На правах рукописи

Алемасов Николай Александрович

КОМПЬЮТЕРНЫЙ АНАЛИЗ СВЯЗИ МЕЖДУ КОНФОРМАЦИОННЫМИ СВОЙСТВАМИ МУТАНТНЫХ ФОРМ БЕЛКА SOD1 И БОКОВЫМ АМИОТРОФИЧЕСКИМ СКЛЕРОЗОМ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ

Специальность 03.01.09 —
«Математическая биология, биоинформатика»

Автореферат
диссертации на соискание учёной степени
кандидата биологических наук

Новосибирск — 2018

Работа выполнена в лаборатории компьютерной протеомики Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», г. Новосибирск.

Научный руководитель: канд. биол. наук, доцент
Иванисенко Владимир Александрович

Официальные оппоненты: **Веселовский Александр Владимирович,**
доктор биологических наук, заведующий лабораторией структурной биоинформатики отдела биоинформатики,
Федеральное государственное бюджетное научное учреждение «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича», г. Москва

Антонец Денис Викторович,
кандидат биологических наук, старший научный сотрудник теоретического отдела,
Федеральное бюджетное учреждение науки «Государственный научный центр вирусологии и биотехнологии «Вектор»» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека, р.п. Кольцово

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, г. Новосибирск

Защита диссертации состоится ___ _____ 201_ г. на утреннем заседании диссертационного совета Д 003.011.01 на базе ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», в конференц-зале Института по адресу:
пр. академика Лаврентьева, 10, г. Новосибирск, 630090,
+7 (383) 363-49-06; факс: +7 (383) 333-12-78;
e-mail: dissov@bionet.nsc.ru

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте Института www.bionet.nsc.ru.

Автореферат разослан ___ _____ 201_ года.

Ученый секретарь
диссертационного совета
доктор биологических наук Хлебодарова Т. М.

Общая характеристика работы

Актуальность темы.

Боковой амиотрофический склероз (БАС) – неизлечимое нейродегенеративное заболевание. Выделяется две его формы: наследственная и спорадическая. Известно, что второй по распространённости и наиболее изученной причиной наследственной формы БАС являются мутации в гене *SOD1*, кодирующем фермент супероксиддисмутазу-1. Одной из гипотез относительно механизма возникновения заболевания на молекулярном уровне является агрегация белка SOD1, вызванная его неверной укладкой в результате мутаций.

Большинство существующих работ, посвященных проблеме агрегации мутантных белков, основаны на экспериментальных данных по изменению их термодинамической стабильности. В гораздо меньшем количестве работ принимается во внимание детальный анализ отдельных факторов, включающих сети водородных связей, солевые мостики и другие физико-химические, структурные и конформационные характеристики белков. Известно, например, что водородные связи играют важную роль, в частности, в поддержании структуры белков. Так, в одном из *in silico* исследований структуры и динамики SOD1 уже анализировались водородные связи и сеть, которую они образуют в этом белке. Другое исследование с применением метода молекулярной динамики (МД) показало, что структура SOD1 становится более гибкой в мутантных белках по сравнению с белком дикого типа. В частности, было показано, что изменение водородных связей в мутантах может приводить к неверной укладке и агрегации данных мутантных форм SOD1.

В рамках настоящей работы сделано предположение, что как стабилизация, так и дестабилизация структуры мутантных SOD1 может влиять на возникновение БАС. Некоторые мутации могут оказывать пространственно-распределенный эффект на физико-химические и структурные характеристики, определяющие подверженность белка агрегации. При этом происходят локальные разнонаправленные изменения характеристик различных участков пространственной структуры белка, компенсирующие друг друга в масштабе всей структуры. Таким образом, такой интегральный показатель как термостабильность не позволяет выявить закономерности агрегации белков. В частности, «патогенные» мутации SOD1, оказывающие влияние на сеть водородных связей путем разрушения одних связей и возникновения других, могут повышать вероятность перехода из нативной конформации SOD1 в её «патогенную» форму. На стабильность белка могут также влиять другие типы связей, например, такие как водородные связи с молекулами воды и водные мостики.

С помощью метода молекулярной динамики в данной работе было исследовано 39 мутаций в белке SOD1. Были обнаружены водородные связи, доля времени существования которых в траектории МД (или стабильность) достоверно коррелирует с дожитием пациентов с БАС. На основе разницы стабильности

найденных водородных связей в мутантах и белке дикого типа были построены регрессионные модели, позволяющие предсказать дожитие пациентов с мутациями в SOD1. Для учёта крупномасштабной динамики белка было также проведено моделирование его структуры с помощью эластичных сетевых (ЭС) моделей. На основе результатов моделирования предложены комплексные регрессионные модели, опирающиеся на стабильность водородных связей, как усреднённую по времени моделирования (из МД), так и по ансамблю конформаций (из ЭС). Коэффициент корреляции между предсказанными и наблюдаемыми значениями дожития пациентов – носителей мутаций в SOD1 – для построенных моделей варьировался от $R = 0.67$ ($R^2 = 0.45$) до $R = 0.98$ ($R^2 = 0.96$). Исследованы области в структуре белка, содержащие водородные связи, стабильность которых коррелирует с патогенностью мутантов SOD1, а также области с повышенной эволюционной консервативностью и механической жёсткостью.

Целью работы является компьютерный анализ связи между конформационными свойствами мутантных белков SOD1 и наследственной формой бокового амиотрофического склероза на основе оценки стабильности водородных связей и водных мостиков.

Для достижения цели были поставлены следующие **задачи**:

1. Разработка метода расчёта изменения конформационных свойств мутантных форм белка по сравнению с белком дикого типа на основе оценки стабильности внутримолекулярных водородных связей, водородных связей белка с молекулами воды и водных мостиков с использованием метода молекулярной динамики и эластичных сетевых моделей.
2. Построение и исследование регрессионных моделей, связывающих конформационные характеристики мутантных форм белка SOD1, ассоциированных с БАС с дожитием пациентов, являющихся носителями данных мутаций. Предсказание дожития пациентов с мутациями в SOD1, данные по которым отсутствуют в литературе.
3. Анализ распределения ключевых водородных связей и водных мостиков, важных для предсказания дожития пациентов с БАС, в пространственной структуре белка SOD1, направленный на выявление их потенциальной роли в структурно-функциональной организации белка.

Научная новизна: Предложен новый метод предсказания влияния мутаций на конформационные свойства мутантных белков, основанный на оценке стабильности внутримолекулярных водородных связей, водородных связей белка с молекулами воды и водных мостиков, с использованием метода молекулярной динамики и эластичных сетевых моделей. Разработанные регрессионные модели позволяют предсказывать дожитие пациентов с БАС на основе изменения стабильности водородных связей в мутантных белках SOD1. Применение данного метода для изучения белков SOD1 позволило впервые получить следующие результаты:

1. Показано, что дожитие пациентов с мутациями в SOD1 имеет высокую корреляцию с конформационными свойствами мутантных форм белка,

полученными из моделирования молекулярной динамики и эластичных сетей.

2. Показано, что водородные связи, стабильность которых достоверно коррелирует с дожитием пациентов с соответствующими мутациями в SOD1, образуются аминокислотными остатками, участвующими в формировании дисульфидной связи, сайта связывания иона меди, электростатической, дисульфидной и цинк-связывающей петли.
3. Обнаружено, что бóльшая часть аминокислотных остатков белка SOD1 (Гис46, Гис71 и Асп124), которые участвуют в образовании водородных связей со значительным отклонением стабильности в мутантах SOD1 от таковой в белке дикого типа участвует в образовании важного контакта между сайтами связывания ионов металлов.
4. Показано, что существует достоверная положительная корреляция ($R = 0.57, p = 2 \times 10^{-14}$) между средней механической жёсткостью позиций в белке и эволюционной консервативностью остатков в данных позициях. На основе анализа механической жёсткости позиций белка сделано предположение о существовании внутримолекулярного взаимодействия между активными центрами обеих субъединиц. Показано, что механическая жёсткость между позициями с изучаемыми мутациями в SOD1 и позицией 46, остаток в которой является частью активного центра белка, достоверно отрицательно коррелирует ($R = -0.41, p = 0.02$) с дожитием пациентов с соответствующими мутациями.

