

Федеральное государственное бюджетное научное учреждение «Федеральный
исследовательский центр Институт цитологии и генетики Сибирского отделения
Российской академии наук»



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»

Диссертация на соискание учёной степени

кандидата биологических наук

Научный руководитель:

канд. биол. наук, доцент

Иванисенко Владимир Александрович

Новосибирск — 2018

Оглавление

	Стр.
Введение	6
Глава 1. Обзор литературы	13
1.1 Боковой амиотрофический склероз	13
1.2 Основные причины наследственной формы БАС	14
1.2.1 Шестинуклеотидные повторы в некодирующем регионе C9orf72	14
1.2.2 Мутации в кодирующих областях генов, ассоциированных с БАС	15
1.3 Конформационные свойства белков	19
1.3.1 Межатомные взаимодействия	20
1.3.2 Промежуточные состояния белков	22
1.3.3 Гидрофобные взаимодействия в белках	23
1.4 Современные гипотезы об агрегации белков	23
1.5 Белок SOD1 и его мутанты, как важный фактор развития БАС	25
1.5.1 Структурные особенности SOD1	25
1.5.2 Нестандартные межатомные взаимодействия в мутантах SOD1	28
1.6 Компьютерные методы молекулярного моделирования белков	30
1.6.1 Метод молекулярной динамики	31
1.6.2 Метод эластичных сетей	34
1.7 Регрессионные модели для предсказания патогенности мутаций в белке SOD1	36
1.8 Заключение по обзору литературы	38
Глава 2. Материалы и методы, применяемые для молекулярного моделирования и анализа конформационных свойств белка SOD1	39
2.1 Протокол моделирования МД	39
2.2 Протокол моделирования ЭС	40
2.3 Анализ водородных связей	41

	Стр.
2.4	Статистический анализ 42
2.5	Эволюционная консервативность позиций в белке SOD1 42
2.6	Анализ структуры белков 43
2.7	Остатки SOD1, участвующие в нестандартных межатомных взаимодействиях 43
2.8	Гидрофобные аминокислотные остатки белка SOD1, доступные растворителю 44
2.9	Водородные связи со значительным отклонением стабильности в мутантах SOD1 от таковой в белке дикого типа 45
2.10	Наблюдаемые значения дожития пациентов с мутациями в SOD1 . 46
2.10.1	Отбор мутаций в SOD1 для моделирования и анализа 46
2.10.2	Обучающие выборки для построения регрессионных моделей 47
Глава 3. Анализ результатов молекулярного моделирования динамики белка SOD1 дикого типа и его мутантов 48	
3.1	Моделирование методом МД 48
3.2	Моделирование ЭС 51
3.3	Анализ водородных связей, формирующихся в белке в ходе молекулярного моделирования 52
3.3.1	Водородные связи в SOD1, обнаруженные с помощью моделирования МД и ЭС 52
3.3.2	Водородные связи со значительным отклонением стабильности 57
3.3.3	Кластеризация мутантов SOD1 по стабильности водородных связей 59
3.4	Заключение по главе 3 63
Глава 4. Построение и исследование регрессионных моделей, связывающих дожитие пациентов с болезнью БАС со структурными характеристиками мутантных форм SOD1 . . . 65	
4.1	Регрессионные модели для предсказания дожития пациентов 65
4.1.1	Построение однофакторных регрессионных моделей 65
4.1.2	Построение многофакторных регрессионных моделей 69

	Стр.
4.1.3	Комбинированные регрессионные модели 73
4.1.4	Исследование точности регрессионных моделей 74
4.1.5	Аминокислотные остатки SOD1, формирующие водородные связи, важные для предсказания дожития пациентов 75
4.2	Исследование регрессионных моделей 78
4.2.1	Анализ точности регрессионных моделей 78
4.2.2	Конформационные свойства белка SOD1, важные для предсказания дожития пациентов 82
4.2.3	Связь конформационных свойств мутантов SOD1 и агрегации 91
4.2.4	Предсказание дожития пациентов с мутациями в SOD1, для которых сведения в литературе неполны 93
4.3	Заключение по главе 4 96
Глава 5.	Исследование крупномасштабной динамики мутантов SOD1 99
5.1	Механическая жёсткость структуры белка SOD1 99
5.1.1	Позиции в белке с наибольшей средней жёсткостью 99
5.1.2	Жёсткость между активными центрами белка 99
5.1.3	Отличие жёсткости к важным остаткам от средней жёсткости в белке 101
5.1.4	Связь механической жёсткости и патогенности мутаций в SOD1 102
5.1.5	Зависимость между механической жёсткостью и эволюционной консервативностью аминокислотных остатков 104
5.2	Важные области в структуре SOD1 104
5.3	Заключение по главе 5 107
Заключение 109
Выводы 113
Список сокращений и условных обозначений 114

	Стр.
Словарь терминов	116
Список литературы	117
Список рисунков	139
Список таблиц	143
Приложение А. Таблица наблюдаемых величин дожития пациентов с мутациями в SOD1	144
Приложение Б. Наблюдаемые величины дожития пациентов с мутациями в SOD1, выбранные для построения регрессионных моделей	150

Введение

Боковой амиотрофический склероз (БАС) – неизлечимое нейродегенеративное заболевание. Выделяется две его формы: наследственная и спорадическая [1]. Известно, что второй по распространённости и наиболее изученной причиной наследственной формы БАС являются мутации в гене *SOD1*, кодирующем фермент супероксиддисмутаза-1 [2]. Одной из гипотез относительно механизма возникновения заболевания на молекулярном уровне является агрегация белка SOD1, вызванная его неверной укладкой в результате мутаций [2—4].

Большинство существующих работ [5—11], 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 [12—15]. Так, в одном из *in silico* исследований структуры и динамики SOD1 уже анализировались водородные связи и сеть, которую они образуют в этом белке [16]. Другое исследование с применением метода молекулярной динамики (МД) [17] показало, что структура SOD1 становится более гибкой в мутантных белках по сравнению с белком дикого типа [18]. 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 [19]. 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 [20; 21].

С помощью метода молекулярной динамики в данной работе было исследовано 39 мутаций в белке SOD1 [22]. 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 [23]. 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) [24; 25]. 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.

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., Иванисенко 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.

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

Глава 1. Обзор литературы

1.1 Боковой амиотрофический склероз

Боковой амиотрофический склероз (БАС) – заболевание, характеризующееся неотвратимой дегенерацией центральных и периферических моторных нейронов, приводящее к прогрессирующему параличу и возможной смерти от дыхательной недостаточности [26—29]. Частота встречаемости заболевания составляет 1-3 случая на 100000 человек в год [26; 30—32].

БАС, как заболевание, был впервые описан в 1824 году Чарльзом Бэллом [33]. Большой вклад в описание патологии БАС внёс Жан Крювелье в 1852 году, а Франсуа Аран ещё в 1848 описал ряд случаев БАС, хотя не разделял заболевание по происхождению на нервное и мышечное [33; 34]. Сам термин «боковой амиотрофический склероз» был впервые упомянут в работе Жана Мартена Шарко в 1874 году [33].

По разным данным от 1% до 25% случаев БАС имеют наследственную природу [26; 35—39]. Практически во всех случаях заболевание наследуется доминантно, чаще с высокой пенетрантностью [39]. По данным секвенирования полного экзона человека, порядка тридцати генов вовлечены в заболевание БАС [40]. Среди этих генов можно выделить следующие: *SOD1*, *TARDBP*, *FUS*, *OPTN* и *VCP*. Взятые вместе, мутации в этих генах объясняют около 25% наследственных случаев БАС [37]. Кроме того, известен некодирующий регион C9orf72, в котором увеличенное количество шестинуклеотидных повторов GGGGCC также приводит к БАС [37; 41].

Средний возраст возникновения БАС по клиническим данным, взятым из базы данных ALSod [42], на сегодняшний день составляет около 47 лет (стандартное отклонение 17.5 лет), а среднее дожитие пациентов варьируется от двух до четырёх лет [32; 43]. Одним из предположений о том, каким образом токсическое влияние от накопления внутри клетки агрегатов мутантных белков проявляется лишь в более зрелом возрасте, является то, что в зрелом возрасте деление клеток замедляется, что приводит к пониженной скорости обновления тканей и компонентов клетки, и, следовательно, к повышенной нагрузке на системы контроля качества клетки [44]. Более того, выдвинута гипотеза о том, что заболевание

БАС, как и ряд разновидностей рака, развивается многоступенчато (по-видимому, существует шесть таких ступеней), а среди факторов риска выделяются: генетическая предрасположенность, условия окружающей среды и образ жизни [32; 45]. Выдвигаются также предположения, что причиной возникновения БАС могут являться вирусные инфекции или активация ретровирусов [32; 39].

В 1995 году на рынок вышел препарат «Рилузол», который продлевает жизнь человека на два-три месяца [46]. В 2016 году было предложено ещё одно лекарство, «Радикава», которое в мае 2017 года было одобрено FDA. Утверждается, что «Радикава» способно облегчить симптомы протекания болезни, снижая окислительный стресс в клетке [47; 48].

1.2 Основные причины наследственной формы БАС

1.2.1 Шестинуклеотидные повторы в некодирующем регионе C9orf72

C9orf72 – некодирующая область хромосомы 9, соответствующая открытой рамке считывания 72 [41]. Было показано, что у здоровых людей максимальное количество повторов GGGGCC составило 23. У людей из выборки VSM-20 (по первым буквам организаций, выполнявших исследование: Vancouver, San Francisco и Mayo, семья 20), страдающих одним из заболеваний: лобно-височная деменция, БАС – или обоими, количество шестинуклеотидных повторов в C9orf72 варьировалось от 700 до 1600. На основе дополнительного анализа данных пациентов с БАС было выявлено 4.1% пациентов, страдающих спорадической формой заболевания и 23.5% пациентов – наследственной формой, которые в то же время имели увеличенное количество повторов GGGGCC в C9orf72.

Данный некодирующий участок хромосомы 9 изучался также другими авторами [37]. При исследовании 405 финских пациентов и 497 здоровых людей обнаружено, что локус на хромосоме 9p21 объясняет до четверти всех случаев БАС и до половины всех случаев его наследственной формы [49]. В последующем исследовании также был выделен участок C9orf72, количество шестинуклеотидных повторов GGGGCC в котором отделяло здоровых людей от больных БАС [37]. На основе анализа данных от нескольких семей с БАС, обнаружено, что у

здоровых людей повторов было меньше 20, а у больных – больше 30. Анализ 402 финских пациентов с БАС и 478 здоровых людей выявил, что количество шести-нуклеотидных повторов увеличено у 28.1% пациентов и 0.4% здоровых людей. Всего пациентов с наследственной формой БАС, у которых увеличено количество повторов GGGGCC, оказалось 46.4%, а среди пациентов со спорадической формой заболевания было 21% с повышенным числом повторов. В среднем количество повторов у пациентов было 53, у здоровых людей – 2. Таким образом, C9orf72 отвечает за наибольшее количество случаев заболевания БАС.

1.2.2 Мутации в кодирующих областях генов, ассоциированных с БАС

Мутации в гене *SOD1*

Второй по распространённости, объясняющей до 20 % случаев, и наиболее изученной причиной наследственной формы БАС являются мутации в гене *SOD1*, кодирующем фермент супероксиддисмутаза-1 [2; 50—52]. На данный момент известно свыше 180 мутаций этого гена, вызывающих различные формы БАС (<http://alsod.iop.kcl.ac.uk/>). Все известные мутанты *SOD1* разделяются на две группы: подобные белку дикого типа и мутации в области связывания ионов металлов [53; 54]. Известно, что потеря ферментативной активности *SOD1* не приводит к БАС [2]. Считается, что мутации в *SOD1* вызывают в белке возникновение новой токсической функции [1]. В настоящий момент предложен ряд гипотез относительно механизма возникновения заболевания на молекулярном уровне [1; 2; 55; 56]. Среди них:

1. повышенный окислительный стресс [1; 2; 50],
2. митохондриальная дисфункция [1; 2; 57],
3. эксайтотоксичность глутамата [1; 2],
4. накопление белковых агрегатов с участием *SOD1*, вызванное неверной укладкой в результате мутаций [1—4; 51; 54; 57—60],
5. самоподдерживающаяся токсичность клеток отличных от нервных (таких, как астроциты и микроглия) [1; 2],

6. нарушение гомеостаза металлов, снижение стабильности мутантов SOD1 при удалении ионов металлов [1; 2; 51; 54; 60],
7. дисфункция натрий/калиевого ионного насоса [1],
8. аутофагия [1; 2],
9. нарушение систем аксонального транспорта [1; 50],
10. посттрансляционные модификации белка, которые снижают стабильность димера SOD1 [61],
11. окислительная модификация SOD1 [51; 57; 60],
12. стресс эндоплазматического ретикулума [50].

Образование агрегатов мутантных SOD1 подтверждается для пациентов с наследственной формой БАС, но не находит подтверждения у пациентов со спорадической формой БАС [60]. Высокий уровень экспрессии человеческого SOD1 дикого типа в линии мышей с мутацией G85R ускоряет появление заболевания [60]. Из этого следует, что сверхэкспрессия SOD1 дикого типа человека сама по себе приводит к появлению токсической функции или усиливает токсическую функцию мутантных белков [60]. Развивая эту идею, мутантным SOD1 приписываются прион-подобные характеристики [62].

