов каталитических сайтов ферментов в третичных структурах белков [8], распознавания сайтов связывания на основе поиска структурных паттернов [9].

Авторами проекта был разработан эффективный алгоритм, заключающийся в структурном сравнении белка не с консенсусом функционального сайта, а с каждым его представителем, что дало возможность получать информацию о структурном сходстве белковых фрагментов с сайтами, представленными малым количеством. [10].

Суть алгоритма состоит в следующем. Рассматриваются всевозможные тройки аминокислотных остатков в различных участках пространственной структуры белка в качестве ядра для искомого сайта. Каждая из троек сопоставляется в пространстве с первыми тремя остатками сайта из базы PDBSite. В случае удовлетворительного структурного сходства соответствующие множества сопоставляемых остатков расширяются до тех пор, пока их число не окажется равным числу остатков в сайте. Результатом являются все найденные множества аминокислотных остатков белка, удовлетворяющие критериям сходства сайт-белок.

Таким образом, программа распознавания функциональных сайтов позволяет аннотировать функцию вновь открытых белков, изучать изменение функции белка при изменении его структуры, моделировать искусственную структуру белка с заранее заданной биологической функции.

2.1.3.1 Работа с программной компонентой PDBSiteScan

Для подготовки к работе с данной программной компонентой следует войти в подсистему «Компьютерная протеомика» и с помощью мыши кликнуть надпись «Распознавание функциональных сайтов в пространственных структурах белков» в списке операций этой подсистемы. В результате появится HTML-страница с интерфейсом этой операции, который показан на Рис.7.



Рисунок 7. HTML-интерфейс операции, подготовленный к началу ее выполнения. Назначения полей: 1) окно для ввода пути к PDB файлу, 2) задание идентификатора цепи, 3) определение порога MDM (Maximum Distance Mismatch) для распознавания функциональных сайтов, 4) определение порога для максимально допустимого различия между показателем доступности для растворителя остатков шаблона функционального сайта и фрагмента белка с которым произведено структурное выравнивание шаблона, 5) задание типов функциональных сайтов, которые будут распознаваться, с помощью кнопок-флажков.

Выполнение операции начинается с ввода пути к файлу PDB исследуемого белка. Для этой цели используется кнопка “Browse”, которая предоставляет стандартный диалог открытия файла, либо пользователь вводит путь в окно “Enter file in PDB format” HTML-интерфейса операции, как это показано на Рис.15. Пользователь определяет идентификатор цепи в белковой структуре файла PDB, пороги MDM и accessibility. Далее пользователь определяет типы функциональных сайтов, которые будут включены в распознавание. Для этой цели необходимо поставить «галочку» в нужном check-box, как это показано на рисунке 8 ниже.

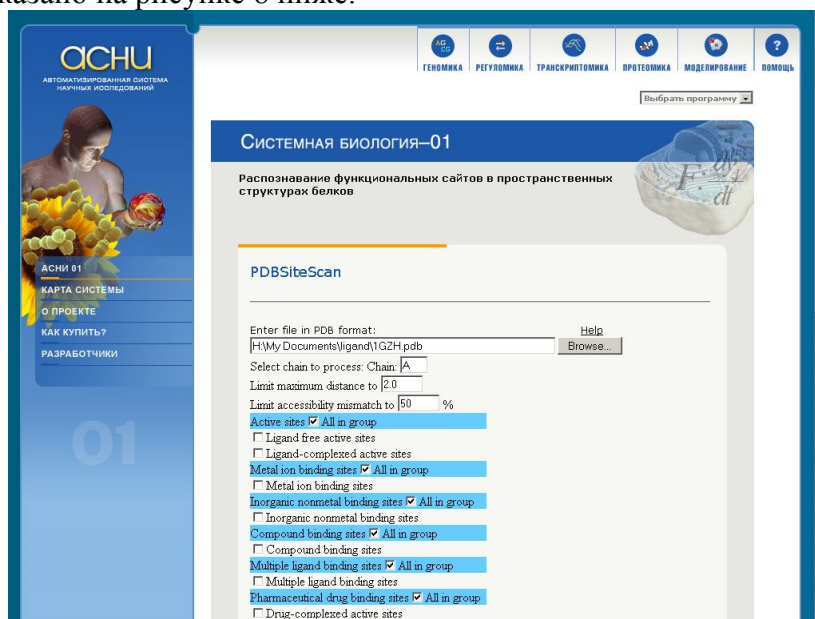


Рисунок 8. Определение параметров для распознавания функциональных сайтов.

После ввода данных как было описано выше необходимо нажать клавишу “Scan”, находящуюся внизу формы, в результате чего появляется HTML-страница с результатом распознавания функциональных сайтов в анализируемом белке, как это показано ниже на рис.9.

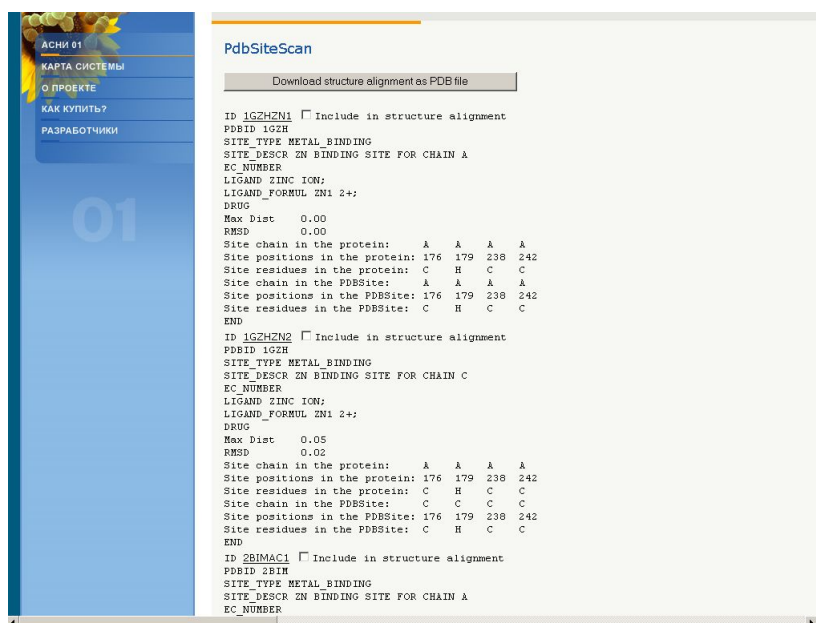


Рисунок 9 Пример HTML-страницы, содержащей результат предсказания функциональных сайтов в пространственной структуре pdb1gzh.ent.

Поля и данные, приведенные на форме имеют следующее значение: 1) ID идентификатор шаблона сайта в базе PDBSite, идентификатор является гиперссылкой на запись в этой базе; кнопка-флажок “Include in structure alignment” – отметить сайт для включения в структурное выравнивание белок-шаблон сайта, 2) PDBID – идентификатор шаблона сайта, 3) SITE_TYPE – тип сайта, 4) SITE_DESCR – описание сайта, 5) EC_NUMBER номер-классификатор фермента, если белок, из которого экстрагирован сайт-шаблон, таковым является, 6) LIGAND – название лиганда, 7) LIGAND_FORMUL – химическая формула лиганда, 8) DRUG – название лекарства, если лиганд является лекарством, 9) Max Dist и RMSD – параметры MDM и RMSD для структурного выравнивания шаблона сайта с белком, 10) Site chain in the protein - идентификатор цепи для распознанного сайта в белке, 11) Site positions in the protein - номера позиций для распознанного сайта в белке, 12) Site residues in the protein - обозначения аминокислот для распознанного сайта в белке, 13) Site chain in the PDBSite - идентификатор цепи для шаблона сайта, 11) Site positions in the PDBSite - номера позиций для шаблона сайта, 14) Site residues in the PDBSite - обозначения аминокислот для шаблона сайта.