Практическая значимость Разработанный метод может иметь высокое прикладное значение для прогнозирования дожития пациентов, несущих патогенные мутации в белке SOD1. Разработанные методы также позволяют оценивать степень влияния на белок ранее неизвестных полиморфизмов в ответственных за болезнь генах, что является крайне важным для учёта в диагностике индивидуального набора полиморфизмов конкретного пациента. Результаты данной работы могут быть использованы для оценки влияния не только мутаций, но и малых химических соединений при их взаимодействии с белком на конформационные свойства белка, что является важным при создании новых лекарственных препаратов, направленных на коррекцию патогенных конформационных изменений мутантов SOD1.

Методы исследования. Для решения поставленных задач в работе используется метод молекулярной динамики, эластичные сетевые модели, регрессионный анализ, кластерный анализ и математическая статистика, а также методы множественного выравнивания аминокислотных последовательностей и анализа их консервативности.

Положения, выносимые на защиту:

1. Стабильность водородных связей в белке SOD1, образованных аминокислотными остатками, участвующими в формировании дисульфидных связей, а также сайтов связывания ионов меди и цинка, достоверно

коррелирует с дожитием пациентов, страдающих боковым амиотрофическим склерозом.
2. Регрессионные модели, построенные на основе оценок стабильности внутри- и межмолекулярных водородных связей, полученных с использованием метода молекулярной динамики и эластичных сетевых моделей, позволяют предсказывать влияние мутаций в белке SOD1 на величину дожития пациентов с боковым амиотрофическим склерозом.
3. Механическая жёсткость между позициями мутаций в SOD1 и позицией 46 в активном центре фермента достоверно коррелирует с дожитием пациентов, носителей соответствующих мутаций.

Апробация работы. Результаты работы докладывались и обсуждались на следующих конференциях:

1. The fifth International German/Russian Workshop in «Integrative Biological Pathway Analysis and Simulation» (Germany, Bielefeld, June 2–3, 2014).
2. VII Российский симпозиум с международным участием «Белки и пептиды» (Новосибирск, 12–17 июля 2015 г).
3. Международная конференция «Актуальные проблемы вычислительной и прикладной математики 2015» (Новосибирск, 19–23 октября 2015 г).
4. Сателлитный симпозиум «Молекулярно-генетические механизмы патогенеза глаукомы» в рамках Международной конференции «Беляевские чтения», посвященной 100-летию академика АН СССР Д.К. Беляева (Новосибирск, 7–10 августа 2017 г).
5. Симпозиум «Системная биология репарации ДНК и программируемой клеточной гибели» в рамках 11 Международной мультиконференции по биоинформатике регуляции и структуры геномов и системной биологии (Bioinformatics of Genome Regulation and Structure\ Systems Biology — BGRS\SB-2018) (Новосибирск, 20–22 августа 2018 г).

Личный вклад. Основные результаты работы получены автором самостоятельно. Автором разработаны и реализованы компьютерные подходы, лежащие в основе исследования, проведено компьютерное моделирование, статистическая обработка расчётов, а также анализ литературных данных и обсуждение полученных результатов.

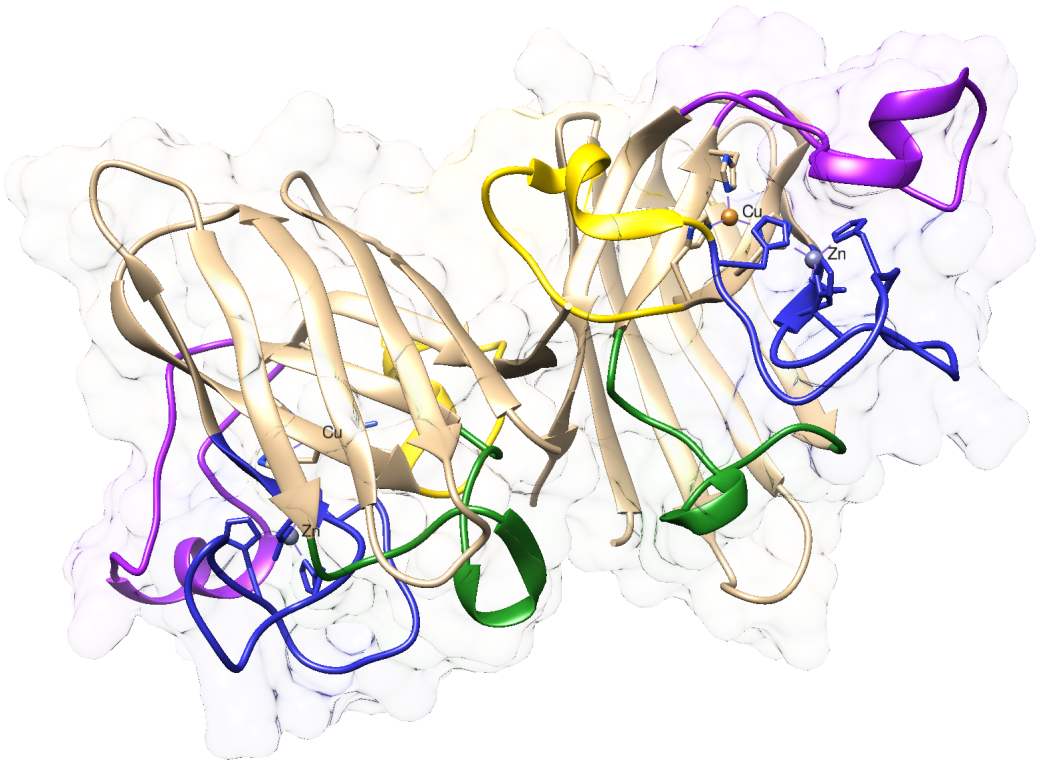
Публикации. Основные результаты по теме диссертации изложены в 9 печатных изданиях, 4 из которых изданы в журналах, рекомендованных ВАК, 5 — в тезисах докладов.

Объем и структура работы. Диссертация состоит из введения, пяти глав, заключения, выводов и двух приложений. Полный объём диссертации составляет 151 страницу, включая 44 рисунка и 6 таблиц. Список литературы содержит 228 наименований.

Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, приводится обзор научной литературы по изучаемой проблеме, формулируется цель, ставятся задачи работы, излагается научная новизна и практическая значимость представляемой работы. В последующих главах описывается метод предсказания влияния мутаций в SOD1 на патогенные свойства мутантов, который используется затем для предсказания дожития пациентов, страдающих БАС, с мутациями в этом белке, для которых такой информации из литературы не известно. Выдвигается гипотеза о механизме развития наследственной формы заболевания БАС у пациентов с мутациями в белке SOD1.

Первая глава посвящена обзору литературы. Описываются наследственные причины возникновения БАС, среди которых – накопление белковых агрегатов с участием белка SOD1, вызванное неверной его укладкой в результате мутаций.



Жёлтым цветом выделена дисульфидная петля, синим – цинк-связывающая, зелёным – «greek-key», фиолетовым – электростатическая петля. Ионы металлов выделены сиреневым (цинк) и оранжевым (медь).
Рис. 1 — Пространственная структура белка SOD1, состоящего из двух мономеров: A и F (PDBID: 2V0A).