Помимо того, что мутации в SOD1 могут повлечь за собой агрегацию данных мутантных форм SOD1, есть сведения, указывающие на связь мутантных SOD1 с апоптозом [63]. В упомянутой работе говорится о прямой ассоциации SOD1, как дикого типа, так и мутантов, с анти-апоптотическим белком BCL-2. Выявлен регион BCL-2, с которым происходит связывание SOD1. При этом, SOD1 дикого типа выполняет анти-апоптотическую функцию, в то время как мутантные SOD1 могут её терять. Кроме того, BCL-2 захватывается мутантными SOD1, агрегированными в нерастворимый комплекс, что ведёт к тому, что BCL-2 теряет возможность выполнять анти-апоптотическую функцию и провоцирует гибель двигательных нейронов. Также известно, что мутации в *SOD1* вызывают первичную открытоугольную глаукому [64].

Мутации в гене *TARDBP*

Мутации в белке TDP-43, кодируемом геном *TARDBP* также приводят в 1% случаев к заболеванию БАС, как наследственной, так и спорадической

формы [65]. Заметные включения, содержащие убиквитин, как считается, являются признаком этого заболевания [50; 66; 67]. В то же время среди белков, составляющих агрегат главный компонент – TDP-43 [66; 68; 69]. TDP-43 – ДНК/РНК-связывающий белок, имеющий множество функций в клетке. Изначально TDP-43 был классифицирован, как репрессор транскрипции, который связывается с TAR-элементом ВИЧ. Из-за своего молекулярного веса он был назван TDP-43 [67]. Белок TDP-43 участвует в регуляции экспрессии генов и сплайсинге, присутствует в комплексе, который производит сплайсинг гена *CFTR* и, вероятно, также участвует в биогенезе микро-РНК, апоптозе и клеточном делении [67; 68].

Известно, что накопление гиперфосфорилированных фрагментов TDP-43 в перикариии нейронов у пациентов с БАС связано со значительной потерей TDP-43 в ядре [68]. За исключением одной, все известные мутации TDP-43 сосредоточены в С-терминальной части белка – глицин-богатом районе, который, вероятно, участвует в связывании с другими белками, включая разнообразные рибонуклеопротеины [65]. Кроме того, подтверждено наличие в образцах мозга пациентов с БАС меньшей фосфорилированной части (25 кДа) С-концевого фрагмента TDP-43, а также крупных убиквитинированных агрегатов, включающих данный белок [65]. Обнаружено, что TDP-43 связывается с высокой аффинностью с убиквилином-2 (UBQLN2) – подобным убиквитину белком семейства UBQLN [70]. Мутации в UBQLN2 известны в качестве генетического маркера доминантного Х-сцепленного варианта возникновения БАС. Вместе с тем UBQLN2 усиливает выведение, как TDP-43, так и С-концевых фрагментов, то есть может влиять на цитотоксичность [70].

Мутации в гене *FUS*

Обнаружена связь гена *FUS*, кодирующего ещё один ДНК/РНК-связывающий белок, мутации в котором объясняют порядка 4–5 % случаев наследственной формы БАС [1; 71—75]. Известно более 50 мутаций в белке *FUS* [76]. Этот белок участвует в процессах метаболизма РНК таких как: транскрипция, сплайсинг, трансляция [76], перенос в цитоплазму, – а также в репарации

ДНК [74; 75; 77]. В нейронах FUS содержится в ядре, но переносится в дендритные шипики в комплексе с РНК или другими РНК-связывающими белками [72; 74]. Мутации в FUS наиболее часто встречаются в С-концевом участке белка; среди таких мутаций – R521C и R521H [74].

В нервной системе пациентов с мутациями в FUS находятся цитоплазматические включения [73]. Показано что тяжесть проявления мутаций в FUS определяет количество цитоплазматических включений и подверженность клеток внешнему стрессу [74]. Показано *in vivo* и *in vitro*, что FUS формирует безмембранные растворимые гранулы, которые при возникновении мутаций, не обязательно в домене низкой сложности, переходят в нерастворимые агрегаты [75; 78]. При этом данные нерастворимые агрегаты FUS также приводят к тому, что в них попадают и другие рибонуклеопротеиды (РНП). В результате нарушается синтез новых белков на концах аксонов, где гранулы РНП регулируют трансляцию и метаболизм локальных РНК [78].

Помимо этого, показано *in vivo*, что мРНК *MECP2* – одного из генов-мишеней FUS – колокализуются с нерастворимыми агрегатами мутантных FUS в цитоплазме [77]. Другими словами, мутации FUS также существенно влияют на экспрессию целевого гена. Сверхэкспрессия немутантных FUS, как утверждается, приводит к дегенерации двигательных нейронов [76]. На мышах показано, что БАС возникает не в результате потери функции мутантами FUS, а из-за приобретения ими новой токсичной функции [76; 79]. Возраст пациента, при котором у него диагностируется заболевание коррелирует со степенью неверной локализацией в цитозоли [79]. Таким образом, мутантные FUS вызывают токсические события, как в двигательных нейронах, так и в соседних с ними клетках, что приводит к БАС [79].

Мутации в гене *OPTN*

Мутации в гене *OPTN*, кодирующем белок оптинейрин (OPTN), также приводят к наследственной форме БАС [80; 81]. До этого было известно, что мутации в *OPTN* также вызывают первичную открытоугольную глаукому. Агрегация белка OPTN является отличительной чертой большого количества случаев БАС [82].

Показано, что гомозиготная нонсенс-мутация Q398X в белке OPTN, а также гетерозиготная миссенс-мутация E478G в убиквитин-связывающем домене OPTN прекращают ингибирование активации ядерного фактора NF-κB [80], хотя это исследование проводилось только *in vitro*, с использованием сверхэкспрессирующих клеточных линий [83]. Оказалось также, что мутант E478G образует OPTN-иммунореактивные цитоплазматические включения [80; 82].

Оптинейрин – положительный регулятор ТВК-1 опосредованного сигнального пути интерферона [83]. Белок дикого типа связывается с белком LC3 аутофагосомы при митофагии, что вызывает сборку аутофагосомы вокруг повреждённой митохондрии [82]. Но мутант E478G оптинейрина, связанный с БАС, препятствует такой сборке, что предлагается в качестве причины заболевания [81; 82]. Подавление экспрессии *OPTN* в линии нервных клеток приводит к увеличению активности NF-κB и последующей гибели клеток, что называется возможным механизмом, приводящем к БАС [83]. Показано также, что аутосомно-доминантный мутант E478G проявляет доминантно-негативные свойства [82].

Мутации в гене *VCP*

Ген *VCP* кодирует валозин-содержащий белок, играющий важную роль в созревании убиквитин-содержащих аутофагосом [84; 85]. Белок *VCP* участвует в освобождении стрессовых гранул, и мутации в нём препятствуют нормальному протеканию аутофагии [85]. Его токсическое для нейронов действие выражается во влиянии на белок TDP-43 – основную составляющую убиквитиновых включений в клетке [84]. Мутации в *VCP* объясняют порядка 1 – 2 % наследственных случаев БАС [84].

1.3 Конформационные свойства белков

Конформация белков – определённая пространственная структура, образующаяся за счёт внутримолекулярных взаимодействий в белках [86]. Белки с одинаковой последовательностью аминокислотных остатков и находящиеся в

одинаковых условиях принимают одинаковую конформацию, которая определяет функцию белка [86]. Конформация белка, в которой происходит его активное функционирование, называется нативной конформацией или нативной структурой [86]. Большинство белков в ходе синтеза и последующего функционирования могут принимать целый ряд промежуточных по отношению к нативной конформаций [87; 88]. В каждой конформации между атомами белков возникают различные взаимодействия, которые влияют на структурные, динамические и функциональные свойства молекулы [86; 88—91]. При исследовании этих свойств в белках обращают внимание на следующие типы межатомных взаимодействий [86; 88—92]:

1. пептидные связи;
2. водородные связи;
3. Ван-дер-Ваальсовы взаимодействия;
4. солевые мостики (или ионные взаимодействия);
5. дисульфидные связи;
6. катион- π взаимодействия;
7. ароматические (или π - π стэкинг) взаимодействия.

1.3.1 Межатомные взаимодействия

Пептидная связь, являясь ковалентной, образуется между α -карбоксильной и α -аминогруппой соседних аминокислот и составляет полипептидный остов белка [86]. Свободная энергия гидролиза пептидной связи составляет порядка 12 кДж/моль [93]. Энергия активации при гидролизе пептидной связи составляет по разным данным 84–120 кДж/моль [88; 94; 95].

Водородная связь – нековалентная межатомная связь [96]. Водородные связи могут быть межмолекулярными или внутримолекулярными [96]. Энергия водородного связывания ($R-X-H \cdots B-Y$) составляет, в среднем, 20 кДж/моль [91; 96]. Водородные связи играют важную роль, в частности, в образовании и поддержании вторичной и третичной структуры белков [12—15; 91; 96; 97].

Ван-дер-Ваальсово взаимодействие является невалентным [91]. В его основе лежат взаимодействия электрических диполей в соседних молекулах [91].

Различается три типа таких дипольных взаимодействий: ориентационные, индукционные и дисперсионные [91]. В среднем, энергия Ван-дер-Ваальсового взаимодействия составляет, порядка, 6 кДж/моль [91]. Характерной особенностью этого взаимодействия является то, что на больших расстояниях преобладает сила притяжения, в то время, как на малых – отталкивания [91]. Ван-дер-Ваальсово взаимодействие участвует в формировании структуры, в частности, биспиральных нуклеотидов, а также влияет на вторичную структуру биомолекулы и даёт заметный вклад в стабилизацию и изменение её конформации [91; 97].

Солевые мостики – невалентные взаимодействия ионизованных боковых групп в белках [90]. Отрицательный заряд карбоксильных групп аспарагиновой и глутаминовой кислот, а также С-концевого остатка может взаимодействовать с положительными зарядами на аминоклуппах боковых цепей лизина и N-концевых аминокислот, гуанидиниевых группах боковых цепей аргинина и заряженных имидазольных кольцах гистидина [90]. Свободная энергия стабилизации белка солевым мостиком составляет порядка 15–20 кДж/моль [97]. Как и другие невалентные взаимодействия, солевые мостики вносят заметный вклад в стабильность белковой структуры, а также в специфичность взаимодействия белков с другими биомолекулами [90; 97].

Дисульфидная связь образуется между парой остатков цистеина при их окислении и относится к ковалентным связям, которые стабилизируют конформацию белков [86; 90; 98]. Образующийся при таком связывании дисульфид называется цистином [90]. Дисульфидные связи чаще всего образуются в белках, функционирующих во внеклеточном пространстве, что, как полагают, связано с необходимостью стабилизации их структуры и предотвращает их денатурацию [86]. Энергия диссоциации дисульфидной связи, в среднем, составляет 260 кДж/моль [98]. Данный тип связи играет важную роль для образования и поддержания третичной структуры белков [98].

Катион- π взаимодействия являются нековалентными и играют важную роль в молекулярном распознавании, влияют на структуру макромолекул и на связывание их с низкомолекулярными соединениями [99; 100]. В белках данный тип межатомных взаимодействий происходит между боковыми группами лизина или аргинина с фенилаланином, тирозином или триптофаном [100]. Энергия связывания лигандов увеличивается на 8–20 кДж/моль, если связывание происходит через катион- π взаимодействие [100].

Ароматические взаимодействия являются нековалентными и в белках возникают между ароматическими аминокислотными остатками – фенилаланином, тирозином, гистидином и триптофаном [101—103]. Эти взаимодействия имеют энергию 8–12 кДж/моль и сравнимы по энергии с водородными связями [104]. Ароматические взаимодействия играют важную роль в укладке белка и поддержании её стабильности [101; 103].

1.3.2 Промежуточные состояния белков

Белки в процессе синтеза проходят через ряд промежуточных состояний [87; 88]. Конформационные перестройки возникают также в процессе выполнения белком своей функции – от передачи сигналов до ионного транспорта [105—107]. Промежуточные состояния проявляют отчасти свойства нативного белка, а отчасти – денатурированного, – и могут являться стабильными [108; 109]. Эти промежуточные конформации могут возникать в результате изменения условий среды (рН, давления или температуры), присутствия некоторых растворителей и солей, а также в результате мутаций [106; 108]. Одним из следствий влияния внешних по отношению к организму (среда) и внутренних (мутации, посттрансляционные модификации) факторов на белок является изменение его конформации с нативной на допускающую его агрегацию с другими биомолекулами [106; 109].

Частным случаем функционального изменения конформации белка может также служить аллостерия – перестройка конформации молекулы в ответ на воздействие на одну из её областей (аллостерический сайт), например, при связывании в этой области некоторого лиганда [110—112]. При этом, между аллостерическим сайтом и областью с изменённой конформацией не обнаруживается прямого взаимодействия, а наблюдается только динамическая связь [112]. Чаще всего белки, подверженные аллостерии, состоят из нескольких субъединиц, хотя аллостерия может наблюдаться и в системах с одним доменом [110; 111].

1.3.3 Гидрофобные взаимодействия в белках

В литературе различается гидрофобный эффект – низкая растворимость гидрофобных растворов в воде, – а также гидрофобная сила или гидрофобное взаимодействие – сильное притяжение гидрофобных поверхностей и групп в воде [113]. Известно, что гидрофобные взаимодействия влияют на структуру белков [114], причём сильнее влияют на стабильность более крупных белков (341 аминокислотный остаток против 36) [115]. Утверждается, что гидрофобные взаимодействия играют важную роль в функционировании систем «контроля качества» белков в клетке [116] и являются причиной агрегации белков [117; 118].