Результат, представленный на HTML-страницы может быть сохранен в файл с помощью стандартных средств Интернет-браузера. Для получения PDB файла со структурным выравниванием белка с шаблонами сайтов необходимо кликнуть мышью на check-box интересующего сайта, а затем нажать на кнопку «Download structure alignment as PDB file». После этого действия будет показано окно для сохранения PDB файла на компьютере пользователя (см. рис. 10 ниже).



Рисунок 10 Пример HTML-страницы для сохранения PDB файла, появляющегося в ответ на нажатие кнопки «Download structure alignment as PDB file».

Для завершения работы с текущей (введенной) третичной структурой белка и перехода к обработке другого белка надо воспользоваться стандартными средствами Интернет браузера для возвращения на предыдущую страницу, в результате чего осуществляется подготовка к повторному выполнению этой операции.

2.1.4 Программные компоненты Subloc и SubPrediction : предсказание внутриклеточной локализации белков прокариот и эукариот.

Пространство внутри живой клетки не однородно, а состоит из замкнутых объемов, разделенных мембранами – компартментов. Прокариотические клетки часто содержат всего один компартмент, содержащий ДНК, РНК, белки. Фотосинтезирующие виды имеют также плоские цистерны - тилакоиды. Эукариотические клетки содержат большой набор компартментов, представленных, в основном, органеллами. Белковые молекулы клетки могут находиться в разных компартментах, а также быть связанными с различными структурами клетки – плазматическими мембранами, цитоскелетом и хроматином, в случае эукариот, или нуклеоидом, в случае эукариот. Каждый компартмент или клеточная структура имеет свой набор функций, который отчасти зависит от белков, находящихся в соответствующей области.

Среди известных серверов, реализующих методы предсказания внутриклеточной локализации белков для бактериальных организмов можно выделить следующие: SubLoc (<http://www.bioinfo.tsinghua.edu.cn/SubLoc/>), LOCTREE (<http://cubic.bioc.columbia.edu/cgi-bin/var/nair/loctree/query>), PSORTB (<http://www.psort.org/psortb/>), Proteome Analyst's Subcellular Localization Server (<http://www.cs.ualberta.ca/%7Ebioinfo/PA/Sub/index.html>), CELLO (<http://cello.life.nctu.edu.tw/>).

Методы предсказания основываются на поиске гомологов среди белков с установленной локализацией, поиске сигналов в последовательностях, характеризующих локализацию, использовании дискриминантного анализа и нейронных сетей, а также существуют алгоритмы, построенные на комбинации нескольких методов.

Алгоритм предсказания внутриклеточной локализации белков состоит из трех разделов: (1) поиск гомологов для анализируемой последовательности среди последовательностей белков с установленной локализацией методом BLAST и PsiBLAST, (2) классификация последовательности методом кластерного анализа КРАБ, и (3) классификация последовательности методом нейронных сетей. В результате классификации будет решаться задача отнесения последовательности к группе белков с заданной внутриклеточной локализацией. В качестве результата будут выдаваться предсказания каждого из трех алгоритмов, а также решение, построенное методом голосования.

Таким образом, определение внутриклеточной локализации вновь открытого белка может указать на область его биологических функций.

Разработаны две программные компоненты для предсказания внутриклеточной локализации белков – Subloc, для прокариот, и SubPrediction, для эукариот.

2.1.4.1 Работа с программной компонентой Subloc

Для подготовки к выполнению операции, соответствующей программной компоненте предсказания внутриклеточной локализации белков прокариот, следует войти в подсистему «Компьютерная протеомика» и с помощью мыши кликнуть надпись «Предсказание внутриклеточной локализации белков у прокариот» в списке операций этой подсистемы. В результате появится HTML-страница с интерфейсом этой операции, который показан на рис.11-А.

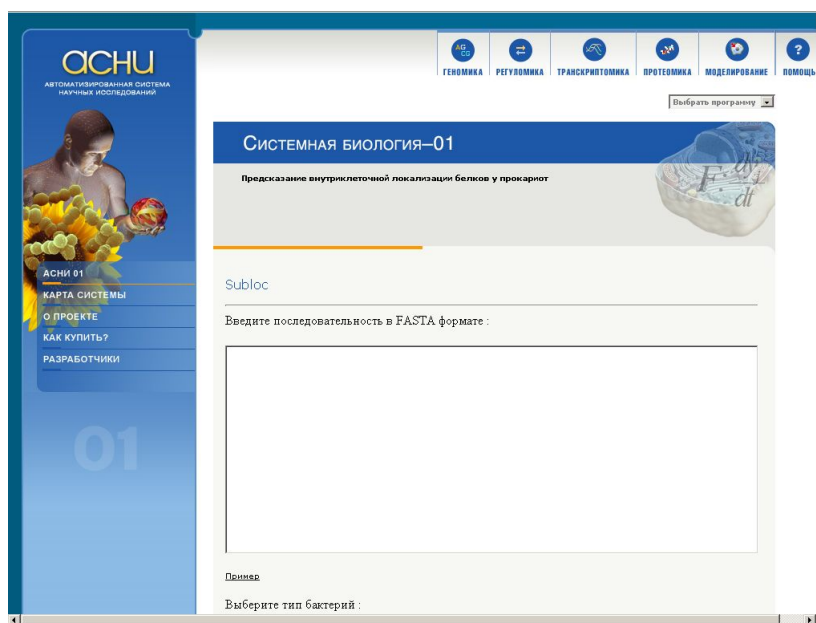


Рисунок 11-А HTML-интерфейс операции, подготовленный к началу ее выполнения.

Выполнение операции начинается с ввода аминокислотной исследуемого белка. Для этой цели используется раздел “Введите последовательность в FASTA формате” HTML-интерфейса операции, как это показано на рис.11-Б. Последовательность может быть набрана в поле экрана для ввода последовательности или скопирована с использованием стандартных операций текстового редактора интернет-браузера.

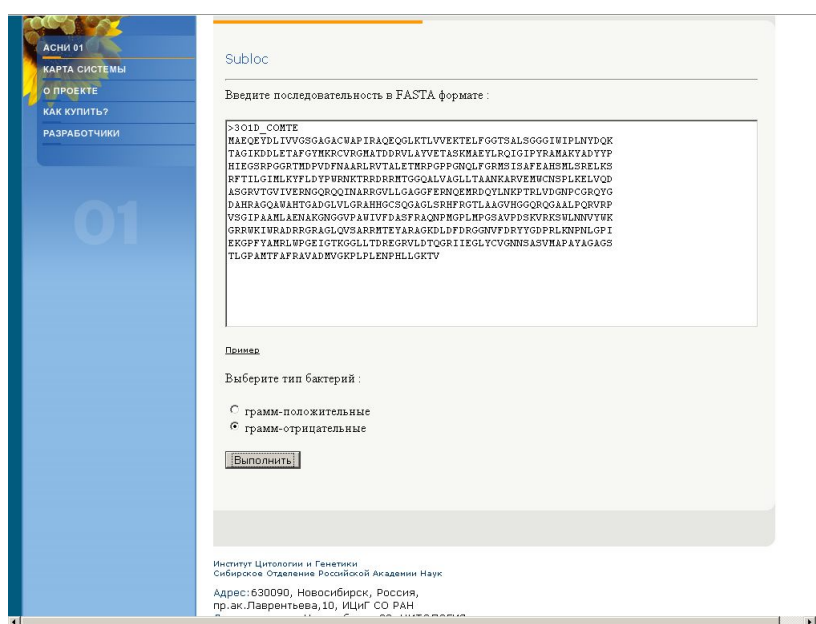


Рисунок 11-Б Ввод аминокислотной последовательности белка.

После ввода данных как было описано выше необходимо указать форму бактерии из которой был выделен анализируемый белок, это грамотрицательные или грамположительные бактерии. Для выбора формы бактерии нужно кликнуть мышкой на радио-кнопке «грамотрицательных», для грамотрицательных, или на «грамположительные», для грамположительных, а затем нажать клавишу “Выполнить”, в результате чего появляется HTML-страница с результатом предсказания внутриклеточной локализации для введенного белка, как это показано ниже на рис.12.