Приводятся современные гипотезы об агрегации белков, сведения о структуре и функции SOD1 (см. рис. 1), а также характеристики структуры белка, которые обычно исследуются в рамках молекулярного моделирования. В этой

же главе описываются компьютерные методы молекулярного моделирования структуры и динамики белков и существующие регрессионные модели для предсказания патогенности мутаций в белке SOD1. Делается вывод, в частности, о том, что до сих пор не создано эффективной регрессионной модели для определения *in silico* патогенности мутантов белка SOD1, которая бы объясняла более 70 % дисперсии в дожитии пациентов с БАС. Кроме того, для компьютерного исследования агрегации белка SOD1, помимо термостабильности его мутантов, следует принимать во внимание и характеристики водородных связей, которые могут играть важную роль в поддержании вторичной и третичной структуры данного белка.

Вторая глава посвящена методам и программным средствам для компьютерного моделирования конформационных свойств белка SOD1, а также литературным данным, используемым для построения регрессионных моделей. Моделирование методом молекулярной динамики (МД) происходило с использованием пакета AMBER 12, силового поля ff99SB, модели воды TIP3P. Время моделирования задавалось равным 50 нс. Для моделирования эластичных сетей (ЭС) применялась система ElNemo.

Для расчёта показателя стабильности водородных связей в SOD1 дикого типа и его мутантах из моделирования МД использовалось количество конформаций, в которых водородная связь наблюдалась, отнесённое к длине фазовой траектории МД ($O_i^m = \frac{1}{n}\sum_{j=1}^{n} o_{ij}^m$). Стабильность водородных связей, оцениваемая на основе анализа ЭС, рассчитывалась как количество конформаций, в которых водородная связь присутствовала, делённое на общее число конформаций. В работе рассматривались водородные связи между атомами внутри белка, между атомами белка и атомами молекул воды и водные мостики. Группа водородных связей внутри белка, полученных при моделировании МД далее будет обозначаться PPhb, $\text{PPhb}_{\text{ЭС}}$ будет обозначать группу таких связей, полученных при моделировании ЭС, PWhb – водородные между атомами белка и атомами молекул воды, а Wbr – водные мостики.

Кластеризация мутантов SOD1 по стабильности водородных связей каждого типа проводилась с помощью метода mean shift. Анализ механической жёсткости между всеми парами позиций в первичной структуре белка SOD1 проводился с использованием системы ProDy. Механическая жёсткость κ_{ij} между парой остатков i и j имеет смысл константы жёсткости эквивалентной «пружины» между ними. Средняя жёсткость характеризует степень зависимости движений некоторого аминокислотного остатка от других остатков в белке или, другими словами, механическую сопротивляемость отдельных остатков деформациям, в целом. Для анализа эволюционной консервативности аминокислотных остатков SOD1 использовалась система ConSurf. В анализе использовалось 150 последовательностей гомологов SOD1 с уровнем идентичности от 35 до 95 %.

При разработке метода предсказания влияния мутаций на патогенные свойства мутантных белков исследовались наблюдаемые значения дожития

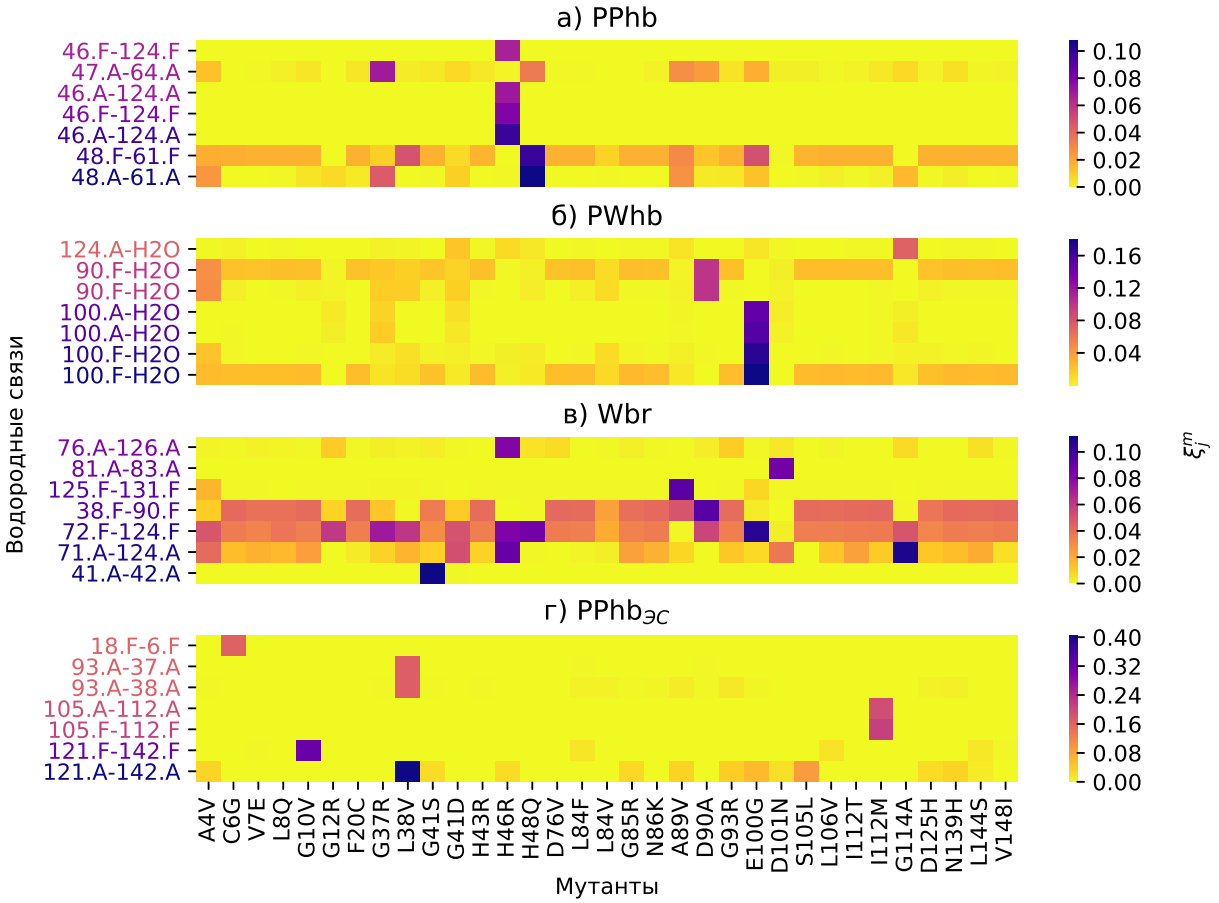
пациентов с мутациями в SOD1, взятые из литературных источников. Из имеющихся в литературе 72 мутаций выбрано 33, для которых описано наибольшее количество случаев у пациентов. Для каждой из этих мутаций в литературе имелась информация о среднем, стандартном отклонении величины дожития пациентов от среднего и количестве известных пациентов с данной мутацией. Для обучения регрессионных моделей использовались, как средние значения (Θ) дожития пациентов с определённой мутацией, так и величины дожития, полученные с помощью выборки из нормального распределения дожития с определёнными по экспериментальным данным средним и стандартным отклонением по каждой анализируемой мутации (Θ_{exp}).

Третья глава посвящена исследованию с помощью молекулярного моделирования (методами МД и ЭС) структурных характеристик белка SOD1 дикого типа и его мутантов.