1.4 Современные гипотезы об агрегации белков

Эукариотические клетки – сложные структуры, способные разграничивать происходящие внутри биохимические реакции в пространстве и времени. Ключевым механизмом такого разграничения является разделение внутриклеточного пространства на функциональные области. Один из хорошо известных примеров таких разделённых областей – клеточные органеллы, отделённые от внешнего пространства внутриклеточными мембранами, действующими, как физический барьер. Последние исследования открыли также немембранный способ разграничения реакций – так называемое «расслоение жидкостей» (liquid demixing) – разделение различных растворов, находящихся в жидкой фазе [75; 119]. Этот способ, по-видимому, используется клеткой для динамической перестройки внутриклеточного пространства. Например, когда требуется временно провести некоторую реакцию в ограниченной от других реакций области. Известны примеры функционального расслоения жидкостей в эукариотических клетках: ядрышко, Р-гранулы (околоядерные РНК-гранулы), стрессовые РНК-гранулы, тельца Кахаля, «ядерные пятна» (paraspeckles), Р-тела (processing (P) bodies) [75; 119]. Фактически, механизм разделения жидких фаз является универсальным, в частности, для образования частиц рибонуклеопротеинов. Другими примерами немембранных многобелковых комплексов являются centrosomes, сигнальные молекулы (мембранные рецепторы) и ядерные поры.

Вне парадигмы разделения фаз агрегация определяется, как любая ассоциация двух или более молекул белка, находящихся в отличной от нативной конформации [120; 121]. Белковые агрегаты характеризуются низкой растворимостью, нарушенной внутри- или внеклеточной локализацией и изменённой вторичной структурой [120]. Предположительно, агрегация происходит в результате взаимодействия частично упорядоченных промежуточных форм белков [120]. Различают, в частности, аморфные, неупорядоченные агрегаты и упорядоченные волокна или амилоиды, с кросс- β структурой [121]. Считается, что подверженность белка агрегации зависит, главным образом, от физико-химических свойств аминокислотных остатков белка и последовательностью остатков в первичной структуре, а также конформационной стабильности белка в нативном состоянии и концентрации белка в клетке [121]. В рамках парадигмы разделения фаз агрегация соответствует одному из трёх различных состояний, в которых могут находиться молекулы в клетке: рассеянное (газообразное), жидкое, твёрдое [119]. Так, применительно к белкам, газообразное состояние соответствует раствору, в котором белки находятся отдельно друг от друга; жидкое состояние – растворимым белковым комплексам, отделённым от других «расслоением жидкости»; твёрдое – нерастворимым белковым агрегатам.

Одним из классов белков, демонстрирующих повышенную подверженность агрегации, являются функционально неупорядоченные белки (от англ. *Intrinsically disordered proteins*, IDP). Для них вероятность фазового разделения, по-видимому, должна быть довольно высокой, поскольку данный тип белков из-за своей структурной гибкости способен создавать множественные взаимодействия с другими белками [119]. Это происходит за счёт конформационной подвижности присутствующих в таких белках неупорядоченных областей и областей с низкой сложностью аминокислотного состава (от англ. *Low complexity regions*, LCR). При этом, известно, что белки с более протяжёнными неупорядоченными участками имеют меньшее время существования в клетке [122; 123]. Подобная же закономерность наблюдается для белков, не подверженных и подверженных агрегации: чем более белок подвержен агрегации, тем более быстрый «круговорот» в клетке он имеет, и тем более низкая его концентрация поддерживается [124].

В здоровой клетке обратимо формируются РНК-гранулы нескольких типов: стрессовые гранулы, Р-тела и нейрональные или транспортные гранулы, – в которых задействованы, наряду с другими РНК-связывающими белками, в частности, белки FUS и TDP-43 [125]. У пациентов с БАС образование РНК-гранул,

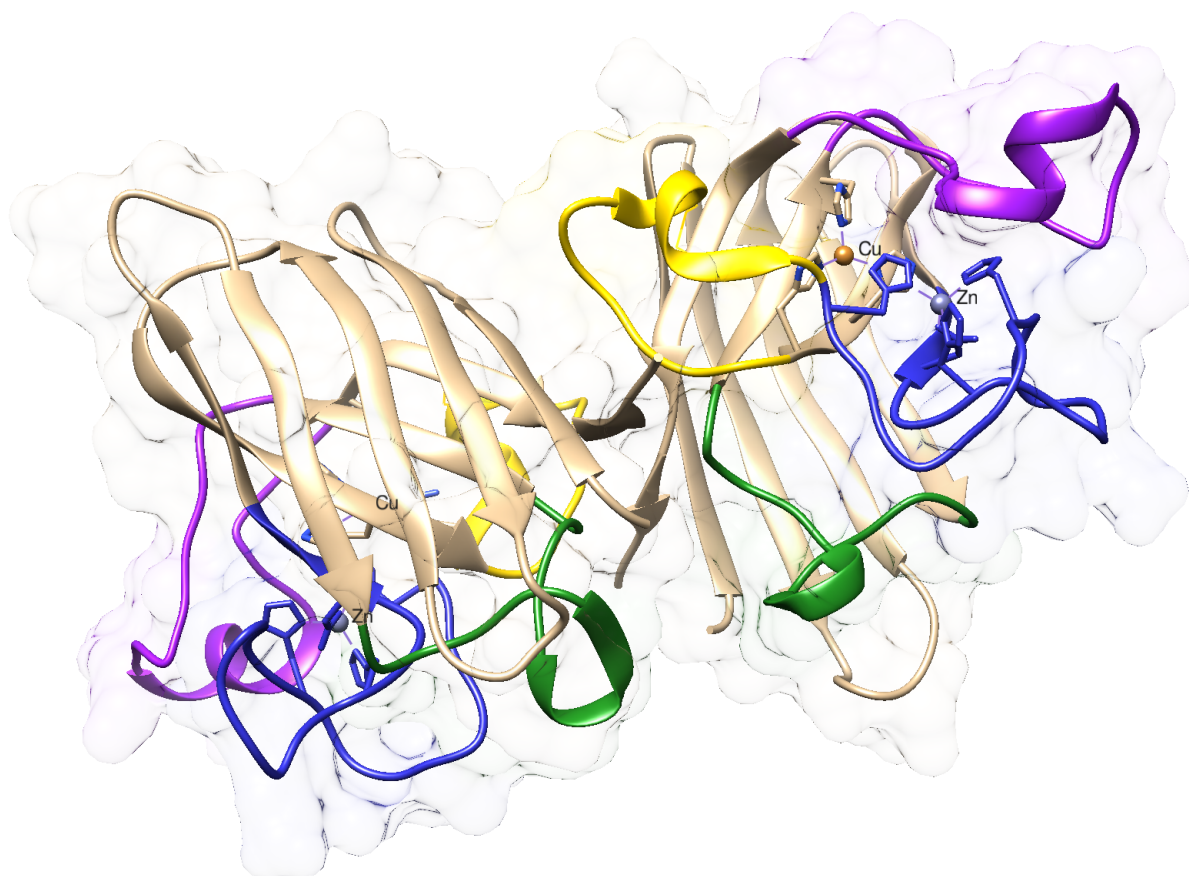
как предполагается, приводит к возникновению нерастворимых включений, состоящих из белковых агрегатов [125]. Присутствие белковых агрегатов в клетке может говорить о дисбалансе системы поддержания протеостаза (от англ. protein homeostasis) [44; 126]. В частности, накопление в клетках убиквитинированных включений является признаком того, что протеасома более не справляется с деградацией повреждённых белков [39; 68; 125; 126]. Нарушения в укладке белков, которые предшествуют их агрегации, происходят с возрастом и являются результатом ослабления в работе систем контроля качества в клетке, включающих шапероны, аутофагию и убиквитин-протеасомную деградацию белков [44].

1.5 Белок SOD1 и его мутанты, как важный фактор развития БАС

1.5.1 Структурные особенности SOD1

Белок супероксиддисмутаза-1 (SOD1) – гомодимер. Он катализирует реакцию превращения анионов супероксида на молекулярный кислород и пероксид водорода. Каждая из двух субъединиц белка состоит из 153 аминокислотных остатков и содержит в себе два иона: медь и цинк [127]. Сайт связывания иона меди представлен остатками 46, 48, 63, 120; сайт связывания иона цинка – 63, 71, 80, 83 [128]. Каждая субъединица имеет в своей вторичной структуре β -бочонок, составленный из 8 β -тяжей [129]. В субъединицах присутствует по одной дисульфидной связи между остатками 57 и 146. В структуре мономера присутствует семь петель, часть из которых выделяется тем, что несёт важную функцию [127; 129; 130]. Так, например, в кристаллической структуре SOD1 (PDBID: 2V0A) выделяются следующие петли:

1. дисульфидная, остатки: 49-62;
2. цинк-связывающая, остатки: 63-85;
3. «greek-key», остатки: 102-115;
4. электростатическая, остатки: 121-142.



Жёлтым цветом выделена дисульфидная петля, синим – цинк-связывающая, зелёным – «greek-key», фиолетовым – электростатическая петля. Ионы металлов выделены сиреневым (цинк) и оранжевым (медь).

Рисунок 1.1 — Пространственная структура белка SOD1.

Интерфейс связывания обоих мономеров представлен аминокислотными остатками: 5, 7, 17, 50-54, 113-115, 148, 150-153 [57]. Известны позиции аминокислотных остатков, находящихся в контакте между агрегатами SOD1: 11-15, 24, 26, 91-92, 97-99, 101-104, 109, 128-131 [131—135].

На молекулярном уровне мутации приводят к различным изменениям в структуре белка SOD1. Известно, что ион цинка стабилизирует структуру SOD1 [136; 137]. Следовательно, дестабилизация сайта связывания цинка, цинк-связывающей или электростатической петель ведёт к потере белком стабильности [138]. То же справедливо и для дисульфидной связи и содержащей один из её концов петли [136; 139]. В свою очередь, ион меди отвечает за ферментативную активность белка [2; 138; 140]. Показано, что ион меди стабилизирует гидрофобное ядро бета-бочонка даже больше, чем ион цинка [141]. Следовательно, дестабилизация сайта связывания иона меди также приведёт к снижению стабильности белка [137; 142]. С другой стороны, показано, что только отсутствие

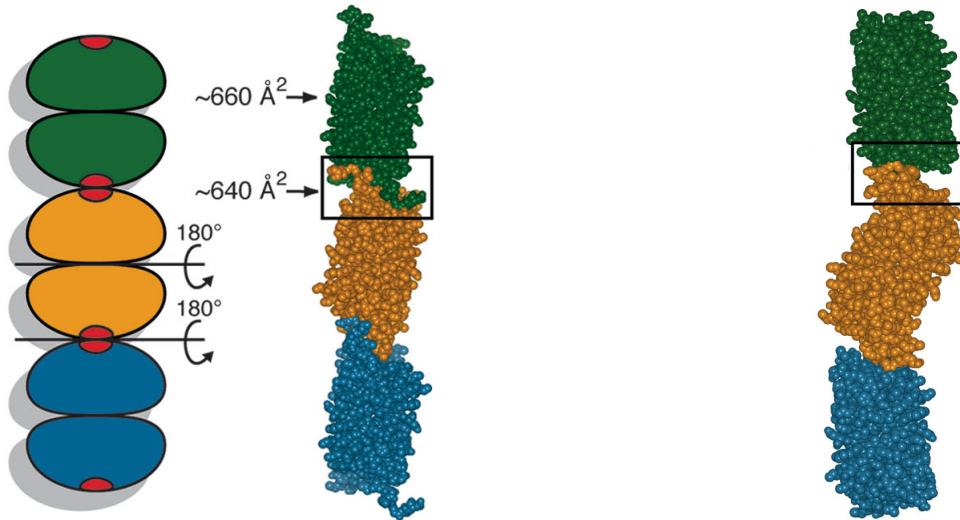
иона меди и снижение ферментативной активности не приводит к заболеванию БАС [2; 140].

Известно, что дестабилизация интерфейса между субъединицами увеличивает вероятность мономеризации SOD1, и, следовательно, может привести к агрегации белка [16; 143]. Также показано, что мономеризация SOD1 сама по себе не приводит к агрегации, но если в тот же момент дестабилизирована дисульфидная связь и нарушено связывание иона цинка, то это приводит к формированию белковых агрегатов и последующему накоплению нерастворимых включений в клетке [144]. Кроме того, показано, что глобальная стабильность SOD1 не связана с БАС, так как известны мутанты белка, стабильность и ферментативная активность которых неотличима от белка дикого типа [10; 145]. В нативной форме SOD1 выполняет антиоксидантную функцию в клетке с привлечением ионов меди для протекания ферментативных реакций [2]. Отсутствие иона меди, включённого в структуру SOD1 наряду со снижением ферментативной активности белка напрямую не приводит к БАС [2]. Однако патогенные мутации могут вызывать агрегацию SOD1 и накопление этих агрегатов в клетке, нарушая, в частности, функционирование протеасомы [134]; некоторые мутанты демонстрируют нормальную ферментативную активность, в то время как другие имеют активность ниже таковой в белке дикого типа или даже совсем теряют её [146]. Все эти, зачастую противоречивые, сведения дают понять, что в структуре белка SOD1 существует сложная сеть межатомных взаимодействий, которая определяет его глобальную и локальную стабильность наряду с подверженностью агрегации.

Сложность экспериментальных исследований агрегации SOD1 заключается в том, что они, чаще всего, проводятся на линиях модельных животных в условиях сверхэкспрессии мутантных белков. Это затрудняет определение действительных механизмов агрегации [147]. Другой сложностью является то, что позиции, в которых находятся известные мутации в SOD1, распределены равномерно по всей его структуре [54; 148]. То есть остаётся неясным, как мутации в различных областях SOD1 могут приводить к повышению подверженности этого белка агрегации.