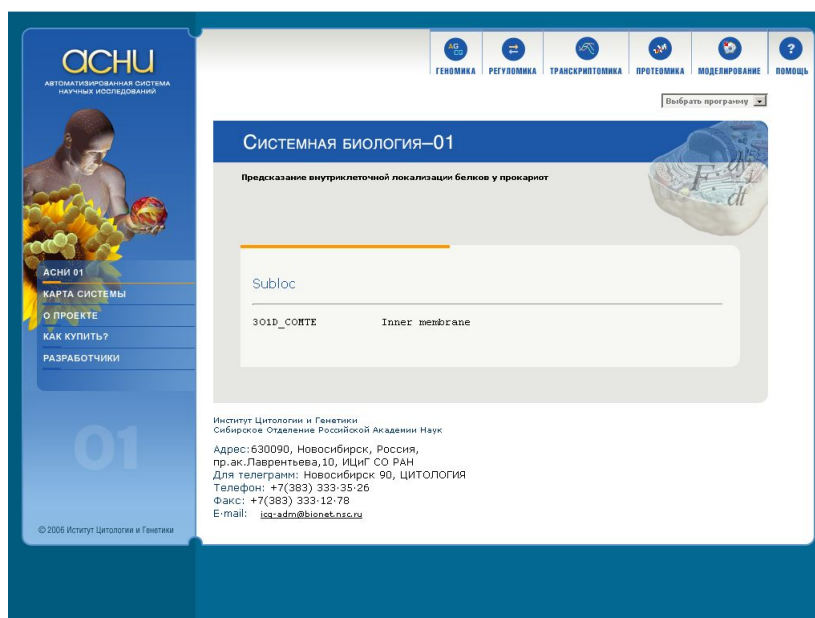


Рисунок 12 Пример HTML-страницы, содержащей результат предсказания внутриклеточной локализации для введенного белка грамотрицательной бактерии.

Для завершения работы с текущей (введенной) последовательности белка и перехода к обработке другого белка надо воспользоваться стандартными средствами Интернет браузера для возвращения на предыдущую страницу, в результате чего осуществляется подготовка к повторному выполнению этой операции.

2.1.4.2 Работа с программной компонентой SubPrediction

Для подготовки к предсказанию внутриклеточной локализации эукариотических белков следует войти в подсистему «Компьютерная протеомика» и с помощью мыши кликнуть надпись «Предсказание внутриклеточной локализации белков у эукариот» в списке операций этой подсистемы. В результате появится HTML-страница с интерфейсом этой операции, который показан на рисунке 13 ниже.

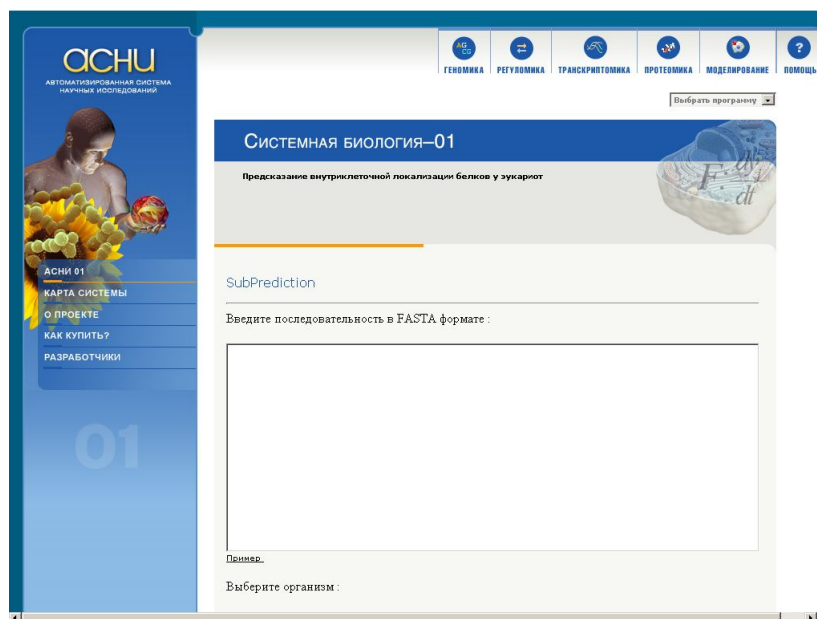


Рисунок 13. HTML-интерфейс операции, подготовленный к началу ее выполнения.

Выполнение операции начинается с ввода аминокислотной исследуемого белка. Для этой цели используется раздел “ Введите последовательность в FASTA формате” HTML-интерфейса операции, как это показано на рис.14. Последовательность может быть набрана в поле экрана для ввода последовательности или скопирована с использованием стандартных операций текстового редактора Интернет-браузера.

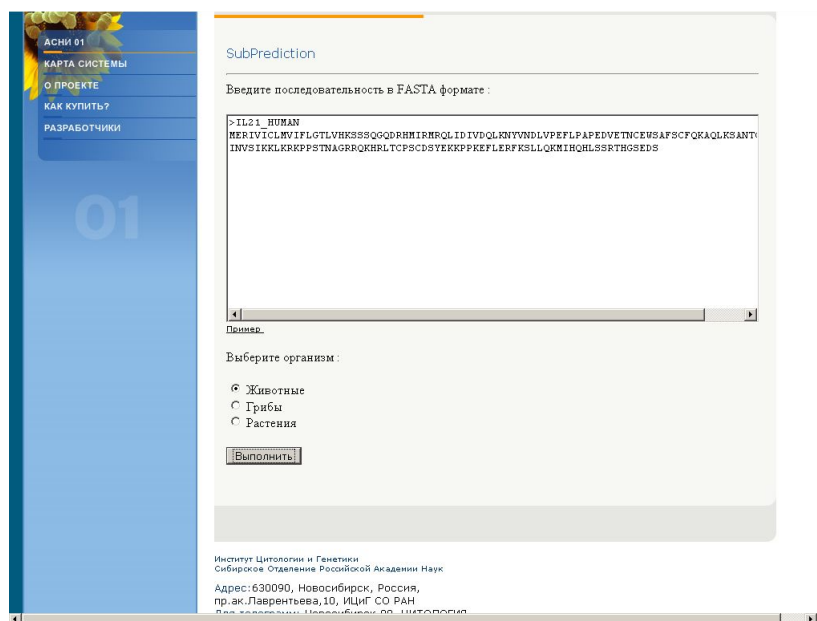


Рисунок 14. Ввод аминокислотной последовательности белка.

После ввода данных как было описано, выше необходимо указать форму организма, из которого был выделен анализируемый белок, это животные, грибы или растения. Для выбора организма нужно кликнуть мышкой на радио-кнопке «Животные», для животных, «Грибы», для грибов или «Растения», для растений, а затем кликнуть кнопку “Выполнить”, в результате чего появляется HTML-страница с результатом предсказания внутриклеточной локализации для введенного белка, как это показано ниже на Рис. 15.

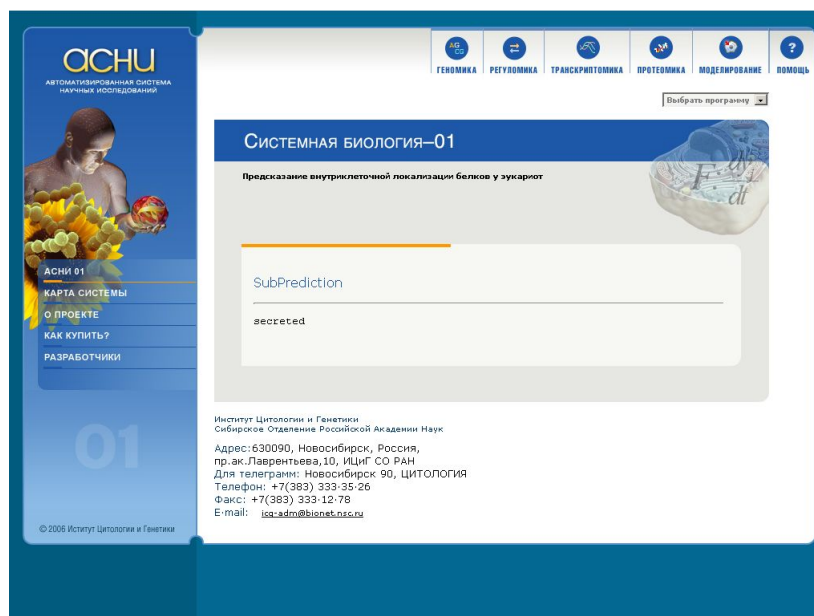


Рисунок 15. Пример HTML-страницы, содержащей результат предсказания внутриклеточной локализации для белка человека. Предсказано, что белок выделяется (секретируется) за пределы клетки

Для завершения работы с текущей (введенной) последовательностью белка и перехода к обработке другого белка надо воспользоваться стандартными средствами браузера Интернет для возвращения на предыдущую страницу, в результате чего осуществляется подготовка к повторному выполнению этой операции.