Было показано, что во многих случаях одни и те же пары аминокислотных остатков в белке SOD1 образуют водородные связи, принадлежащие разным группам (PPhb, Wbr и $\text{PPhb}_{\text{ЭС}}$). Такие водородные связи могут иметь важное значение для структуры белка, поскольку они обеспечивают более стабильный контакт между соответствующей парой остатков. Можно предположить, что изменение стабильности этих водородных связей в результате мутаций SOD1 имеет более сильные последствия для структуры белка по сравнению с уникальными водородными связями, непересекающимися между группами.

Среди водородных связей со значительным отклонением стабильности в мутантах SOD1 от таковой в белке дикого типа (деструктивные водородные связи) чаще других встречались остатки в позициях 46 и 124, и их стабильность наиболее сильно отличалась от стабильности этих связей в белке дикого типа в мутантах Гис46Арг и Гис48Глн (см. рис. 2). С другой стороны, стабильность водных мостиков с участием остатков в позициях 38, 71–72, 90 и 124 заметно отличается от стабильности этих связей практически во всех изучаемых мутантах. Известно, что остатки в позициях 46, 71 и 124 образуют важный контакт между сайтами связывания ионов металлов. Этот обнаруженный факт может говорить о важной роли данных остатков SOD1 в патогенезе БАС.

В результате кластеризации мутантов SOD1 по стабильности водородных связей было обнаружено, что наибольшая часть мутантов попадает в один (в случае $\text{PPhb}_{\text{ЭС}}$) и два (в остальных случаях) крупнейших кластера, а остальные мутанты образовывали кластеры, состоящие только из одного мутанта SOD1 (см. рис. 3). Оказалось, что аминокислотные остатки, формирующие водородные связи с наивысшими вкладами ξ_j^m в мутантах из крупнейших кластеров также участвуют в формировании различных элементов вторичной структуры белка SOD1. В случае PPhb наиболее отличающимися по стабильности водородными связями в мутантах из кластера, содержащего мутант Цис6Гли, от стабильности этих связей в белке дикого типа были связи, аминокислотные остатки которых участвуют также в формировании элементов вторичной структуры, частично содержащих сайты связывания иона меди и цинка (β-тяжи 4 и 7).



Значение ξ_j^m – вклады водородных связей i в среднеквадратичное отклонение стабильностей всех водородных связей мутанта m от таковых для SOD1 дикого типа.

Рис. 2 — Семь водородных связей PPhb (а), PWhb (б), Wbr (в) и $\text{PPhb}_{\text{ЭС}}$ (г) с наивысшим отклонением стабильности в мутантах SOD1 по сравнению с белком дикого типа.

В мутантах из второго крупного кластера (содержащего Ала4Вал) водородные связи с наивысшими вкладами ξ_j^m формировались аминокислотными остатками, которые составляли другую часть сайтов связывания иона меди и цинка (цинк-связывающая петля и β-тяж 4). В случае PWhb в мутантах крупнейшего кластера (с Цис6Гли) дестабилизированными оказались водородные связи, аминокислотные остатки которых содержались в «greek-key» и электростатической петлях, а остатки, формирующие водородные связи с наивысшими ξ_j^m в мутантах второго по величине кластера (с Ала4Вал) расположились в другой части электростатической петли. В случае Wbr мутанты крупнейшего кластера (с Цис6Гли) имели нарушенную структуру в области остатков, которые могут участвовать в формировании агрегатов SOD1, а мутанты другого крупного кластера (с Ала4Вал) формировали отличающиеся по стабильности от белка дикого типа водородные связи в области электростатической петли. В случае $\text{PPhb}_{\text{ЭС}}$ мутанты из единственного крупнейшего кластера имели изменённую по сравнению с белком дикого типа стабильность в области остатков 90–94, которые, как

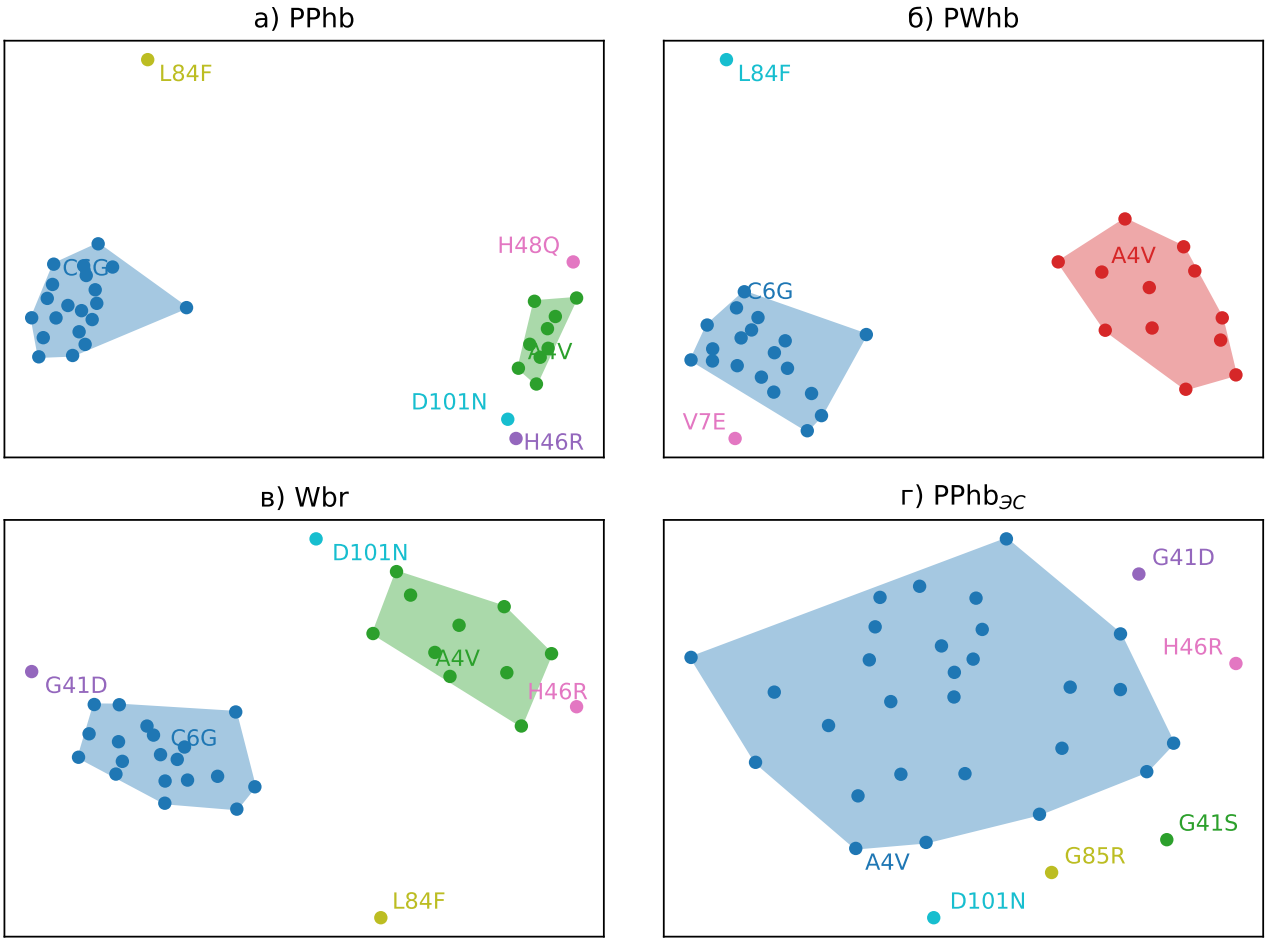


Рис. 3 — Кластеризация мутантов SOD1 по стабильности всех обнаруженных водородных связей PPhb (а), PWhb (б), Wbr (в) и $\text{PPhb}_{\text{ЭС}}$ (г).