1.5.2 Нестандартные межмолекулярные взаимодействия в мутантах SOD1

В ряде работ упоминается, что некоторые мутации в SOD1 могут приводить к возникновению новых, отсутствующих в белке дикого типа, контактов с другими молекулами SOD1 [131—135].



а) Линейные волокна

б) Волокна типа «зигзаг»

Разным цветом выделены смежные молекулы SOD1, составляющие агрегат.

Рисунок 1.2 — Пример амилоидных волокон (а) линейного вида (б) в виде «зигзага». По рис. 2 и 3 из работы [131].

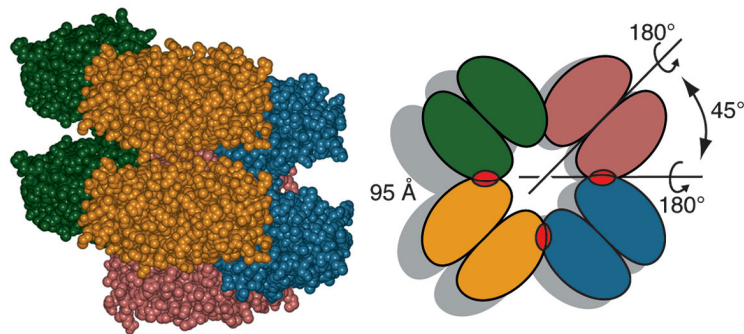


Рисунок 1.3 — Пример агрегата в виде спирали. По рис. 4 из работы [131].

В частности, существуют несколько подобных взаимодействий, в результате которых, как считается, формируются амилоиды в виде линейных нитей и «зигзагов», а также в виде спиралей [131] (см. рис. 1.2 (а, б) и 1.3). Ниже перечислены подобные контакты в порядке: аминокислотный остаток одной молекулы и контактирующий с ним остатком другой.

1. Лей42 с Лей126;

2. Гли129 с Вал88;
3. Гли130 с Фен45;
4. Гли130 с Вал87;
5. Гли130 с Тре88;
6. Гли130 с Вал97;
7. Гли130 с Сер98;
8. Гли130 с Иле99;
9. Арг79 с Асп101;
10. Арг79 с Вал103.

В более поздней работе этих же авторов указываются вероятные взаимодействия между смежными молекулами мутантов Гис46Арг SOD1, образующих агрегаты в виде волокон, а также взаимодействия между волокнами линейного типа и типа «зигзаг» [132]. В таких нестандартных контактах обнаруживаются следующие аминокислотные остатки и их группы:

1. Асп11 с Лиз36;
2. Про13 с Про13;
3. Асп109 с Асп109;
4. Асп92-Лиз36;
5. Гли12-Про13 и Гли37-Лей38-Тре39 с теми же остатками другой молекулы;
6. Про13-Вал14-Гли15, Лиз36-Гли37-Лей38-Тре39 и Лиз91-Асп92 с теми же остатками другой молекулы.

В работе [133] также сообщается о возможных контактах между молекулами SOD1, образующих агрегат в виде «зигзага». Такие контакты включают следующие пары взаимодействующих остатков:

1. Асн26 с Проб6;
2. Асн26 с Сер102-Вал103-Иле104;
3. Глу24 с Асп109;
4. Лиз128 с Асн86;
5. Гли129 с Асн86 и Тре88;
6. Гли130 с Асн86 и Сер98;
7. Асн131 с Сер98.

Мутации в SOD1 приводят к формированию агрегатов в виде волокон из-за заметного изменения в структуре белка [134]. Авторами обнаружено три области

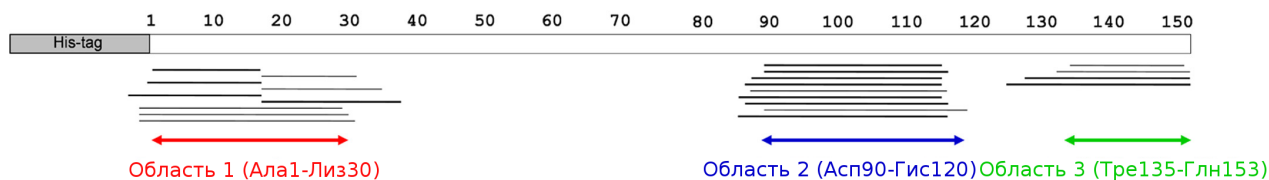


Рисунок 1.4 — Области в первичной структуре белка, соответствующие пептидам, устойчивым к действию протеазы. По рис. 1 из работы [134].

в первичной структуре SOD1, которые играют роль «строительных блоков» при образовании агрегатов (см. также рис. 1.4):

1. Ала1–Лиз30;
2. Асп90–Гис120;
3. Тре135–Глн153.

В этой же работе предлагается три модели, описывающих образование агрегатов трёх разных типов в зависимости от контактирующих областей белка:

1. Область 1 формирует контакт с областью 2 в одном белке и через данную область 2 взаимодействует с областью 3 в другом.
2. Область 1 формирует контакт с областью 3 в одном белке и через данную область 3 взаимодействует с областью 1 в другом.
3. Область 1 формирует контакт с областью 2 в одном белке и через данную область 2 взаимодействует с областью 1 в другом.

Другими авторами выделяется область Ала1–Гис63, которая в эксперименте *in vitro* образовывала агрегаты в виде волокон [135].

1.6 Компьютерные методы молекулярного моделирования белков

Существует, по крайней мере, два источника, на основе которых можно изучать межатомные взаимодействия в белке: статическая структура (например, кристаллическая из базы данных PDB) и ансамбль конформаций, полученный с помощью одного из методов моделирования динамики белка или из экспериментов ЯМР. На данный момент в базе данных PDB содержится ограниченное количество структур белков из экспериментов ЯМР, что ограничивает их использование для изучения белка дикого типа и мутаций в нём. Среди наиболее часто используемых методов моделирования динамики выделяется метод молекулярной динамики [17], метод Монте-Карло [149] и крупнозернистое моделирование

(см., например, обзор [150]). Для получения крупномасштабной динамики белка в рамках методов полноатомного моделирования – молекулярной динамики и Монте-Карло – требуются, как правило, исключительные вычислительные ресурсы [150]. Одним из подходов к крупнозернистому моделированию белков является метод моделирования эластичных сетей [23]. Показано, что данный подход может применяться для исследования крупномасштабной динамики белков [151—153], которая труднодостижима при использовании МД.

1.6.1 Метод молекулярной динамики

Важнейшим средством теоретического исследования структуры, динамики и термодинамических свойств комплексов биомакромолекул является метод МД, который является методом компьютерного моделирования, позволяющим в течение заданного периода времени проследить эволюцию системы взаимодействующих атомов с помощью численного интегрирования уравнений движения [17; 154]. Метод МД широко используется для решения задач исследования термостабильности белков, конформационных переходов, транспорта молекул, белкового фолдинга и прочих.

Разброс временных масштабов различных физических явлений огромен, от 10^{-8} сек (времена движения доменов) до 1-10 сек (время самоорганизации белков и нуклеиновых кислот). Ещё 10 лет назад исследования конформационных переходов макромолекул могли быть сделаны только для систем малых белков и пептидов [155; 156], а для больших систем, размером более 1 миллионов атомов, фазовые траектории исследовались в ограниченном временном интервале не превышающем 50 нс [157]. В последнее время с ростом производительности вычислительной техники и суперкомпьютеров, с помощью МД исследуются модели биологических систем, состоящих из миллионов и даже миллиардов атомов [158].

Метод молекулярной динамики состоит в численном интегрировании уравнений движения Ньютона для системы взаимодействующих частиц (атомов), находящихся в некоторых начальных условиях (заданы начальные координаты и скорости каждого атома):

$$F_i = -\frac{\partial U(r_1, \dots, r_N)}{\partial r_i}$$

где $i = 1 \dots N$,—координаты атома i ; N —число атомов.

Движение атомов описывается уравнениями:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \quad (1.1)$$

где m_i —масса атома i .

Типичное МД-исследование предполагает расчёт траектории развития системы во времени и её изучение на предмет интересующих свойств. При этом точность расчётов зависит от выбранной функции потенциальной энергии U , форма которой вместе с необходимыми для её использования параметрами составляют силовое поле.

Силовое поле

В соответствии с квантовой теорией, для небольших молекул энергия системы может быть определена методами квантовой химии. Но при числе атомов более десяти задача становится трудно вычислимой [159]. Поэтому в МД применяются такие потенциалы, которые зависят только от положения атомов, что позволяет исследовать системы с десятками-сотнями тысяч и даже миллионами атомов.

В существующих силовых полях МД функция потенциальной энергии разбивается на сумму вкладов от различных типов взаимодействий: $U = U_b + U_\theta + U_\phi + U_{el} + U_{LJ}$. Среди них взаимодействия:

1. Связующие

а) растяжение связей:

$$U_b = \frac{1}{2} \sum K_b (r - b_0)^2$$

где b_0 —равновесная длина связи; K_b —силовая константа.

б) изменение валентных углов:

$$U_\theta = \frac{1}{2} \sum K_\theta (\theta - \theta_0)^2$$

где θ_0 —равновесное значение угла; K_θ — силовая константа.

в) изменение двугранных углов:

$$U_{\Phi} = \sum K_{\Phi} [\cos(n\Phi - \delta) + 1]$$

где n —кратность торсионного барьера; δ —сдвиг фазы; K_{Φ} —константа, определяющая высоту потенциальных барьеров двугранных углов.

2. Несвязующие

Возникают между парами атомов, принадлежащими различным молекулам. Либо принадлежащие той же самой молекуле, но разделённые, по крайней мере, тремя связями.

а) электростатические:

Кулоновский потенциал может быть разбит на две части — ближнюю и дальнюю:

$$U_{el} = \sum \frac{1 - \operatorname{erf}(\beta r_{ij})}{\epsilon r_{ij}} q_i q_j + \sum \frac{\operatorname{erf}(\beta r_{ij})}{\epsilon r_{ij}} q_i q_j$$

где q_i, q_j —парциальные заряды на атомах; ϵ —диэлектрическая проницаемость среды.

Дробь в первой сумме ведёт себя как $1/r$ при $r \rightarrow 0$, убывает экспоненциально при $r \rightarrow \infty$ и отвечает за ближнюю часть взаимодействий, вторая дробь стремится к $1/r$ при $r \rightarrow \infty$ и отвечает за дальнюю часть. β —параметр, определяющий относительный вклад между прямой суммой и суммой в обратном пространстве, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ —функция ошибки.

б) ван-дер-ваальсовы (потенциал Леннарда-Джонса):

$$U_{LJ} = \sum \left[\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right]$$

где A, B зависят от типов атомов i и j ; r_{ij} —расстояние между этими атомами.

Параметры силовых полей для потенциалов взаимодействий определяются из различных экспериментальных данных (спектральные, калориметрические, кристаллографические) и квантово-химических расчётов. Наиболее часто используются такие семейства полей, как: AMBER [160—166], CHARMM [167—170], GROMOS [171], OPLS [172], CVFF (CFF, PCFF) [173].

1.6.2 Метод эластичных сетей

Для изучения коллективных движений в макромолекулах применяются методы анализа нормальных мод [151]. Для описания коллективных движений в белке обычно достаточно бывает от одной до трёх мод с наименьшей частотой [151; 152]. Показано, что анализ нормальных мод может быть проведён с использованием простого гармонического потенциала с одним параметром [23]. В дальнейшем было показано, что даже упрощённой модели белка, в которой атомы аминокислотного остатка в белке заменены точечной массой, оказывается достаточно для получения корректных низкочастотных движений [152]. Такие упрощённые модели называются эластичными сетевыми моделями и входят в разновидность крупнозернистых (coarse grained) моделей [174]. Замена целой группы атомов одной частицей позволяет снизить количество степеней свободы всей молекулярной системы и, таким образом, перейти на более крупный масштаб динамики белка.

Эластичные сетевые модели применяются, например, для описания аллостерических переходов в двигательных белках, изучения броуновской динамики, изучения изменения структуры белков в процессе эволюции, структурного выравнивания [175], а также докинга [176], определения промежуточных состояний белков [107] и изучения эволюционных аспектов функционирования белков [177]. Кроме того, применение эластичных моделей нашлось для эффективного исследования конфигурационного пространства, предсказания путей конформационных переходов в белке, сравнения динамики эволюционно схожих белков, предсказания динамики связывания, при аллостерической регуляции активного сайта [178]. Эластичные модели позволяют проводить исследования, которые потребовали бы исключительных вычислительных ресурсов при использовании традиционных полноатомных методов моделирования [136]. Потенциал, использованный в [23] задан таким образом, что минимален в случае, когда моделируемая система находится в позиции, соответствующей термодинамическому равновесию, что позволяет избежать этапа минимизации структуры, необходимого, например, при моделировании методом МД [175].

Анализ нормальных мод

Суть анализа нормальных мод заключается в разложении функции потенциальной энергии $U(x)$ в ряд Тейлора в окрестности точки x_0 [179]. Если градиент потенциала стремится к 0 в этой точке и если отбросить множители третьего порядка и выше, то динамика системы может быть описана в терминах направлений Q_i и частот ω_i нормальных мод:

$$M^{-\frac{1}{2}} F M^{-\frac{1}{2}} Q_i = \omega_i^2 Q_i$$

$$Q_i \cdot Q_j = \delta_{ij}$$

где M – матрица, содержащая на диагонали массы атомов, а гессиан F – содержащая вторые производные потенциала в точке x_0 . Для молекулы, состоящей из N атомов нахождение нормальных мод равносильно численной диагонализации матрицы размера $3N \times 3N$.