2.1.5 Программная компонента PDBSiteDocking: реконструкция пространственных структур комплексов "белок-ион металла", "белок-низкомолекулярные лиганды", "белок-белок", "белок-ДНК/РНК"

Многие белки выполняют свои биологические функции, будучи нековалентно связанными с различными молекулами: ионами металлов, низкомолекулярными соединениями, а также молекулами ДНК, РНК или других белков, то есть образуют комплексы, которые и обладают биологической активностью. Конкретный механизм работы такого комплекса зависит не только от его качественного состава, но и от его пространственной структуры, то есть от взаимного расположения элементов комплекса, структуры участков контакта его элементов и влияния контактирующих элементов на структуру друг друга. Реконструкция пространственных структур комплексов белков с другими молекулами позволяет определять молекулярные механизмы сложных биологических процессов, протекающих с участием белков [11].

Алгоритм программы построен на взаимодействии программного модуля распознавания функциональных сайтов в пространственной структуре белков (см. описание выше), базы данных пространственных структур функциональных сайтов (см. описание выше, база PDBSite) и библиотеки лигандов. Библиотека лигандов содержит координаты атомов для лигандов, которые по данным рентгеноструктурного анализа находятся в комплексе с сайтами из пространственных структур функциональных сайтов. Сайты связывания лигандов в базе PDBSite определялись из молекулярных комплексов белок-лиганд, которые приведены в базе PDB, поэтому в базе лигандов содержатся координаты атомов именно для комплексов сайт-лиганд. На первом этапе производится структурное выравнивание сайта связывания некоего лиганда из базы PDBSite с пространственным участком анализируемого белка с помощью программного модуля

распознавания функциональных сайтов в пространственной структуре белков. Затем преобразования поворота и сдвига (см. описание программного модуля распознавания функциональных сайтов в пространственной структуре белков), которые были произведены для совмещения сайта и фрагмента белка применяются к лиганду, с которым этот сайт образует комплекс. Таким образом, лиганд оказывается позиционирован с белком [1].

Программа реконструкции может применяться в области протеомики, молекулярной биологии, молекулярной генетики для исследования регуляции биологических процессов на различных уровнях: от экспрессии генов, до переключения активности белковых продуктов посредством связывания низкомолекулярных лигандов.

2.1.5.1 Работа с программной компонентой PDBSiteDocking

Для подготовки к работе с программной компонентой следует войти в подсистему «Компьютерная протеомика» и с помощью мыши кликнуть надпись «Реконструкция пространственных структур комплексов "белок-ион металла"; "белок-низкомолекулярные лиганды"; "белок-белок"; "белок-ДНК/РНК"» в списке операций этой подсистемы. В результате появится HTML-страница с интерфейсом этой операции, который показан на рисунке 16 ниже.

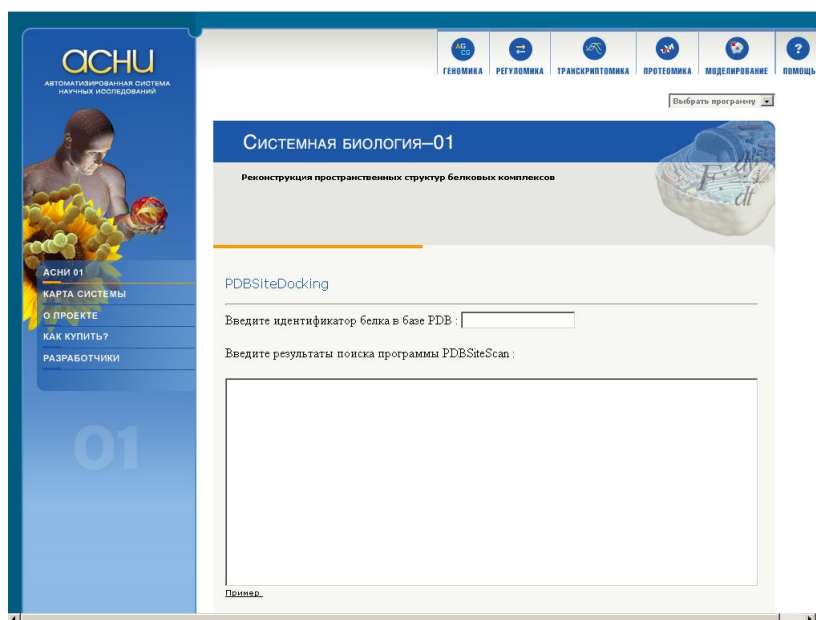


Рисунок 16. HTML-интерфейс операции, подготовленный к началу ее выполнения.

Выполнение операции начинается с ввода в окно формы «Введите идентификатор белка в базе PDB» PDB идентификатор анализируемого белка и ввода в окно «Введите результаты поиска программы PDBSiteScan» результата распознавания функционального сайта, выполненного программой PDBSiteScan, как показано на Рис.17. Заполнение данными окна «Введите результаты поиска программы PDBSiteScan» осуществляется с помощью стандартных операций интернет браузера копирования из буфера обмена.

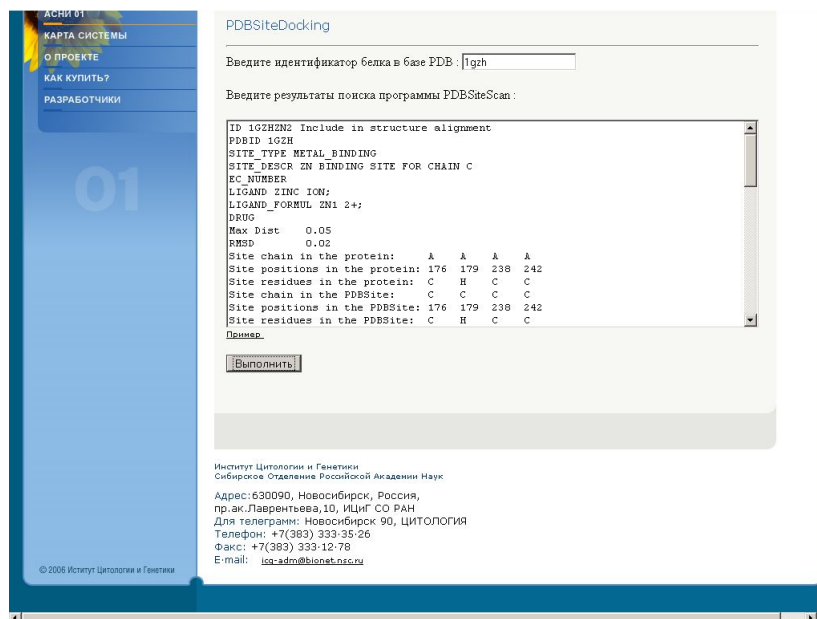


Рисунок 17. Определение входных данных для реконструкции молекулярных комплексов белок-лиганд.

После ввода данных как было описано выше необходимо нажать клавишу “Выполнить”, находящуюся внизу формы, в результате чего появляется HTML-страница с координатами атомов реконструированного комплекса белок-лиганд в PDB формате, как это показано ниже на рис.18.