известно из литературы, могут участвовать в образовании нестандартных взаимодействий между мутантными SOD1.

В **четвертой главе** описано построение и приведены результаты исследования регрессионных моделей, связывающих дожитие пациентов с БАС со структурными характеристиками мутантных форм SOD1.

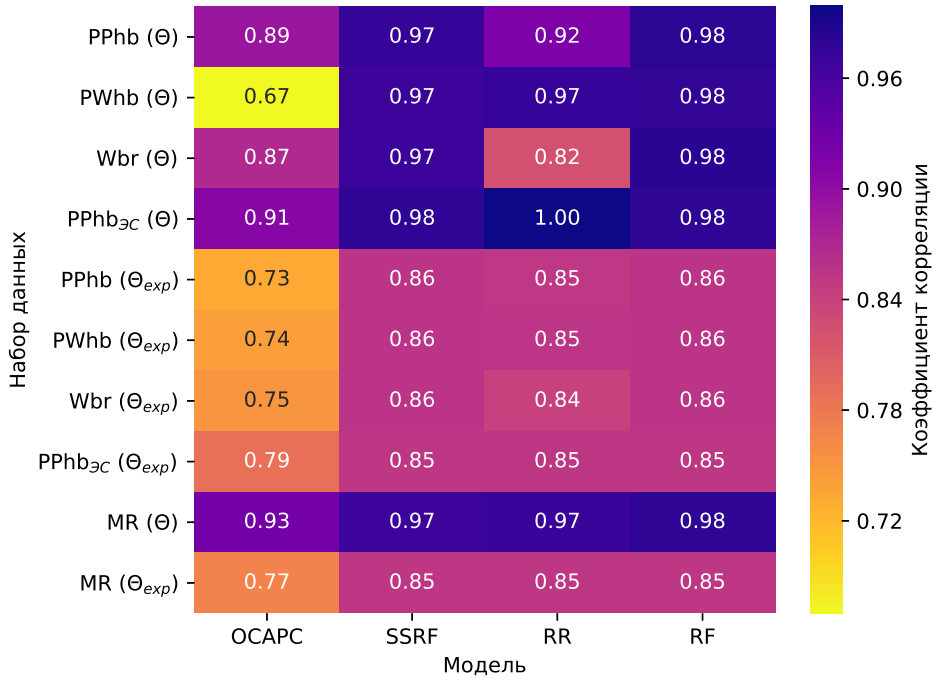
Для предсказания дожития пациентов с мутациями в SOD1 рассматривались как однофакторные, так и многофакторные регрессионные модели. Всего было построено восемь однофакторных и 32 многофакторных регрессионных модели. Однофакторные регрессионные модели строились с использованием линейной регрессии, в которой в качестве единственного фактора выступала оптимальная сумма абсолютных разностей стабильностей водородных связей в белке SOD1 дикого типа и его мутантах (модель ОСАРС). При построении модели ОСАРС строилась таблица $\overline{O}$ абсолютных различий между стабильностью водородных связей в мутантах и белке дикого типа (ДТ): $\overline{O}_i^m = \left| O_i^m - O_i^{\text{ДТ}} \right|$. Строки таблицы $\overline{O}$ сортировались по убыванию абсолютного значения коэффициента корреляции Пирсона между $\overline{O}_i^m$ и дожитием пациентов Θ^m (или ${\Theta_{exp}}^m$), полученным из литературы; порядок строк записывался в вектор $\mathfrak{I}$. Рассчитывалась мера патогенности $\Omega_{\mathfrak{K}}^m$ мутантов ($\mathfrak{K} \subseteq [1,N]$): $\Omega_{\mathfrak{K}}^m = \sum_{k \in \mathfrak{K}, R_k > 0} \overline{O}_k^m -$

$\sum_{k\in\mathfrak{K},R_k<0}\overline{O}_k^m$, – где R_k – коэффициент корреляции строки $\{\overline{O}_k^m\}$ с $\{\Theta^m\}$ (или $\{{\Theta_{exp}}^m\}$), $\mathfrak{K}$ – подмножество номеров связей, удовлетворяющих условию: $|R_k| > R_{th}$. Для определения R_{th} применялась итерационная процедура, которая максимизировала абсолютное значение коэффициента корреляции между $\Omega_{\mathfrak{K}}^m$ и Θ^m. Последовательно для каждого $K \in [1; N]$, на основе набора первых K компонент $\{k\}$, полученного из вектора $\mathfrak{I}$, составлялось множество $\mathfrak{K}_K$ и вычислялось значение ${\Omega_{\mathfrak{K}}^m}_K$. Оценивался коэффициент корреляции R_K между ${\Omega_{\mathfrak{K}}^m}_K$ и Θ^m (или ${\Theta_{exp}}^m$) и в качестве $\mathfrak{K}$ бралось такое $\mathfrak{K}_K$, коэффициент R_K для которого был максимален, а пороговый коэффициент R_{th} в этом случае принимался равным коэффициенту корреляции между строкой $\overline{O}_{\mathfrak{I}_K}^m$ и Θ^m (или ${\Theta_{exp}}^m$). Однофакторные модели были отдельно построены для каждой группы водородных связей (PPhb, $\text{PPhb}_{\text{ЭС}}$, PWhb и Wbr) на основе обучающих выборок Θ или Θ_{exp}.

Многофакторные регрессионные модели включали модель, использующую метод random forest (модель RF) и гребневую регрессию (RR), а также модель, в которой в качестве факторов выступала совокупная стабильность водородных связей в элементах вторичной структуры белка SOD1 (SSRF). Для построения модели SSRF с помощью программы DSSP была получена вторичная структура белка SOD1 (PDBID: 2V0A), в которой для каждой из двух субъединиц белка выделено 23 элемента: три α-спирали, восемь β-тяжей и 12 петель. Все обнаруженные в исследуемых мутантах SOD1 и белке дикого типа водородные связи группировались в зависимости от принадлежности формирующих их аминокислотных остатков одному из элементов вторичной структуры. Водородная связь, образованная различными остатками, попадала в группу водородных связей, соответствующую элементам вторичной структуры, содержащим остаток-донор, а также в группу остатка-акцептора. Если водородная связь была образована между атомами одного и того же остатка, она учитывалась единожды. Для каждой водородной связи, входящей в одну группу рассчитывалась стабильность O. Полученные стабильности O_i^m суммировались для водородных связей из одной группы. Строилась таблица O_s^m стабильностей элементов вторичной структуры SOD1. Всего для четырёх групп водородных связей на основе обучающих выборок Θ^m или ${\Theta_{exp}}^m$ построено 24 таких многофакторных моделей. Для того, чтобы интегрировать результаты предсказаний этих 24 многофакторных моделей и восьми однофакторных моделей строились специальные регрессионные модели MR, основанные на гребневой регрессии. На вход таких моделей подавались результаты предсказания дожития, полученные с помощью моделей RF, RR, OCAPC и SSRF. Например, на вход модели MR_{RF} подавались значения дожития, являющиеся выходными данными четырёх моделей RF, которые были построены на основе четырёх групп водородных связей (PPhb, $\text{PPhb}_{\text{ЭС}}$, PWhb и Wbr). Всего было построено восемь таких гребневых регрессий MR, рассматривающих все отдельно взятые комбинации между четырьмя моделями (RF, RR, OCAPC и SSRF) и двумя обучающими выборками (Θ и Θ_{exp}).