Моды колебаний называются нормальными, поскольку являются ортогональными по отношению к тензору кинетической энергии системы [178]. Из нормальности моды следует, что положение атома в пространстве является линейной комбинацией положений для отдельных мод, а также полная энергия может быть представлена в виде суммы энергий для каждой из мод [178].

Коллективность движений атомов

Коллективность отражает долю атомов молекулы, которые в значительной степени затронуты выбранной модой колебаний [151]. Максимальная доля коллективности составляет единицу для совместных движений целой группы атомов, в то время как локальные флуктуации лишь нескольких атомов выражаются в коллективности, стремящейся к нулю. Формально индекс коллективности выражается следующей формулой [180; 181]:

$$\kappa = \frac{1}{N} \exp\left(-\sum_{i=1}^N \Delta \bar{r}_i^2 \log \Delta \bar{r}_i^2\right)$$

где $\Delta \bar{r}_i^2$ – изменение координат атома i для выбранной моды колебаний, N – количество атомов; \bar{r}_i^2 имеют такой масштаб, чтобы $\sum_{i=1}^N \Delta \bar{r}_i^2 = 1$ [181].

Механическая жёсткость

Одним из подходов к изучению механической жёсткости между различными областями в белке является моделирование ЭС с помощью анизотропных сетевых моделей (АСМ) [182]. Такой подход рассматривает ситуацию, в которой к определённой паре аминокислотных остатков белка, представленных в виде ЭС, прикладывается сила в направлении прямой, проходящей через центры данных остатков [183]. Тогда механическая жёсткость κ_{ij} между парой остатков i и j имеет смысл константы жёсткости эквивалентной «пружины» между ними. Проведён ряд исследований, применявших механическую жёсткость, в частности, к изучению гибкости структурно близких белков [184], механическому описанию амилоидных фибрилл [185], моделированию микротрубочек [186], изучению чувствительности белков к растягивающей силе [187], и определению механического ответа белков на анизотропные деформации [188].

1.7 Регрессионные модели для предсказания патогенности мутаций в белке SOD1

В литературе предложено несколько регрессионных моделей для предсказания дожития пациентов с мутациями в SOD1, страдающими БАС [7; 10; 189]. В одной из этих моделей в качестве независимых факторов применялись две меры: подверженность мутантов SOD1 агрегации (вероятность, что белок будет агрегировать) и «нестабильность» этих мутантов (вероятность, что белок разупорядочится), что позволило объяснить около 69 % дисперсии в дожитии пациентов [7]. Подверженность белка агрегации определялась из изменений в физико-химических свойствах таких, как гидрофобность, вторичная структура (включая потерю α -спиралей и возникновение β -тяжей), а также изменение заряда молекулы в результате мутаций.

В рамках другой модели было показано, что дожитие пациентов со дня первой диагностики заболевания отрицательно коррелирует ($R = -0.78$) с потерей термостабильности мутантов, носителями которых они являлись [10]. Однако эта закономерность хорошо работала только для ограниченного круга мутаций, в то время как эффекты большого количества мутаций, связанных с изменением заряда аминокислотных остатков, не подчинялись построенной зависимости. Интересно, что годом ранее в исследовании на клеточных культурах не удалось объяснить, в общем, связь агрегации с термостабильностью, изменением заряда и др. [9].

В третьей модели в качестве независимой переменной выступала энергия гетеродимеризации ($\Delta\Delta G_{Het}$), и её применение для пяти из шести исследованных мутантов SOD1 позволило объяснить до 98 % дисперсии в дожитии пациентов с соответствующими мутациями [189]. Тем не менее, значение переменной для одного мутанта выбивалось из построенной зависимости, из чего авторы работы сделали вывод о том, что энергия гетеродимеризации – только один из важных факторов в определении токсичности SOD1. Другими факторами могут выступать различные меж- и внутримолекулярные взаимодействия [189].

Кроме того, недавно было проведено исследование восьми компьютерных подходов к предсказанию дожития пациентов с БАС с мутациями в SOD1, в котором наибольший полученный коэффициент корреляции не превысил 0.33 [11]. В упомянутом исследовании, в частности, изучалась корреляция между экспериментально измеренной для 30 мутантов SOD1 величиной изменения свободной энергии фолдинга ($\Delta\Delta G$) и каждым из трёх известных из литературы показателей патогенности этих мутантов (возрастом диагностики заболевания, дожитием пациентов и возрастом их смерти). Корреляция обнаружилась между экспериментально измеренной $\Delta\Delta G$ и дожитием пациентов ($R = 0.4$), а также между $\Delta\Delta G$ и возрастом смерти пациентов ($R = 0.4$). Рассчитанный с помощью программы I-mutant 3.0 показатель $\Delta\Delta G$ коррелировал с дожитием пациентов лишь с коэффициентом корреляции $R = 0.33$, а наибольший коэффициент корреляции ($R = 0.47$, программа ENCoM) рассчитанного значения $\Delta\Delta G$ был получен в его корреляции с возрастом смерти пациента.

1.8 Заключение по обзору литературы

Существует ряд наследственных причин возникновения БАС, из которых наиболее изученной являются мутации в гене *SOD1*. В литературе имеются, зачастую противоречивые, экспериментальные сведения об основном механизме развития БАС. Большое количество публикаций указывает на то, что накопление белковых агрегатов с участием белка *SOD1*, вызванное неверной укладкой в результате мутаций, является одним из важнейших аспектов в патогенезе БАС. Агрегация белков, как таковая, может вызываться нарушением в работе различных подсистем в клетке, включая аутофагию и убиквитин-протеасомную деградацию белков.

В литературе упоминаются аминокислотные остатки, участвующие в нестандартных межатомных контактах между мутантными *SOD1*. Предполагается, что данные контакты приводят к формированию в нервных клетках белковых агрегатов с участием *SOD1* и возникновению заболевания БАС. Данные взаимодействия между белками часто опосредованы водородными связями, имеющими одну из самых высоких энергий и наибольшее распространение среди нековалентных типов связей. Поэтому анализ сети водородных связей, образующихся в белке, может представлять интерес для выявления механизма влияния мутаций на его структуру и конформацию.

Для компьютерного исследования агрегации белка *SOD1*, кроме термостабильности его мутантов, по-видимому, следует принимать во внимание характеристики водородных связей, которые также играют важную роль в поддержании вторичной и третичной структуры белка. Чаще всего для теоретического компьютерного исследования структуры, динамики и термодинамических свойств комплексов биомакромолекул используется метод МД. Более упрощённым, с физической точки зрения, теоретическим методом, но позволяющим исследовать обширный набор конформаций белков, является метод моделирования ЭС. При этом возникает возможность сопоставить результаты моделирования упомянутых двух методов с тем, чтобы получить дополнительное описание конформационных и динамических свойств моделируемой биологической молекулы.

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.

Глава 2. Материалы и методы, применяемые для молекулярного моделирования и анализа конформационных свойств белка SOD1

2.1 Протокол моделирования МД

В работе для моделирования в рамках метода МД был применён программный комплекс AMBER 12 [190]. В качестве аппаратной платформы выступал гибридный высокопроизводительный кластер ЦКП «Биоинформатика» (<http://bioinformatics.bionet.nsc.ru/>), который содержит ускорители NVIDIA Tesla M2090. При моделировании МД применён следующий набор параметров:

- силовое поле – ff99SB [163];
- шаг интегрирования – 2 фс;
- радиус отсечения взаимодействий – 10 Å;
- температура – 300 К;
- время моделирования – 50 нс.

В качестве модели воды использовалась модель TIP3P [191]. Кубическая ячейка имела сторону 12 Å и состояла из более чем 20000 молекул воды, растворённого в ней белка (PDB ID: 2V0A) из 4376 атомов. В зависимости от заряда мутантного белка в «раствор» добавлялось порядка 36 и 30 соответственно ионов Na^+ и Cl^- , что соответствовало концентрации 0.137 моль. Для обеспечения достоверности результатов моделирования каждая траектория МД повторялась 5 раз.

Белок SOD1 для нормального функционирования в клетке требует включения в свой состав ионов меди и цинка. Параметры сайта связывания атома цинка получены Иванисенко Н.В. с помощью расчётов в рамках теории функционала плотности (DFT). Был применён обменный функционал M06 с базисным набором 6–31+G* [192] из пакета Gaussian 09 [193] с эталонной экспериментальной структурой (PDBID: 2V0A). Далее, параметры были адаптированы для силового поля с помощью модуля МТК++/MCPB [194] из пакета программ AmberTools14. Известно, что на один димер приходится ~ 0.2 атома меди, а цинка ~ 1.5 [195]. Поэтому моделирование осуществлялось без учёта иона меди.

2.2 Протокол моделирования ЭС

Получение пространственных структур мутантных белков SOD1 производилось с помощью моделирования мутаций в структуре белка дикого типа (PDBID: 2V0A) с использованием программы FoldX [196]. Проверка и исправление полученных структур мутантов, включая удаление пересечений координат аминокислотных остатков и оптимизация ориентации остатков также проводилась с помощью FoldX.

Построение модели ЭС для белка дикого типа и его мутантов проводилось с помощью системы ElNemo [151] с параметрами по умолчанию:

- Количество нормальных мод с наименьшей частотой (NMODES) было установлено в значение 5.
- Минимальное и максимальное значение параметра возмущения (DQMIN/DQMAX) и шага между DQMIN и DQMAX (DQSTEP) было установлено, соответственно, в -100, 100 и 20.
- Количество остатков для группировки внутренней утилитой diagrtb (NRBL) было установлено в «auto», что означало, что это значение определялось автоматически для оптимизации скорости вычислений.
- Радиус отсечения для определения эластичных взаимодействий (CUTOFF) был установлен в 8 Å.

В рамках каждой такой модели был получен ансамбль из 11 конформаций для каждой из 5 нетривиальных мод колебаний с наивысшим показателем коллективности. Таким образом, для белка дикого типа и каждого из изучаемых мутантов были рассчитаны 55 конформаций, представляющих крупномасштабные флуктуации структуры этих молекул. С помощью пакета программ AMBER структура, соответствующая каждой конформации, впоследствии минимизировалась с целью избежать физически недостижимых торсионных углов, длин связей и других параметров SOD1. Водородные связи в каждой из 55 конформаций белка находились с помощью утилиты crrtraj из пакета программ AMBER.

Для проверки корректности моделирования ЭС рассчитывалась корреляция между собственными векторами для одной и той же моды колебаний для белка дикого типа и изучаемых мутантов SOD1. Проверялось значение коэффициентов корреляции с тем, чтобы удостовериться, что структурные изменения именно

вследствие мутаций привели к различиям в компонентах собственных векторов, а не в результате альтернативного порядка собственных значений.

Механическая жёсткость κ_{ij} между всеми парами позиций в первичной структуре белка рассчитывалась с помощью пакета ProDy для Python [182]. Также рассчитывалась величина средней жёсткости каждой позиции по-отдельности, как средняя жёсткость всех возможных пар позиций в белке, в которые входит данная позиция:

$$\bar{K}_i = \frac{1}{153} \sum_{j=1}^{153} \kappa_{ij} \quad (2.1)$$

2.3 Анализ водородных связей

Водородные связи были получены путём анализа ансамблей конформаций (из траекторий МД или конформаций ЭС) с помощью утилиты `srptraj` из комплекса программ AMBER. Утилита `srptraj` обнаруживает водородные связи, образующиеся внутри мутантной формы $m \in [1, M]$ белка SOD1, между атомами мутанта m и окружающими её молекулами воды (только для МД) и водные мостики (только для МД), возникающие между аминокислотными остатками в m . Программа `srptraj` возвращает число $o_{ij}^m \in [0; 1]$ – время существования (или стабильность) каждой обнаруженной связи $i \in [1, N]$, нормированное на количество конформаций в ансамбле $j \in [1, n]$. Здесь N – количество водородных связей, образующихся в процессе динамики белка SOD1, M – количество исследуемых форм белка SOD1 – дикого типа и мутантов, n – количество ансамблей j конформаций ($n = 5$ для МД, $n = 1$ для ЭС). В случае связей между атомами белка и молекулами воды o_{ij}^m может превышать 1 и достигать 4, поскольку молекулы воды могут формировать до четырёх водородных связей [88]. Водородная связь в `srptraj` определялась по межатомному расстоянию, которое не должно превышать 3.0 \AA , и величине угла, образованного связями донор-водород и водород-акцептор, которая, в свою очередь, не должен превышать 135° .

С помощью МД была исследована стабильность водородных связей между атомами белка (PPhb), атомами белка и атомами молекул воды (PWhb), а также

стабильность водных мостиков (Wbr). С помощью ЭС исследовалась стабильность водородных связей только между атомами белка (PPhb_{ЭС}).

Для белка SOD1 дикого типа и его мутантов строилась таблица стабильностей водородных связей O . В ячейках таблицы содержались средние значения стабильности связей по ансамблям конформаций белка дикого типа (первый столбец) и его мутантов (остальные столбцы):

$$O_i^m = \frac{1}{n} \sum_{j=1}^n o_{ij}^m \quad (2.2)$$

2.4 Статистический анализ

Статистический анализ стабильности водородных связей проводился с помощью библиотек `numpy`, `scipy`, `scikits`, `scikit-learn` (`sklearn`) 0.19.1 [197], `pandas` [198], `seaborn`, `matplotlib` [199] и `matplotlib_venn` для языка программирования Python 2.7.