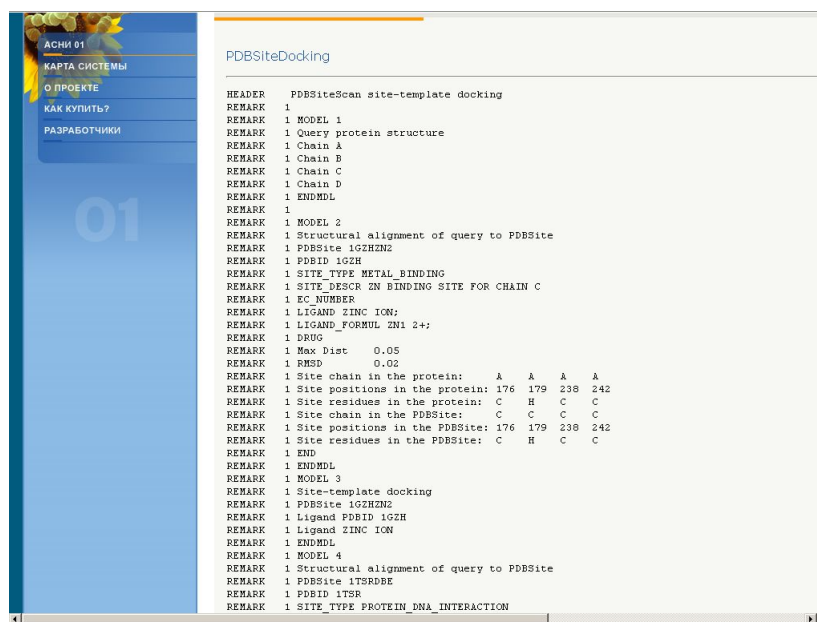


Рисунок 18. Пример HTML-страницы, содержащей результат реконструкции пространственной структуры молекулярного комплекса белок-лиганд.

Результат, представленный на HTML-странице может быть сохранен в файл с помощью стандартных средств Интернет-браузера.

Для завершения работы с текущей (введенной) третичной структурой белка и перехода к обработке другого белка следует воспользоваться стандартными средствами Интернет браузера для возвращения на предыдущую страницу, в результате чего осуществляется подготовка к повторному выполнению этой операции.

2.1.6 Программная компонента Veritope: предсказание В-клеточных эпитопов в аминокислотной последовательности

В-лимфоциты, или, проще, В-клетки несут на своей поверхности антитела, конститутивные (Fc) домены которых прикреплены к плазматической мембране, а вариабельные (Fab) выставлены наружу и могут связывать фрагменты антигенов, в том числе и белковой природы – эпитопы. Эпитопы имеют физико-химические особенности – как правило, они выставлены на поверхности и гидрофильны. Наиболее антигенные эпитопы расположены на выдающихся частях молекулы и не имеют жесткой третичной структуры. Последовательность аминокислот, образующих эпитоп, может быть разрывна, но в третичной структуре соответствующие остатки собраны вместе.

Предсказание В-клеточных эпитопов осуществляется на основе анализа профилей физико-химических свойств аминокислот. Профили строятся по следующим алгоритмам:

1. Антигенность по Веллингу [1];
2. Гидрофобность по Хоппу и Вудсу [2];
3. Гидрофобность по Кайту и Дулитлу [3];
4. Конформационная гибкость по Карплюсу и Шульцу [4].

Производится усреднение рассчитанных профилей. На результирующем профиле выделяются пики. Фрагменты последовательности в окрестности пиков, длиной 20 остатков с центром, соответствующим положению пика, принимаются за антигенные детерминанты [1].

Предсказание В-клеточных эпитопов в аминокислотной последовательности может применяться в области иммунологии для предсказания антигенной активности белка, структура которого неизвестна, или для моделирования искусственных белков, не обладающих антигенной активностью. Также возможно применение этого метода в области молекулярной генетики для предсказания или объяснения эффекта мутаций на антигенную активность белка.

2.1.6.1 Работа с программной компонентой Veritope

Для подготовки к работе с программной компонентой следует войти в подсистему «Компьютерная протеомика» и с помощью мыши кликнуть надпись «предсказание В-клеточных эпитопов» в списке операций этой подсистемы. В результате появится HTML-страница с интерфейсом этой операции, который показан на рисунке 19 ниже.

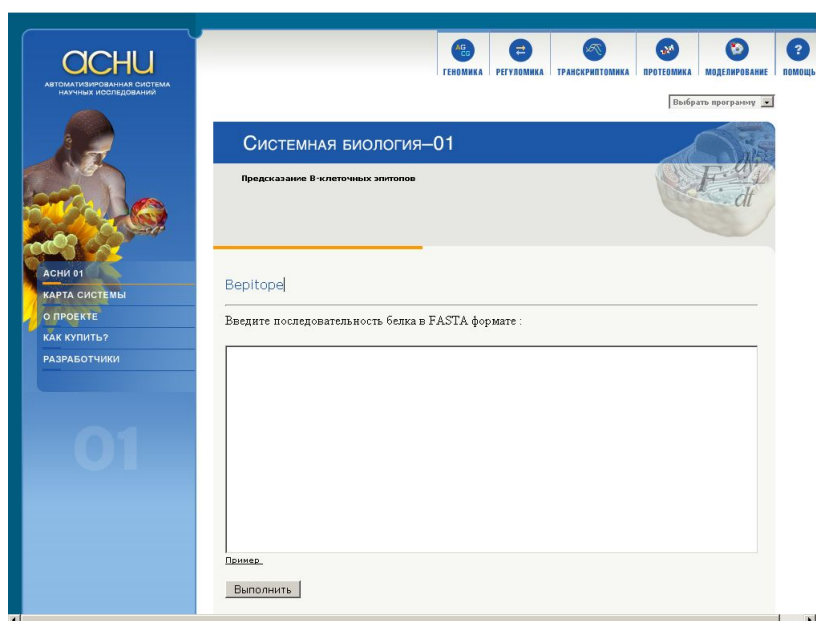


Рисунок 19. HTML-интерфейс операции, подготовленный к началу ее выполнения.

Выполнение операции начинается с ввода аминокислотной исследуемого белка. Для этой цели используется раздел “Введите последовательность белка в FASTA формате” HTML-интерфейса операции, как это показано на рис.20. Последовательность может быть набрана в поле экрана для ввода последовательности или скопирована с использованием стандартных операций текстового редактора Интернет-браузера.

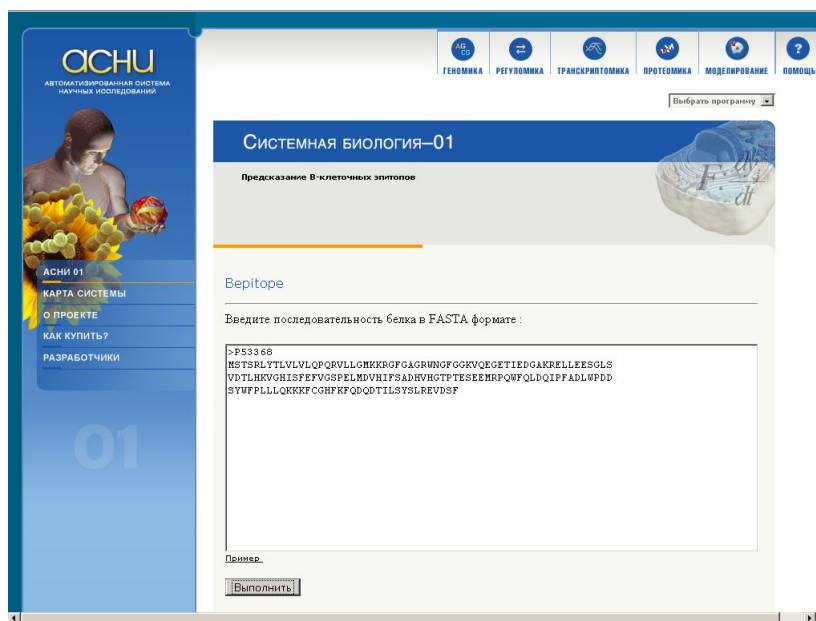


Рисунок 20 Ввод аминокислотной последовательности белка.