Среднеквадратичная ошибка предсказаний регрессионных моделей оценивалась с помощью метода bootstrap. Дополнительно для оценки точности

результаты предсказаний дожития пациентов сравнивались с результатами, полученными с помощью регрессионных моделей, построенных на наборах случайно выбранных водородных связей (Rhb). Для оценки важности водородных связей для предсказания дожития пациентов с использованием методов RF и SSRF рассчитывалась величина feature importance, которая отражает усреднённую величину снижения среднеквадратичной ошибки предсказаний при перемешивании значений выбранного признака в сравнении с ошибкой предсказаний на основе исходной выборки. При использовании метода OCAPC важность водородных связей определялась с помощью коэффициента корреляции стабильности этих водородных связей с дожитием пациентов. При использовании метода RR важность водородных связей характеризовалась значением коэффициентов уравнения множественной регрессии. С помощью статистических методов исследовалось расположение аминокислотных остатков $\mathfrak{T}$, формирующих водородные связи с наивысшей важностью, в пространственной структуре белка SOD1.



Уровень значимости коэффициентов корреляции для всех регрессионных моделей не превышает 0.001.

Рис. 4 — Значения коэффициентов корреляции между конформационными свойствами мутантов белка SOD1 и дожитием пациентов с БАС с соответствующими мутациями.

Всего на основе анализа стабильности водородных связей в белке SOD1 дикого типа и его мутантах было построено 40 регрессионных моделей для предсказания дожития пациентов с БАС (см. рис. 4). Наиболее точные предсказания были получены с помощью моделей RF_{Θ} ($R = 0.98$, $p < 10^{-22}$) и $\mathrm{RF}_{\Theta_{exp}}$ ($R = 0.85$, $p < 10^{-160}$). Данные модели были построены с помощью метода

random forest. Точность этих моделей в результате перекрёстной проверки составила наименьшую среднеквадратичную ошибку предсказаний, по сравнению с другими моделями, равную 4 года.

Анализ модели SSRF показал, что совокупная стабильность водородных связей, формирующихся с участием элементов вторичной структуры белка, является важным признаком для предсказания дожития пациентов с мутациями в SOD1. Данный результат хорошо согласуется с результатами анализа и других моделей, построенных в данной работе.

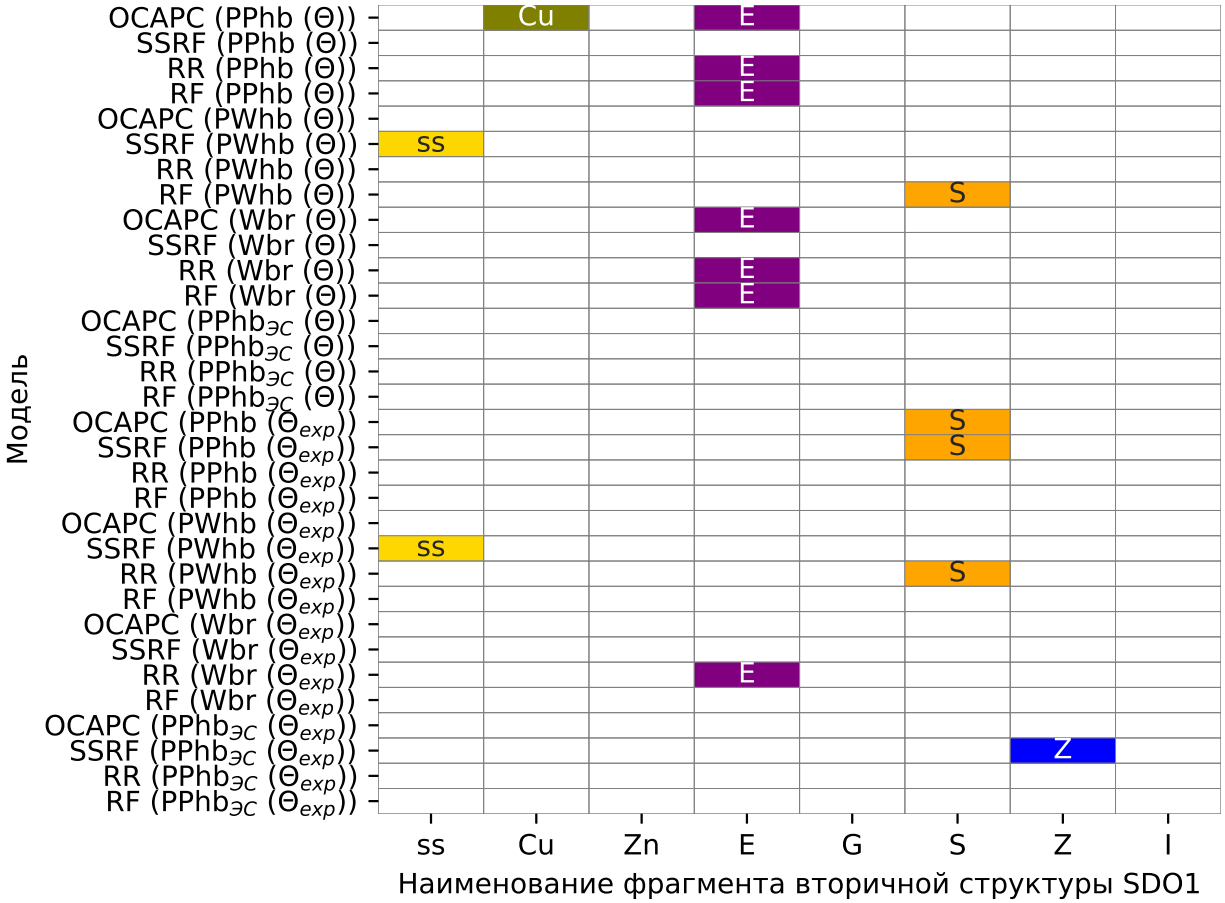
Таблица 1 — Параметры гипергеометрического распределения и уровень значимости пересечения наборов $\mathfrak{T}$ и $\mathfrak{S}^9_{\text{агр}}$.

Набор $\mathfrak{T}$	M	n	N	k	p-значение	p-значение (попр.)
$\mathfrak{T}_{\text{SSRF}}$ (Wbr (Θ))	153	9	9	4	2.2×10^{-5}	0.0008
$\mathfrak{T}_{\text{SSRF}}$ (PPhb (Θ_{exp}))	153	9	5	3	2.8×10^{-5}	0.0008
$\mathfrak{T}_{\text{SSRF}}$ (Wbr (Θ_{exp}))	153	9	9	4	2.2×10^{-5}	0.0008

Примечание: Столбцы «M», «n», «N» и «K» соответствуют параметрам гипергеометрического распределения для сопоставления набора $\mathfrak{T}$ из первого столбца таблицы с набором $\mathfrak{S}^9_{\text{агр}}$. Столбец «p-значение» содержит значимость пересечения позиций остатков из заданной пары наборов $\mathfrak{T}$ и $\mathfrak{S}^9_{\text{агр}}$. Столбец «p-значение (попр.)» содержит p-значения с поправкой Беньямини-Йекутиэли на множественное сравнение.

Следует отметить, что элементы вторичной структуры, важные для предсказания дожития, в основном располагаются в интерфейсе между агрегатами SOD1 (см. таблицу 1). Кроме того, такие элементы вторичной структуры, как оказалось, хоть и располагаются на поверхности белка, но часто содержат гидрофобные остатки, что хорошо объясняется повышенной гидрофобностью интерфейса в целом.