Мутанты m белка SOD1 были кластеризованы по вектору значений $\{O_i^m\}_i$ с использованием метода `mean shift` [200], реализованного с помощью пакета `sklearn`. Параметр `bandwidth`, который влияет на количество получаемых кластеров, оценивался встроенной в реализацию `sklearn` процедурой. Параметры `bandwidth` для PPhb, PWhb, Wbr и PPhb_{ЭС} имели следующие значения: 1.82, 4.5, 0.74 и 4.86, соответственно.

Везде, где возникало множественное сравнение p -значений применялась поправка Беньямини-Йекутиэли [201] с заранее заданным уровнем ложноположительных результатов.

2.5 Эволюционная консервативность позиций в белке SOD1

Эволюционная консервативность позиций в SOD1 была получена с помощью системы `ConSurf` [202] с настройками по умолчанию. В качестве исходной

структуры была выбрана цепь А в структуре SOD1 (PDBID: 2V0A). Выбран автоматический поиск гомологов в базе данных UNIREF-90 с помощью алгоритма HMMER (1 итерация, порог отсечения E-value задан равным 0.0001). До 150 последовательностей гомологов с подобием от 35 до 95% учитывалось в ConSurf. Множественное выравнивание производилось с помощью метода MAFFT-L-INS-i. Сама консервативность рассчитывалась в ConSurf с помощью байесовского метода.

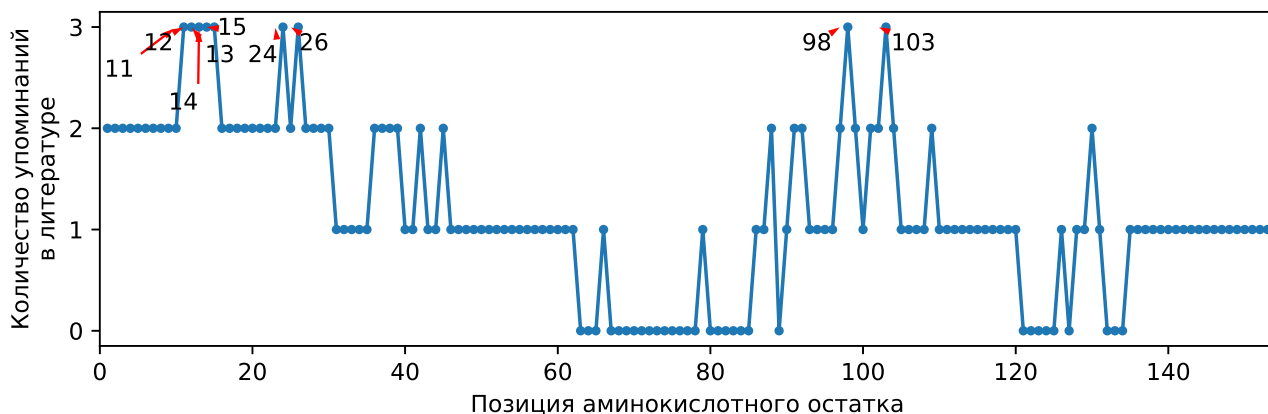
2.6 Анализ структуры белков

Графика подготовлена с помощью пакета matplotlib для Python и UCSF Chimera [203].

2.7 Остатки SOD1, участвующие в нестандартных межатомных взаимодействиях

В разделе 2.7 уже упоминалось, что мутации в SOD1 могут приводить к возникновению контактов с другими молекулами SOD1, отсутствующих в белке дикого типа. В ряде работ указываются конкретные аминокислотные остатки, участвующие в образовании этих нестандартных контактов [131—135]. Все эти остатки далее будут упоминаться как \mathcal{S}_{agr} .

На графике рис. 2.1 продемонстрирована величина, характеризующая количество упоминаний остатка в определённой позиции в качестве участвующего в образовании нестандартных контактов между SOD1 в работах [131—135]. Из рисунка видно, что в указанных публикациях девять остатков в позициях 11–15, 24, 26, 98 и 103 упоминаются трижды. Набор этих остатков далее будет называться набором $\mathcal{S}_{\text{agr}}^9$.



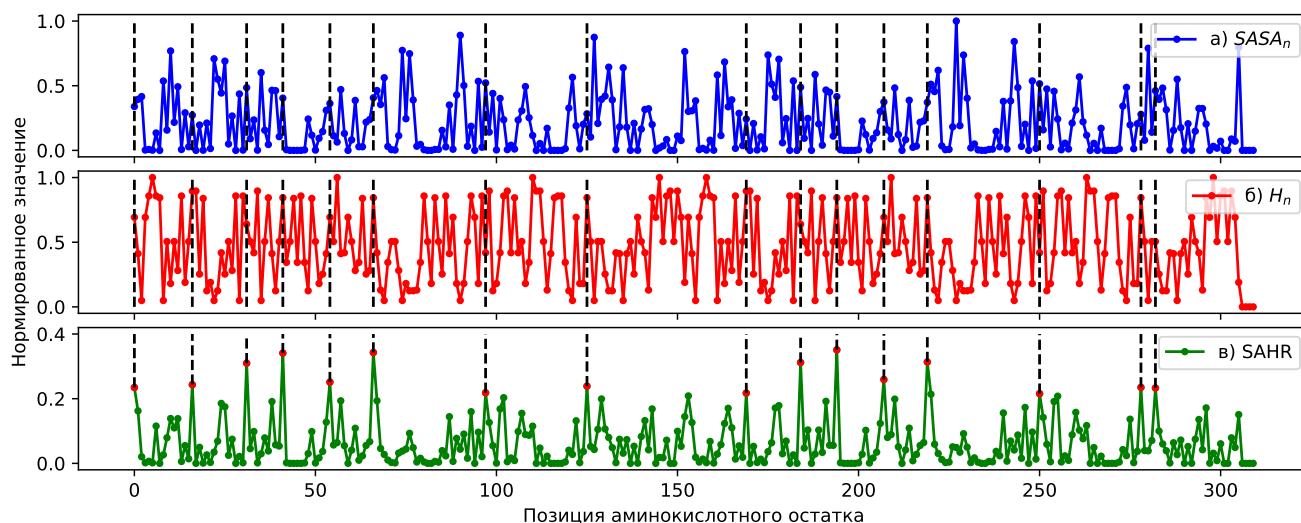
Наиболее часто упоминаемые аминокислотные остатки отмечены стрелками с указанием позиции остатка в последовательности белка.

Рисунок 2.1 — Зависимость между позицией аминокислотного остатка в последовательности SOD1 и количеством упоминаний остатка в этой позиции в качестве участвующего в образовании нестандартных контактов между SOD1 в литературе.

2.8 Гидрофобные аминокислотные остатки белка SOD1, доступные растворителю

Одной из мер, которая может указывать на область в структуре белка, участвующую в образовании белковых агрегатов, является доступность гидрофобных остатков молекулам растворителя. В настоящей работе предложено рассчитывать такую меру следующим образом. На основе кристаллической структуры белка методом Шрейка-Рапли [204] при использовании пакета mdtraj 1.9.0 [205] для Python для каждого аминокислотного остатка находится площадь поверхности, доступной растворителю ($SASA$). Полученные значения нормируются таким образом, чтобы результирующая величина лежала в промежутке от 0 до 1 от минимальной до максимальной для остатков данного белка площади поверхности ($SASA_n$). С помощью шкалы гидрофобности аминокислотных остатков в белке, предложенной в работе [206], рассчитывается гидрофобность каждого остатка (H), которая также нормируется до отрезка $[0; 1]$ (H_n). Мерой доступности гидрофобных остатков молекулам растворителя является величина SAHR, равная произведению нормированных величин $SASA_n$ и H_n :

$$\text{SAHR} = \text{SASA}_n * H_n \quad (2.3)$$



Пунктиром выделены позиции аминокислотных остатков SOD1, для которых значения SAHR лежат в 1-% хвосте распределения SAHR.

Рисунок 2.2 — Графики величин SASA_n , H_n и SAHR для аминокислотных остатков белка SOD1.

Для остатков белка SOD1 (кристаллическая структура PDBID: 2V0A) значения данных трёх величины приведены на графике рис. 2.2. Среди аминокислотных остатков с наибольшими значениями произведения SAHR (взятыми из 10-% хвоста распределения SAHR) находятся остатки: Ала1, Иле17, Трп32, Лей42, Ала55, Лей67, Сер98, Лей126 и Гли130. Другими словами, перечисленные девять остатков, будучи гидрофобными, являются в то же время заметно доступными молекулам растворителя. Данный набор аминокислотных остатков далее будет называться $\mathfrak{S}_{\text{гфоб}}^9$.

2.9 Водородные связи со значительным отклонением стабильности в мутантах SOD1 от таковой в белке дикого типа

Среди всех водородных связей, формирующихся в мутантах SOD1, был проведён поиск таких, чья стабильность наиболее существенно отличалась от стабильности тех же связей в белке дикого типа. Такие водородные связи в дальнейшем упоминаются как деструктивные. Поиск деструктивных водородных связей состоял из следующих шагов. В первую очередь для каждого мутанта m

рассчитывалось отклонение стабильности всех водородных связей в нём от таковой в белке дикого типа (ДТ):

$$D^m = \sqrt{\sum_i (O_i^m - O_i^{\text{ДТ}})^2} \quad (2.4)$$

Далее рассчитывался вклад квадрата разницы между стабильностью связи j в мутанте m и белке дикого типа в D^m :

$$\xi_j^m = \frac{(O_j^m - O_j^{\text{ДТ}})^2}{(D^m)^2} \quad (2.5)$$

После этого исследовалось распределение величин ξ_j^m , на основании которого находились водородные связи j имеющие наибольший по сравнению с остальными вклад квадрата разницы между стабильностью связи j в мутанте m и белке дикого типа в отклонение стабильности всех водородных связей в мутанте m от стабильности связей в белке дикого типа. Отбирались водородные связи с ξ_j^m из правого 1-% хвоста распределения всех вкладов ξ_j^m . Аминокислотные остатки, формирующие деструктивные водородные связи далее будут упоминаться как $\mathfrak{S}_{\text{дест}}$.

2.10 Наблюдаемые значения дожития пациентов с мутациями в SOD1

2.10.1 Отбор мутаций в SOD1 для моделирования и анализа

В работе [7] опубликованы данные о 72 патогенных мутациях в SOD1, для каждой из которых представлена информация о средней величине дожития пациентов, стандартном отклонении её от среднего, количестве известных пациентов с данной мутацией, а также среднем возрасте диагностики БАС у таких пациентов.

Поскольку расчёт конформационных свойств мутантов требует длительных вычислений, а также некоторые мутации из числа приведённых в работе [7] имели малое количество данных по пациентам, то для построения регрессионных моделей из имеющихся 72 мутаций было выбрано 33 мутации (см. таблицу 6 в приложении Б). Средняя величина дожития для этих пациентов оказалась

равной 5.0 лет со стандартным отклонением 5.1 года. Выбор мутаций осуществлялся с помощью критериев, учитывающих количество пациентов, являющихся носителями конкретной мутации, а также распределение мутаций по различным позициям в последовательности. Кроме этого, во внимание принималось распределение величины дожития пациентов.

Позиции в аминокислотной последовательности SOD1, в которых имелись выбранные мутации, далее будут упоминаться как $\mathfrak{S}_{\text{мут}}$.

2.10.2 Обучающие выборки для построения регрессионных моделей

В работе [7] было показано, что регрессионные модели, учитывающие количество пациентов для каждой мутации, обладали большей точностью предсказания дожития по сравнению с моделями, не учитывающими эту информацию. Коэффициенты корреляции между предсказанной и наблюдаемой величиной дожития с учётом и без учёта количества пациентов составили $R = 0.83$ ($p < 0.001$) и 0.47 ($p = 0.01$), соответственно. В настоящей работе для построения регрессионных моделей использовалось два подхода. Первый подход, в котором использовались оригинальные данные о дожитии пациентов с 33 мутациями, не учитывал количество пациентов. Во втором подходе, использующем сгенерированную обучающую выборку (566 мутаций) на основе оригинальных данных по 33 мутациям, было учтено количество пациентов.

Генерация обучающей выборки осуществлялась следующим образом. В таблице (см. рис. A.1–A.5 в приложении A), взятой из работы [7], для каждой мутации приведены данные о средней величине дожития (μ), количестве пациентов с данной мутацией (n) и стандартном отклонении величины дожития от среднего (σ), которые авторы собрали из разных публикаций. Как можно видеть из таблицы, для одной мутации, как правило, было приведено несколько различных значений μ , σ и n . Например, для мутации Ала4Вал указывается девять различных значений μ , σ и n . В таких случаях строилось нормальное распределение с параметрами $\bar{\mu} = \frac{1}{K} \times \sum_{i=1}^K \mu_i$, $\bar{\sigma} = \frac{1}{K} \times \sqrt{\sum_{i=1}^K \sigma_i^2}$, где K – количество значений μ , σ и n для заданной мутации (см. рис. A.1–A.5 в приложении A).

Глава 3. Анализ результатов молекулярного моделирования динамики белка SOD1 дикого типа и его мутантов

Используя существующие методы и подходы, а также вновь разработанные в рамках данной работы и описанные в главе 2, в настоящей главе представлены результаты моделирования структуры и динамики мутантных форм белка SOD1. Исследованы структурные характеристики этого белка.