После ввода данных как было описано выше необходимо нажать клавишу “Выполнить”, в результате чего появляется HTML-страница с результатом предсказания сайтов протеосомного протеолиза для введенной последовательности, как это показано ниже на рис.21.

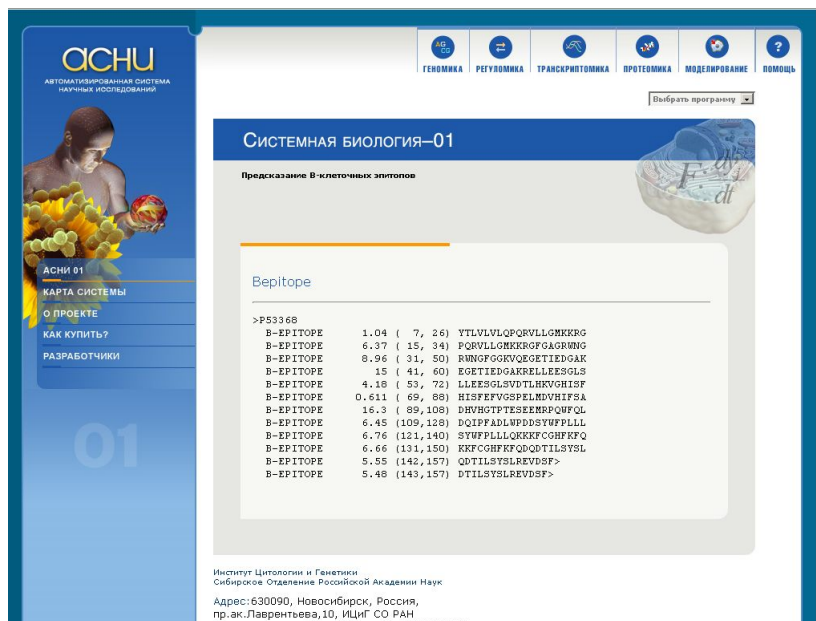


Рисунок 21. Пример HTML-страницы, содержащей результат предсказания B-эпитопов для введенной последовательности белка.

Для завершения работы с текущей (введенной) последовательности белка и перехода к обработке другого белка надо воспользоваться стандартными средствами

Интернет браузера для возвращения на предыдущую страницу, в результате чего осуществляется подготовка к повторному выполнению этой операции.

2.1.7 Программная компонента Cleavage: предсказание сайтов протеасомного протеолиза в аминокислотной последовательности

Действие подавляющего большинства белков в клетке ограничено во времени. По истечению определенного, часто довольно короткого периода белок подвергается протеасомному протеолизу – разрушению с помощью специального комплекса ферментов, представленного особой клеточной органеллой – протеасомой. Таким образом осуществляется регуляция количества и функции многих белков. Протеасомная деградация необходима для нормального проведения в клетке таких важных процессов, как регуляция клеточного цикла, транскрипция, участие в иммунном ответе, тепловой шок, апоптоз. Нарушение процесса протеолиза в протеасоме связано с различными заболеваниями, в частности злокачественным перерождением клетки и болезнью Паркинсона. Некоторые мутантные формы белков избегают протеасомной деградации, в результате чего происходит накопление белка в количествах, несовместимых с жизнью клетки. Избегать протеолиза способны и белки некоторых вирусов, что предотвращает уничтожение зараженных клеток иммунной системой хозяина

Скорость деградации белка зависит от N-концевых аминокислот, что было впервые установлено А. Варшавским [16]. Алгоритм основан на весовой матрице, построенной с использованием данных из работы [17]. Производится усреднение рассчитанных профилей. На результирующем профиле выделяются пики. Фрагменты последовательности в окрестности пиков, длиной 20 остатков с центром, соответствующим положению пика, принимаются за сайты протеолиза.

Определение сайтов протеасомного протеолиза позволяет оценить способность белка к регулируемому протеолизу и скорость этого процесса. Предсказание таких сайтов может также применяться при исследовании влияния мутаций на структуру и функцию белка, а также для более глубокого изучения механизмов протеолиза. Также этот метод может применяться в области вирусологии для изучения путей приспособления вирусов к иммунной системе хозяина и разработки возможных методов их предотвращения. Метод также применим в области медицинской генетики для исследования наследственных заболеваний, связанных с нарушением протеасомной деградации.

2.1.7.1 Работа с программной компонентой Cleavage

Для подготовки к работе с программной компонентой следует войти в подсистему «Компьютерная протеомика» и с помощью мыши кликнуть надпись «Предсказание сайтов конституитивного и иммунопротеасомного протеолиза» в списке операций этой подсистемы. В результате появится HTML-страница с интерфейсом этой операции, который показан на рисунке ниже.

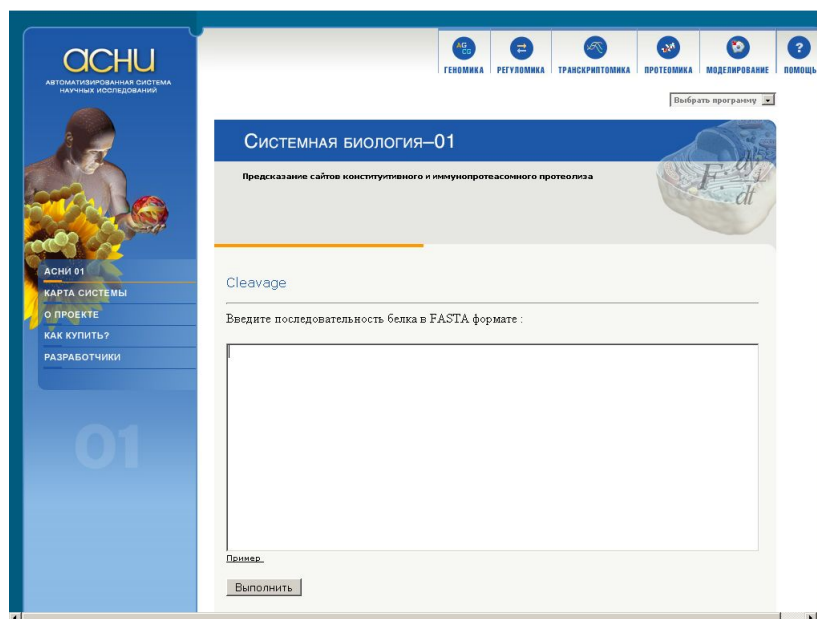


Рисунок 22. HTML-интерфейс операции, подготовленный к началу ее выполнения.

Выполнение операции начинается с ввода аминокислотной исследуемого белка. Для этой цели используется раздел “Введите последовательность белка в FASTA формате” HTML-интерфейса операции, как это показано на Рис.21. Последовательность может быть набрана в поле экрана для ввода последовательности или скопирована с использованием стандартных операций интернет браузера.

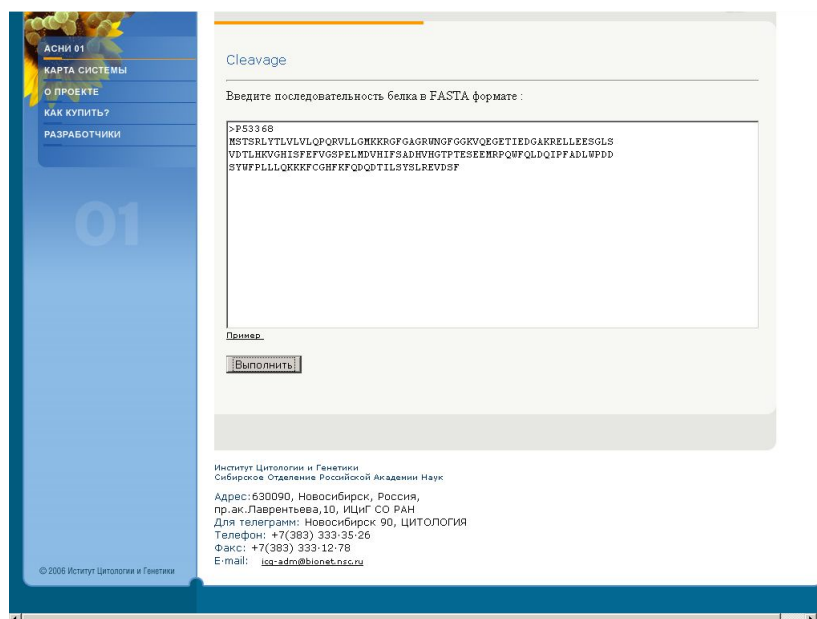


Рисунок 23. Ввод аминокислотной последовательности белка.