Анализ моделей показал, что в образовании водородных связей, важных для предсказания дожития, часто участвуют аминокислотные остатки, формирующие дисульфидную связь, а также остатки, располагающиеся в сайте связывания иона меди, электростатической, дисульфидной и цинк-связывающей петлях (см. рис. 5). Также следует отметить, что позиции остатков, формирующих важные водородные связи, статистически значимо ($p < 0.02$) пересекаются



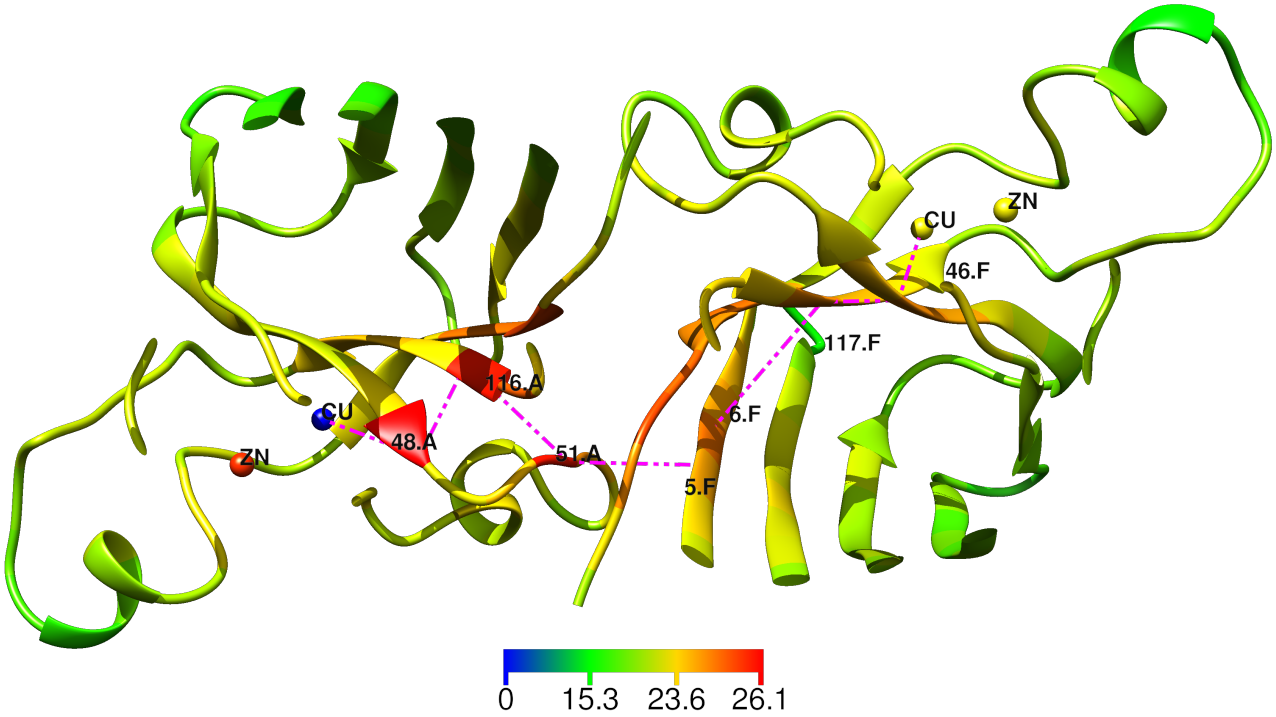
Подписи по оси абсцисс соответствуют элементам вторичной структуры: «ss» – дисульфидная связь; «Cu» и «Zn» – сайты связывания иона меди и цинка; «E», «G», «S» и «Z» – электростатическая, «greek-key», дисульфидная и цинк-связывающая петли; «I» – интерфейс между субъединицами.

Рис. 5 — Диаграмма, отражающая элементы структуры SOD1, позиции остатков которых значимо пересекаются с остатками из $\mathfrak{T}$.

с мутантными позициями. Интересным оказался факт, что остатки, формирующие деструктивные водородные связи также значимо пересеклись с позициями мутаций ($p < 0.003$).

В частности, среди мутаций в области интерфейса между агрегатами SOD1 можно выделить мутацию Вал94Ала. Для этой мутации нет данных о дожитии пациентов. Оказалось, что, согласно предсказаниям, дожитие пациентов – носителей данной мутации – может составить 7.3 ± 2.1 лет. Таким образом, эта мутация может в дальнейшем рассматриваться, как кандидат для генотипирования.

В **пятой главе** приведены результаты исследования крупномасштабной динамики мутантов SOD1 с помощью метода моделирования ЭС. Исследована механическая жёсткость структуры SOD1. Показано, что позиции, обладающие наибольшей средней жёсткостью, располагаются в сайте связывания иона меди, интерфейсе между субъединицами, «greek-key» петле, а также β-тяжах 4, 7 и 8.



Линиями показана возможная цепочка нековалентных взаимодействий между сайтами связывания иона меди обеих субъединиц. Цветом показана величина механической жёсткости.

Рис. 6 — Структура белка SOD1 с отмеченной цветом механической жёсткостью $\kappa_{Cu,i}$ между ионом меди и аминокислотными остатками белка

В структуре белка можно выделить цепочку наиболее близко расположенных в пространстве аминокислотных остатков с повышенной жёсткостью к иону меди, которая может служить путём передачи механических колебаний от сайта связывания иона меди одной субъединицы к сайту связывания иона меди другой субъединицы: Cu.A-48.A-116.A-51.A-5.F-6.F-117.F-46.F-Cu.F (см. рис. 6). Аналогичная картина наблюдалась также для жёсткости между ионом меди субъединицы F и остальными остатками в SOD1. Эти аминокислотные остатки можно рассматривать как интерфейс взаимодействия между активными центрами обеих субъединиц.

Оказалось, что механическая жёсткость между позициями с изучаемыми мутациями в SOD1 и позицией 46, остаток в которой является частью активного центра белка, достоверно отрицательно коррелирует ($R = -0.41$, $p = 0.02$) с дожитием пациентов с соответствующими мутациями (см. рис. 7). Также было показано, что существует достоверная положительная корреляция ($R = 0.57$, $p = 2 \times 10^{-14}$) между средней механической жёсткостью позиций в белке и эволюционной консервативностью остатков в данных позициях. Показано, что позиции с наибольшей эволюционной консервативностью пересекались с позициями из интерфейса между активными центрами SOD1 ($p < 1.0 \times 10^{-6}$, пересечение – остатки: 45–47, 117–118, 148–150).