3.1 Моделирование методом МД

Стабильность водородных связей, выявленных в ходе анализа траектории МД для мутантов SOD1 и белка дикого типа рассчитывалась, как среднее по времени моделирования МД по формуле 2.2. Рассматривались водородные связи между атомами белка, атомами белка и атомами молекул воды, а также водные мостики. Соответствующие группы водородных связей далее будут обозначены, как PPhb, PWhb и Wbr. Для анализа того, насколько оценка стабильности водородных связей устойчива в зависимости от периода моделирования, были построены графики изменения стабильности от длины траектории МД (см. рис. 3.1–3.3).

Для построения данных графиков область значений стабильности всех водородных связей $[0; Q]$ (Q – максимальное значение стабильности) была разбита на 30 равных интервалов. Для каждого интервала был определён набор водородных связей, которые наблюдались в траектории с 10 по 14 нс. Каждая линия на приведённых рисунках отражает усреднённую стабильность водородных связей в соответствующем наборе водородных связей из этих 30 наборов.

Как видно из рис. 3.1–3.3, динамики стабильностей связей при увеличении продолжительности моделирования выходят на плато. Таким образом, выбранная длительность моделирования в 50 нс может считаться достаточной для оценки стабильности связей. Всего обнаружено 5266 водородных связей внутри белка, 1637 водородных связей с молекулами воды и 2575 связей, составляющих водные мостики.

На рис. 3.4 представлен график RMSD для шести мутантов SOD1 и белка дикого типа. Для каждого из этих белков было смоделировано пять траекторий

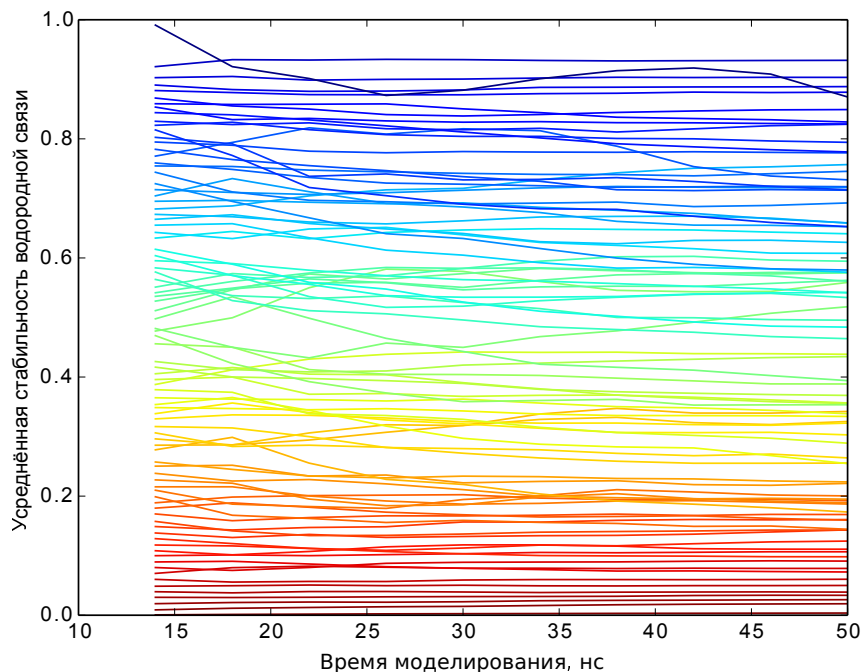


Рисунок 3.1 — Стабильность водородных связей между атомами белка.

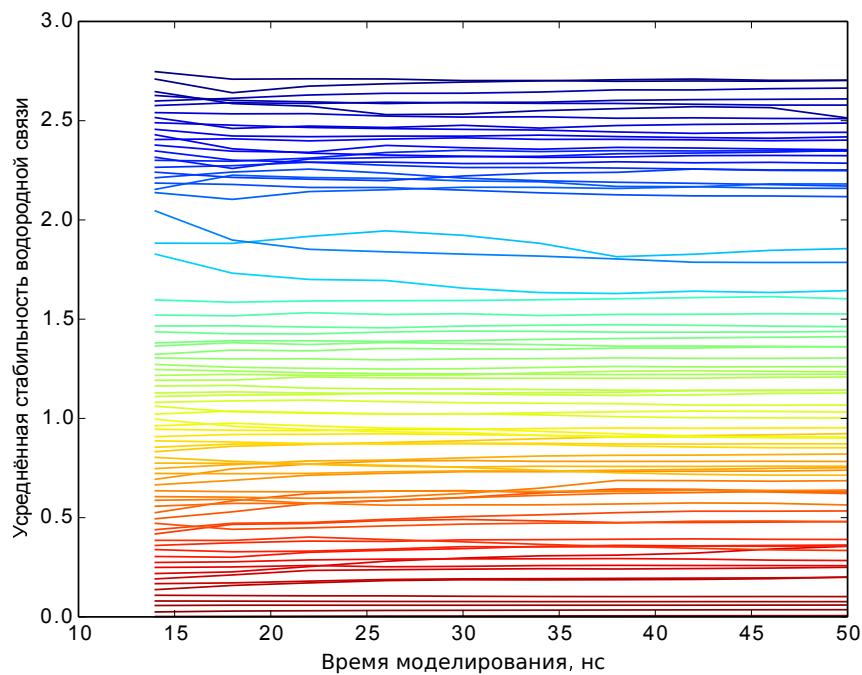


Рисунок 3.2 — Стабильность водородных связей между атомами белка и атомами молекул воды.

МД с длительностью 40 нс (рассматривался интервал моделирования от 10 до 50 нс). На оси абсцисс цифрами обозначены номера этих отдельных траекторий.

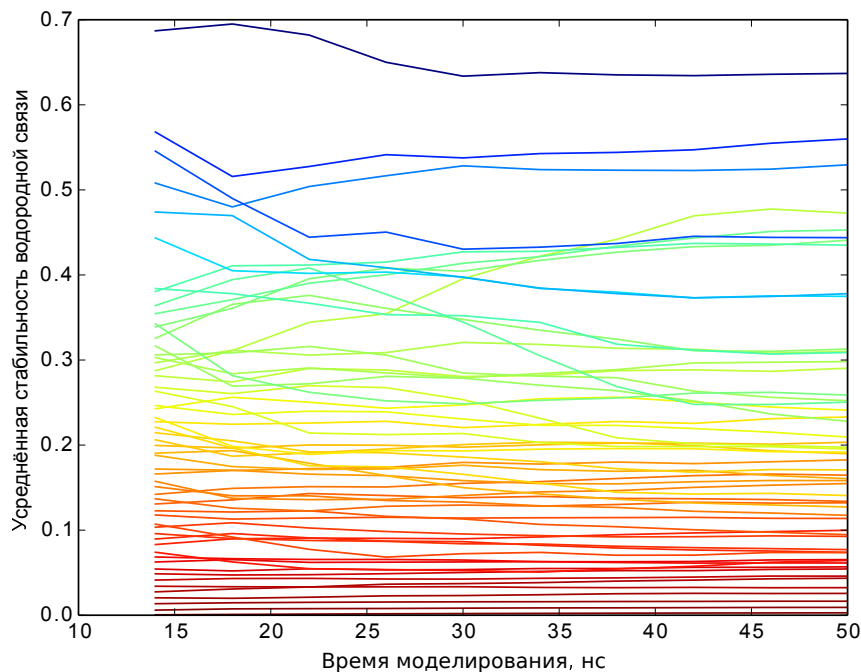
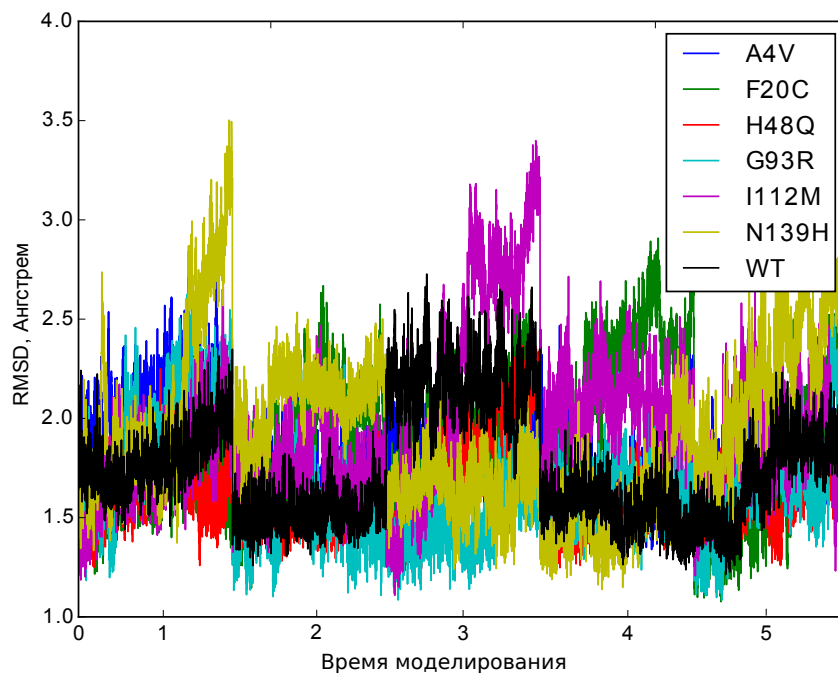


Рисунок 3.3 — Стабильность водных мостиков.



На оси абсцисс цифрами обозначены номера отдельных траекторий МД.
 Рисунок 3.4 — Зависимость RMSD пространственной структуры мутантов SOD1 и белка дикого типа в ходе моделирования МД.

Как видно из рисунка, значение RMSD в процессе моделирования для белка дикого типа не превосходит 3 \AA , а для мутанта Асн139Гис эта величина достигает

значения 3.5 Å, что может свидетельствовать в пользу меньшей структурной стабильности последнего по сравнению с белком дикого типа.

3.2 Моделирование ЭС

В данной работе в качестве одного из подходов оценки стабильности водородных связей используется анализ ансамблей пространственных структур, полученных с помощью моделирования ЭС. Такое моделирование является важной составляющей в анализе стабильностей водородных связей поскольку оно позволяет учесть крупномасштабные изменения в структуре белка. Основным результатом моделирования ЭС являются моды гармонических колебаний атомов.

С использованием FoldX проводилась оптимизация пространственных структур белка дикого типа и моделирование 35 выбранных мутантных SOD1. В качестве исходной структуры использовалась структура белка дикого типа (PDBID: 2V0A). Далее с использованием ЭС было проведено моделирование крупномасштабной динамики мутантов и белка дикого типа. Для каждой эластичной модели был получен ансамбль из 11 конформаций для каждой из 5 нетривиальных мод колебаний с наивысшим показателем коллективности. Таким образом, для белка дикого типа и каждого из изучаемых мутантов были рассчитаны 55 конформаций, представляющих крупномасштабные флуктуации структуры этих молекул.

Для того, чтобы оценить, насколько мутации могли повлиять на кооперативные движения атомов, были рассчитаны коэффициенты корреляции между собственными векторами для одних и тех же мод разных мутантов и белка дикого типа. Оказалось, что все коэффициенты корреляции находятся в диапазоне от 0.97 до 0.99. Это говорит о том, что конформационная динамика при моделировании ЭС у мутантов сходна с динамикой белка дикого типа. Таким образом, оценка изменения стабильности водородных связей у мутантов и белка дикого типа на основе анализа ансамбля пространственных структур, полученных из моделирования ЭС, может считаться корректной.

С помощью утилиты crrtraj производился поиск водородных связей в каждой структуре, соответствующей анализируемым конформациям, для мутанта и белка дикого типа. Для каждой обнаруженной водородной связи по формуле 2.2

рассчитывалась её стабильность в имеющемся ансамбле конформаций, как среднее по этому ансамблю конформаций. Всего обнаружено 934 водородных связи, которые образовывались в белке дикого типа и изучаемых мутантах SOD1.

Следует подчеркнуть, что в ходе моделирования МД было выявлено 5266 водородных связей, что больше, чем количество водородных связей, обнаруженных на основе анализа пространственных структур, полученных с помощью ЭС. Такая заметная разница может быть связана с разницей в количестве анализируемых конформаций. В случае МД для анализа использовалось 200 структур, а в случае ЭС – 55. При этом, часть водородных связей, выявленных с помощью МД, была образована в результате небольших локальных флуктуаций атомов (см. рис. 3.7). Однако, не смотря на различия в количестве обнаруженных водородных связей, значительное их количество было образовано между теми же парами остатков, как в случае МД, так и ЭС.

3.3 Анализ водородных связей, формирующихся в белке в ходе молекулярного моделирования

3.3.1 Водородные связи в SOD1, обнаруженные с помощью моделирования МД и ЭС

Анализ водородных связей в пространственной структуре мутантов SOD1, найденных в результате моделирования ЭС и МД, показал, что значительная доля водородных связей, найденных в рамках ЭС была обнаружена также в моделях МД.

На рис. 3.5 изображена диаграмма Венна, демонстрирующая пересечения трёх групп (PPhb, Wbr и PPhb_{ЭС}) водородных связей, образованных между аминокислотными остатками при рассмотрении структур всех мутантов SOD1 и белка дикого типа. Если между заданной парой остатков было обнаружено несколько водородных связей, образованных разными атомами этих остатков, то все эти связи считались как одна связь. Кроме этого, если одна и та же водородная связь была обнаружена в разных мутантах, она также учитывалась единожды.