После ввода данных как было описано выше, необходимо нажать клавишу “Введите последовательность белка в FASTA формате”, в результате чего появляется HTML-страница с результатом предсказания сайтов протеасомального протеолиза для введенной последовательности, как это показано ниже на рис.24.

Position	Residue	Score
>P53368		
0	<	*
1	M	*
2	S	*
3	T	*
4	S	*
5	R	*
6	L	0.599
7	Y	-2.054
8	T	-8.843
9	L	-0.219
10	V	-4.029
11	L	-1.233
12	V	-4.26
13	L	-0.052
14	Q	-4.766
15	P	*
16	Q	-5.664
17	R	-2.795
18	V	-3.711
19	L	-0.774
20	L	-3.534
21	G	-2.892
22	M	-2.41
23	K	-7.942
24	K	-2.557
25	R	0.754
26	G	-1.317
27	F	1.111
28	G	0.249
29	A	-5.803
30	G	-5.618
31	R	-5.13
32	W	-0.076
33	N	-6.072
34	G	-3.906
35	F	0.818
36	G	-0.53
37	G	-1.651
38	K	-4.677
39	V	-1.301
40	Q	-2.103

Рисунок 24. Пример HTML-страницы, содержащей результат предсказания сайтов протеасомального протеолиза для введенной последовательности белка.

Для завершения работы с текущей (введенной) последовательностью белка и перехода к обработке другого белка надо воспользоваться стандартными средствами Интернет браузера для возвращения на предыдущую страницу, в результате чего осуществляется подготовка к повторному выполнению этой операции.

2.1.8 Программная компонента WebProAnalyst: предсказание мутаций, направленно меняющих биологические активности и свойства белков на основе анализа взаимосвязи между активностями белков и физико-химическими характеристиками аминокислот сайтов в выравнивании белковых последовательностей

При несинонимичной мутации в кодирующей части гена соответствующая аминокислота белкового продукта заменяется новой аминокислотой, имеющей свои, новые, физико-химические свойства. При изменении физико-химических свойств остатков, составляющих аминокислотную последовательность белка, изменяются физико-химические свойства всей белковой молекулы и ее третичная структура, что отражается на биологической активности белка. Как и физико-химические свойства, активность белка может быть измерена количественно.

Таким образом, имеется возможность установить количественное отношение между первичной структурой и активностью заданного белка, используя известные статистические модели. Так как биологическая активность белка во многом определяется его функциональными сайтами, то можно устанавливать характер зависимости активности белка от структуры функциональных сайтов и их набора в белке, определяемых мутациями заданной белковой формы. Также можно решать и обратную задачу – предсказывать, какие мутации могут привести к изменению структуры белка, определяющему заданное изменение его биологической активности.

Проблема взаимосвязи структуры и функции биологических молекул всегда была центральной в молекулярной биологии. Для белков эта взаимосвязь наиболее интересна и, по-видимому, наиболее сложна. Компьютерные методы анализа взаимосвязи структура-активность белков дают возможности для решения проблемы поиска сайтов в белковых последовательностях, аминокислоты которых определяют величину активности белков и позволяют строить количественные модели, связывающие значение активности с физико-

химическими свойствами данных аминокислот. Такие модели служат необходимым компьютерным инструментом в руках исследователей для рационального планирования белково-инженерных экспериментов.

Из анализа научной литературы можно выделить два основных подхода, используемых для решения данной проблемы, каждый из которых имеет свои положительные стороны и недостатки.

Изменения активностей белков в результате мутаций, как правило, сопряжено с конформационными изменениями белка, которые могут быть выявлены методами молекулярной динамики (Prevost et al., 1991).

Следующий подход к анализу взаимосвязи структура-активность базируется на установлении статистических зависимостей между активностью белков и характеристиками сайтов, определяемых через физико-химические, структурные или эволюционные свойства аминокислот.

В ИЦИГ СОРАН активно развиваются подходы для этой актуальной области молекулярной биологии. Разработана программа WebProAnalyst, которая автоматически генерирует и проверяет гипотезы по связи между физико-химическими свойствами аминокислот участков в выровненных последовательностях белков и их активностями (Ivanisenko et al, 2005).

Алгоритм компьютерного анализа количественной взаимосвязи структура-активность в семействах гомологичных и/или мутантных белков базируется на автоматической генерации и проверки гипотез по связи между физико-химическими свойствами аминокислотных участков в выровненных последовательностях белков и белковых активностей. Программа сканирует множественное выравнивание белков подвижной рамкой. Рамка задает сайты последовательностей белков для анализа. Для каждого сайта рассчитываются физико-химические свойства, и строится регрессионная модель (множественный регрессионный анализ или нейронная сеть), связывающая активности белков с физико-химическими свойствами. Далее оценивается статистическая значимость модели. Построенная модель используется для предсказания активности мутантных или вновь секвенированных белков.

Данный метод позволяет планировать получение искусственные белки с заранее заданной функцией путем направленного мутагенеза и может применяться в молекулярной биологии. Также метод может быть применен в эволюционной биологии для исследования влияния предполагаемых мутаций на эволюцию белков.

2.1.8.1 Работа с программной компонентой WebProAnalyst

Для подготовки к работе с программной компонентой следует войти в подсистему «Компьютерная протеомика» и с помощью мыши кликнуть надпись «Предсказание мутаций, направленно меняющих биологические активности и свойства белков» в списке операций этой подсистемы. В результате появится HTML-страница с интерфейсом этой операции, который показан на рис 25.

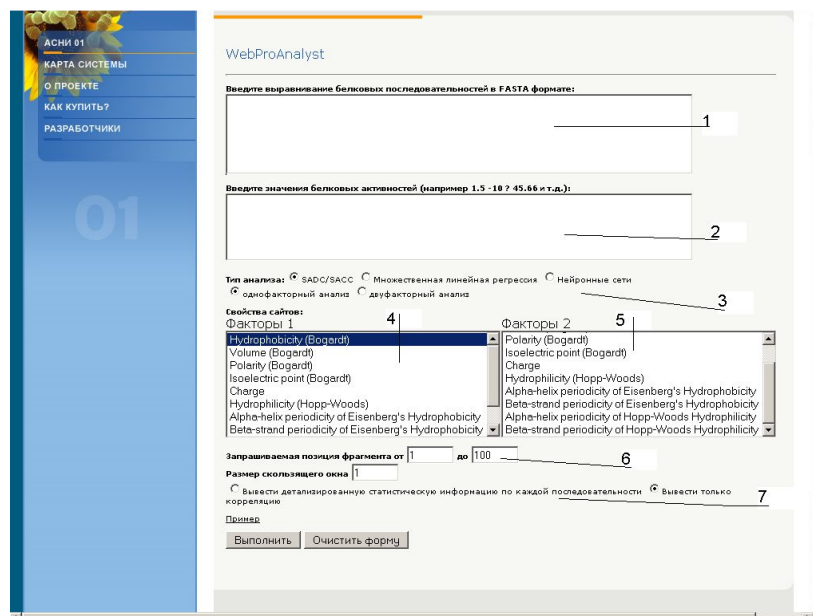


Рисунок 25. HTML-интерфейс операции, подготовленный к началу ее выполнения. Назначение полей на HTML форме: 1) Окно ввода множественного выравнивания, 2) Окно ввода величин активностей, 3) Тип статистической модели, 4, 5) Физико-химические свойства сайта, которые будут использоваться при построении модели структура-активность, 6) Фрагмент множественного выравнивания для анализа, 7) Тип выходных данных (только значения корреляций или полная статистическая модель).