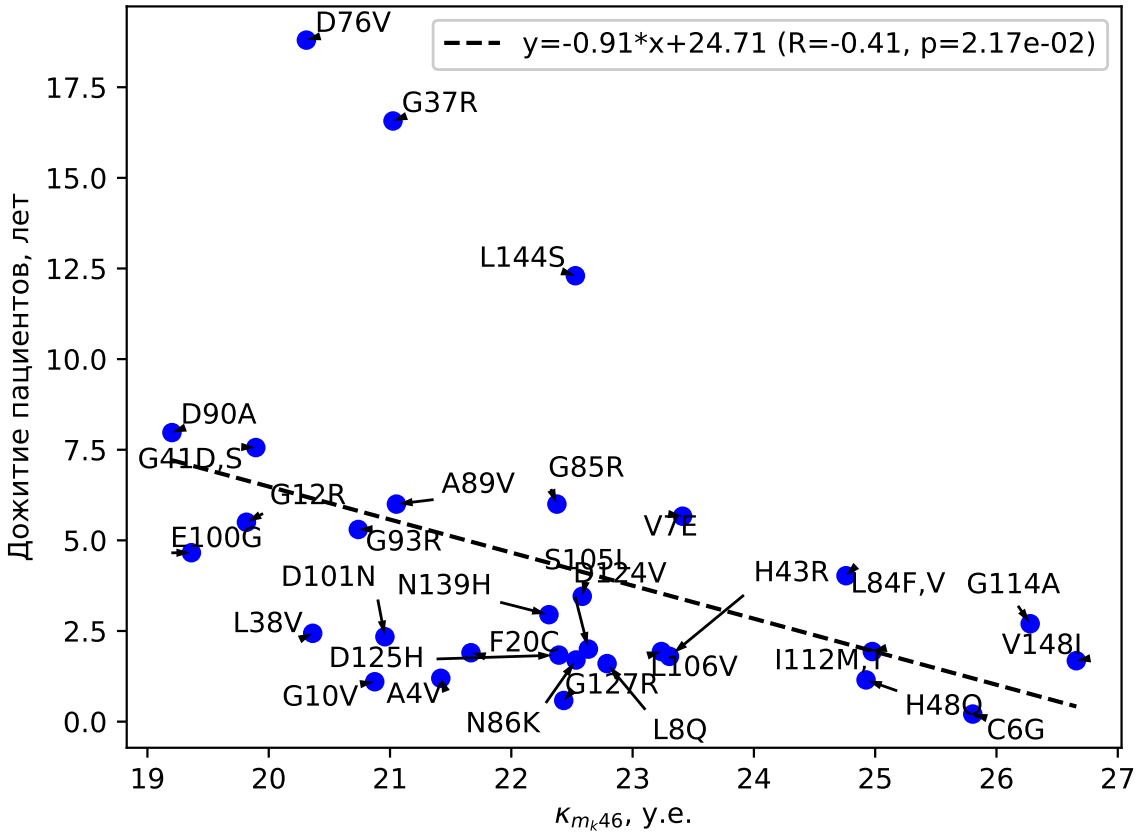


Рис. 7 — График зависимости между механической жёсткостью для пар $(m_k, 46)$ и дожитием пациентов с мутациями в SOD1 в соответствующих позициях m_k

Оказалось, что средняя оценка эволюционной консервативности остатков, в которых имелись изучаемые мутации была значимо выше средней эволюционной консервативности всех остатков SOD1, а средняя оценка эволюционной консервативности остатков, встречающихся в области контактов между агрегирующими SOD1 была значимо ниже средней эволюционной консервативности всех остатков SOD1.

В **заключении** приведено обсуждение механизмов конформационных нарушений в мутантах белка SOD1, связанных с изменением стабильности водородных связей, и их потенциальная роль в нарушении стабильности и влияние на агрегацию мутантов SOD1, ассоциированных с БАС.

Перспективным направлением дальнейшего развития настоящей работы является разработка новых моделей, основанных на анализе стабильности водородных связей для оценки возможного влияния малых химических соединений на конформационные свойства мутантных белков с целью разработки лекарств, корректирующих патогенность мутантов. Кроме того, результаты расширенного анализа мутантов, для которых нет данных о дожитии могут быть использованы для предсказания тяжести заболевания пациентов – носителей данных мутаций.

Особо следует отметить, что разработанная модель предназначена для предсказания тяжести заболевания только для тех мутаций, для которых известно, что они связаны с развитием заболевания БАС.

Выводы

1. Разработан метод расчёта изменения конформационных свойств мутантов белка на основе оценки стабильности внутримолекулярных водородных связей, водородных связей белка с молекулами воды и водных мостиков с использованием метода молекулярной динамики и эластичных сетевых моделей.
2. Построено 40 регрессионных моделей, связывающих дожитие пациентов с БАС со стабильностью водородных связей в белке SOD1, основанных на разных регрессионных методах (линейная регрессия, гребневая регрессия и random forest), а также учитывающих разные группы водородных связей (внутримолекулярные связи, связи между белком и молекулами воды и водные мостики). Наиболее точной оказалась модель random forest ($R = 0.85$, $p < 0.01$), построенная на основе стабильности внутримолекулярных водородных связей, средняя ошибка предсказаний которой при перекрёстной проверке составила 4 года.
3. Показано, что важными для предсказания дожития пациентов являются водородные связи, образуемые остатками, расположенными в интерфейсе между агрегатами SOD1, в сайте связывания иона меди, электростатической, дисульфидной и цинк-связывающей петлях, а также участвующие в формировании дисульфидной связи.
4. На основе анализа механической жёсткости выявлены аминокислотные остатки, которые потенциально могут составлять интерфейс между сайтами связывания ионов меди в димере SOD1. Данный интерфейс может служить путём передачи механических колебаний от сайта связывания иона меди одной субъединицы к сайту связывания иона меди другой субъединицы через остатки Гис48, Тре116, Гли51 (субъединица A) и остатки Вал5, Цис6, Лей117, Гис46 (субъединица F).

Публикации автора по теме диссертации

1. Алемасов Н., Иванисенко Н., Иванисенко В. Структурные и динамические особенности мутантов белка SOD1, ассоциированных с боковым амиотрофическим склерозом // Вавиловский журнал генетики и селекции. — 2014. — Т. 18, № 4/2. — С. 920—927. — (Scopus).
2. Dynamic properties of SOD1 mutants can predict survival time of patients carrying familial amyotrophic lateral sclerosis / N. A. Alemasov [et al.] // Journal of Biomolecular Structure and Dynamics. — 2017. — Vol. 35, no. 3. — P. 645–656. — (Web of Science).

3. Alemasov N. A., Ivanisenko N. V., Ivanisenko V. A. Regression model for predicting pathogenic properties of SOD1 mutants based on the analysis of conformational stability and conservation of hydrogen bonds // Journal of Molecular Graphics and Modelling. — 2017. — Vol. 77. — P. 378–385. — (Web of Science).
4. Molecular mechanisms underlying the impact of mutations in SOD1 on its conformational properties associated with amyotrophic lateral sclerosis as revealed with molecular modelling / N. Alemasov [et al.] // BMC Structural Biology. — 2018. — Vol. 18(Suppl 1), no. 1. — P. 1–14. — (Web of Science).
5. Alemasov N., Ivanisenko N., Ivanisenko V. Structural and dynamical properties of SOD1 protein mutants related to familial amyotrophic lateral sclerosis // German/Russian Network of Computational Systems Biology. — Aachen : Shaker Verlag, 2014. — P. 62–63.
6. Алемасов Н., Иваниесенко Н., Иванисенко В. Молекулярно-динамическая модель для оценки влияния мутаций в белке на фенотипические признаки организма // Тезисы Международной конференции «Актуальные проблемы вычислительной и прикладной математики – 2015». — Новосибирск : Академиздат, 2015. — С. 170.
7. Алемасов Н., Иванисенко Н., Иванисенко В. Структурные и динамические особенности мутантов белка SOD1, ассоциированных с боковым амиотрофическим склерозом // Материалы VII Российского симпозиума «Белки и пептиды». — Новосибирск : ЗАО ИПП «Офсет», 2015. — С. 380.
8. Алемасов Н., Иванисенко В. Раскрытие общих механизмов агрегации мутантов белка SOD1 с помощью эластичного сетевого моделирования // Беляевские чтения. Тезисы докладов Международной конференции, посвященной 100-летию со дня рождения академика АН СССР Д.К. Беляева. Новосибирск, 07-10 августа 2017 г. — Новосибирск : ИЦиГ СО РАН, 2017. — С. 226.
9. Alemasov N., Ivanisenko V. Topological properties of graph of hydrogen bonds forming in SOD1 protein indicate critical regions in its structure // Systems Biology of DNA Repair Processes and Programmed Cell Death (SbPCD-2018) : Symposium (22 Aug. 2018, Novosibirsk, Russia). — Novosibirsk : ICG SB RAS, 2018. — P. 7.