Выполнение операции начинается с ввода множественного выравнивания аминокислотных последовательностей белков в Fasta формате, значения величины их активностей, выбор типа статистической модели для анализа, физико-химических свойств сайта, указания позиций фрагмента выравнивания для анализа, типа выходных данных. Для этой цели используются разделы 1 – 7 HTML-интерфейса операции, как это показано на Рис.26. Значение активности для мутантной последовательности, активность которой требуется предсказать, указывается знаком ‘?’.

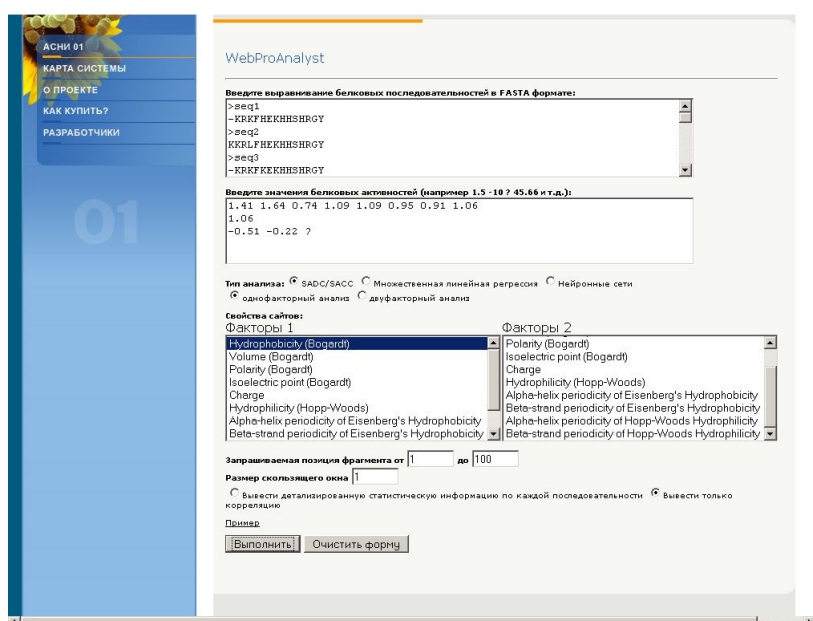


Рисунок 26. Ввод данных для анализа.

После ввода данных как было описано выше необходимо нажать клавишу “Выполнить”, в результате чего появляется HTML-страница с результатом анализа количественной взаимосвязи структура-активность для введенных последовательностей

белков с предсказанием активности для мутантной последовательности, как это показано ниже на рис.27.

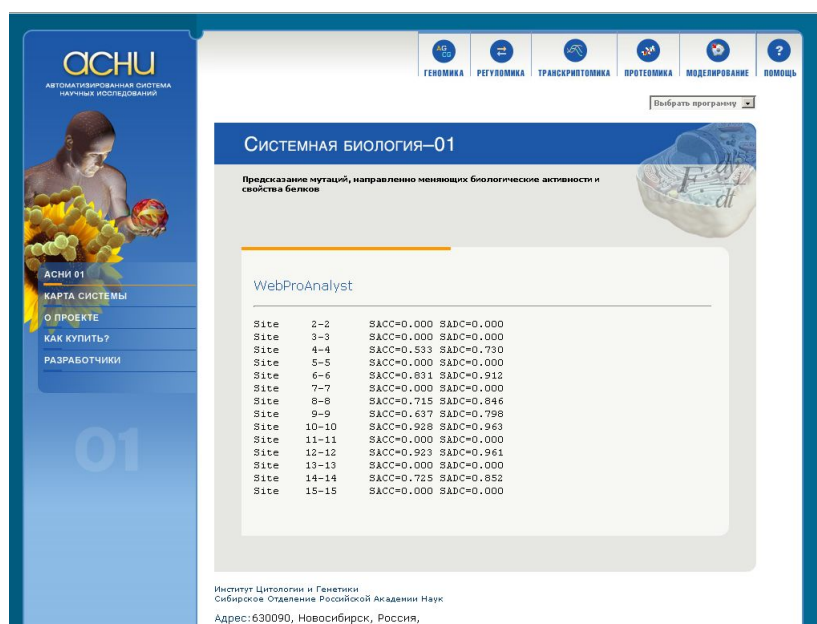


Рисунок 27. Пример HTML-страницы, содержащей результат анализа количественной взаимосвязи структура-активность для введенных последовательностей белков.

Для завершения работы с текущей (введенной) последовательности белка и перехода к обработке другого белка надо воспользоваться стандартными средствами Интернет браузера для возвращения на предыдущую страницу, в результате чего осуществляется подготовка к повторному выполнению этой операции.

3 Полезные ссылки.

1. V.A. Ivanisenko, S.S. Pintus, P.S. Demenkov, M.A. Krestyanova, E.K. Litvenko, D.A. Grigorovich, V.A. Debelov Fastprot: a Computational Workbench for Recognition of the Structural and Functional Determinants in Protein Tertiary Structures. //Bioinformatics of Genome Regulation and Structure II. (Eds. N.Kolchanov, R. Hofstaedt, L.Milanesi) Springer Science+Business Media, Inc. 2006. P. 305-316.
2. Bairoch A, Bucher P. PROSITE: recent developments. //Nucleic Acids Res. 1994. V. 22. N. 17. P. 3583-9.
3. Kalinina OV, Novichkov PS, Mironov AA, Gelfand MS, Rakhmaninova AB. // Nucleic Acids Res. 2004. V. 32. W424-428.
4. Aloy P., Querol E., Aviles F.X., Sternberg M.J. //J Mol Biol. 2001. 311:395-408.
5. Bairoch A, Bucher P. PROSITE: recent developments. //Nucleic Acids Res. 1994. V. 22. N. 17. P. 3583-9.
6. Schneidman-Duhovny D, Nussinov R, Wolfson HJ. Predicting molecular interactions in silico: II. Protein-protein and protein-drug docking. //Curr Med Chem. 2004 V. 11. N. 1. P. 91-107.
7. Ondrechen MJ, Clifton JG, Ringe D. THEMATICs: a simple computational predictor of enzyme function from structure. //Proc Natl Acad Sci U S A. 2001 V. 98. N. 22. P. 12473-8.

8. Barker JA, Thornton JM. An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis. // *Bioinformatics*. 2003. V. 19. N. 13. P.1644-9.
9. Hendlich M, Bergner A, Gunther J, Klebe G. Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. // *J Mol Biol*. 2003. V. 326. N. 2. P.607-20.
10. Ivanisenko VA, Pintus SS, Grigorovich DA, Kolchanov NA. PDBSiteScan: a program for searching for active, binding and posttranslational modification sites in the 3D structures of proteins. // *Nucleic Acids Res*. 2004. V. 32. P. W549-54.
11. Shoichet BK, Stroud RM, Santi DV, Kuntz ID, Perry KM. Structure-based discovery of inhibitors of thymidylate synthase. // *Science*. 1993. V. 259. N. 5100. P.1445-50.
12. Welling et al. Prediction of sequential antigenic regions in proteins. // *FEBS Letters*. 1985. V. 188. P. 215-218
13. Hopp TP, Woods KR (1981), *Proc. Natl Acad. Sci. USA* 78, 3824
14. Kyte J, Doolittle RF (1982), *J.Mol.Biol* 157, 105
15. Karplus PA, Schultz GE Prediction of chain flexibility in proteins. // *Naturwissenschaften*. 1985. V. 72. P. 212-213
16. Bachmair A., Finley D., Varshavsky A. // *Science*. 1986. V.234. P.179-186 .
17. R.E.M. Toes, et al. Discrete Cleavage Motifs of Constitutive and Immunoproteasomes Revealed by Quantitative Analysis of Cleavage Products. // *J. Exp. Med*. 2001. V. 194. N. 1. P. 1-12
18. Ivanisenko VA, Eroshkin AM, Kolchanov NA. WebProAnalyst: an interactive tool for analysis of quantitative structure-activity relationships in protein families. // *Nucleic Acids Res*. 2005 V. 33 P. W99-104.