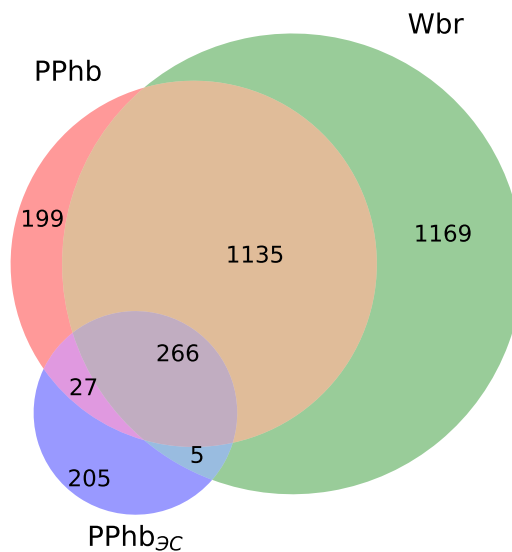
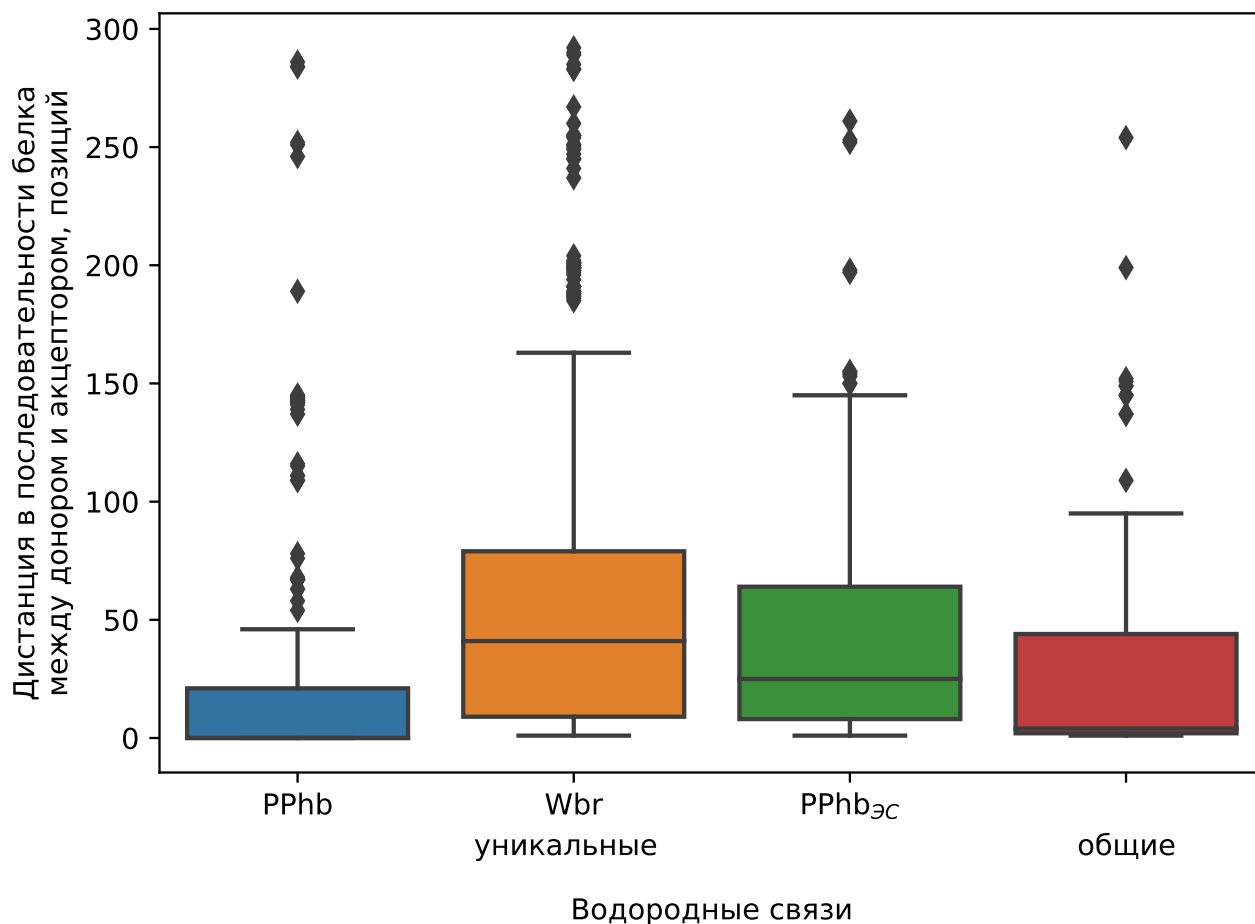


Рисунок 3.5 — Диаграмма Венна для пересечения множеств контактов через водородные связи PPhb, Wbr и PPhbЭС между аминокислотными остатками донора и акцептора.

Из рисунка можно видеть, что наибольшее число связей (2575) являлось водными мостиками, а наименьшее количество (503) соответствовало внутрибелковым водородным связям, полученным с помощью ЭС. Среди всех трёх групп водородных связей присутствует 266 общих связей, то есть образованных между одной и той же парой остатков, что составляет около 9 % от общего количества связей. Оказалось, что в каждой группе также присутствуют уникальные связи, то есть связи между неповторяющимися парами остатков. Наибольшее число таких связей присутствует в группе Wbr, что можно ожидать, поскольку количество водных мостиков превышает количество связей в остальных группах.

Интересным оказался факт, что число уникальных связей в группе PPhbЭС превышает количество таковых в группе PPhb, хотя размер последней больше размера первой в три раза.

Анализ расположения водородных связей в пространственных структурах анализируемых белков показал, что уникальные водородные связи PPhb чаще образуются между соседними в последовательности аминокислотными остатками (медиана дистанции равнялась 0). В то же время, уникальные для Wbr связи образовывались между остатками в позициях, заметно удалённых в первичной структуре SOD1 (медиана дистанции равнялась 41). Для уникальных водородных связей из группы PPhbЭС медиана расстояний между позициями пары остатков



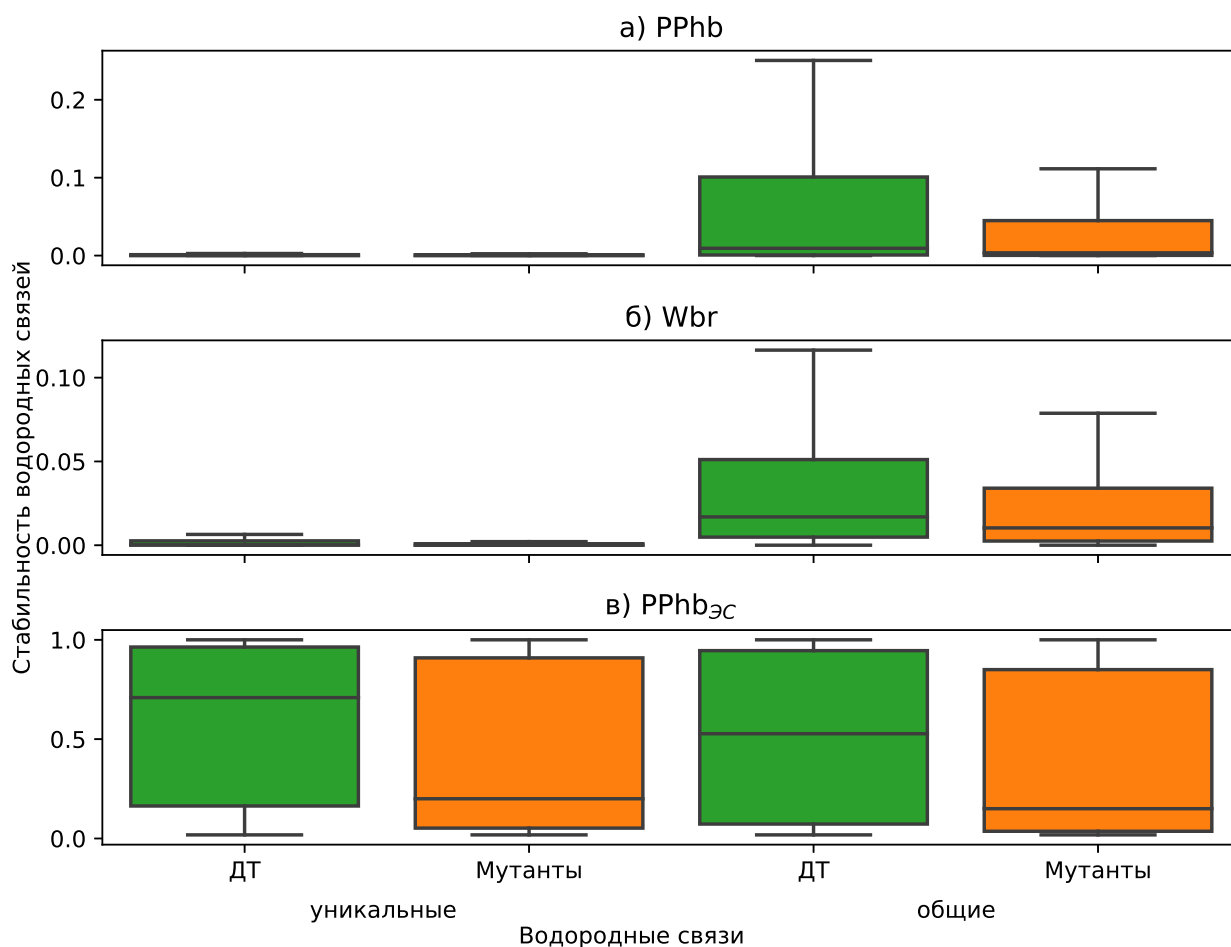
Первые три графика соответствуют уникальным для PPhb, Wbr и PPhbЭС контактам. Четвёртый график соответствует контактам, которые обнаружены среди водородных связей каждого типа.

Рисунок 3.6 — Средняя дистанция в белковой последовательности между позициями аминокислотных остатков-доноров и акцепторов из контактов через водородные связи PPhb, Wbr и PPhbЭС.

оказалась равной 25. То, что медиана расстояний между парами остатков для водных мостиков оказалась максимальной хорошо согласуется с природой этих связей, поскольку они образуются с участием остатков, располагающихся на поверхности белка. Поверхностные остатки, сами по себе, в большом количестве случаев представлены удалёнными в первичной структуре.

Различие в значении медиан, наблюдаемое для групп связей PPhb и PPhbЭС, можно объяснить особенностями структур белков, полученных в результате МД и ЭС. Как упоминалось в разделе 3.2 выше, ансамбль структур, полученных из МД, был больше чем таковой из ЭС. При этом, многие конформации имели относительно небольшие отклонения друг от друга в связи с тем, что они были сближены

в траектории МД. В то же время, структуры, попавшие в ансамбль из ЭС, соответствовали гораздо более сильным конформационным изменениям по сравнению с МД. Однако следует ещё раз подчеркнуть, что большое количество общих связей для рассмотренных трёх групп указывает на наличие общей основы в этих конформациях. Учёт уникальных водородных связей позволяет получить более полную картину о конформационных состояниях анализируемых белков при интеграции результатов моделирования МД и ЭС.



Левая часть каждой диаграммы построена для уникальных для каждой группы (PPhb, Wbr и PPhbЭС) водородных связей, правая – для общих для всех групп связей. Обозначение «ДТ» – белок SOD1 дикого типа, «Мутанты» – патогенные мутанты SOD1.

Рисунок 3.7 — Стабильность водородных связей PPhb, Wbr и PPhbЭС, уникальных для каждой группы связей и общих для всех групп связей.

Было интересно сравнить между собой стабильности общих и уникальных водородных связей, образующихся в белке SOD1 дикого типа и его мутантах (см. рис. 3.7). Для этого также была рассчитана значимость отличия стабильности

водородных связей из различных групп, рассчитанная в соответствии с непараметрическим критерием Манна-Уитни-Вилкоксона [207; 208] (см. таблицу 1).

Таблица 1 — Значимость отличия стабильности водородных связей в различных группах по критерию Манна-Уитни-Вилкоксона.

Тест	р-значение (группа 1)	р-значение (группа 2)
Водородные связи PPhb		
Мутанты < ДТ	^a 0.11	^б 8.82×10^{-10}
Уникальные < общие	^в 6×10^{-25}	^г 2.3×10^{-43}
Водородные связи Wbr		
Мутанты < ДТ	^a 2.6×10^{-5}	^б 9.2×10^{-4}
Уникальные < общие	^в 5×10^{-47}	^г 1.1×10^{-71}
Водородные связи PPhb _{ЭС}		
Мутанты < ДТ	^a 1.2×10^{-5}	^б 1.4×10^{-5}
Уникальные > общие	^в 0.05	^г 0.034

Примечание: ДТ – дикий тип; ^a – уникальные; ^б – общие; ^в – ДТ; ^г – мутанты.

Из рисунка видно, что уникальные водородные связи из групп PPhb и Wbr имеют заметно более низкую стабильность, чем общие для всех групп водородные связи ($p < 0.01$), а уникальные для PPhb_{ЭС} водородные связи напротив – имеют более высокую стабильность ($p \leq 0.05$). Также рисунок демонстрирует, что уникальные и общие водородные связи в мутантах, как правило, менее стабильны, чем водородные связи в белке дикого типа ($p < 0.01$). Исключение – в группе PPhb стабильность уникальных водородных связей в мутантах отличается незначимо от таковой для белка дикого типа ($p = 0.11$).

3.3.2 Водородные связи со значительным отклонением стабильности

В соответствии с методом, описанным в разделе 2.9 находились водородные связи со значительным отклонением стабильности в мутантах SOD1 по сравнению с белком дикого типа. Для этого, по формуле 2.5 для каждой водородной связи j рассчитывался вклад ξ_j^m квадрата разницы между стабильностью этой связи в мутанте m и белке дикого типа в отклонение стабильности всех водородных связей в мутанте m от стабильности связей в белке дикого типа.

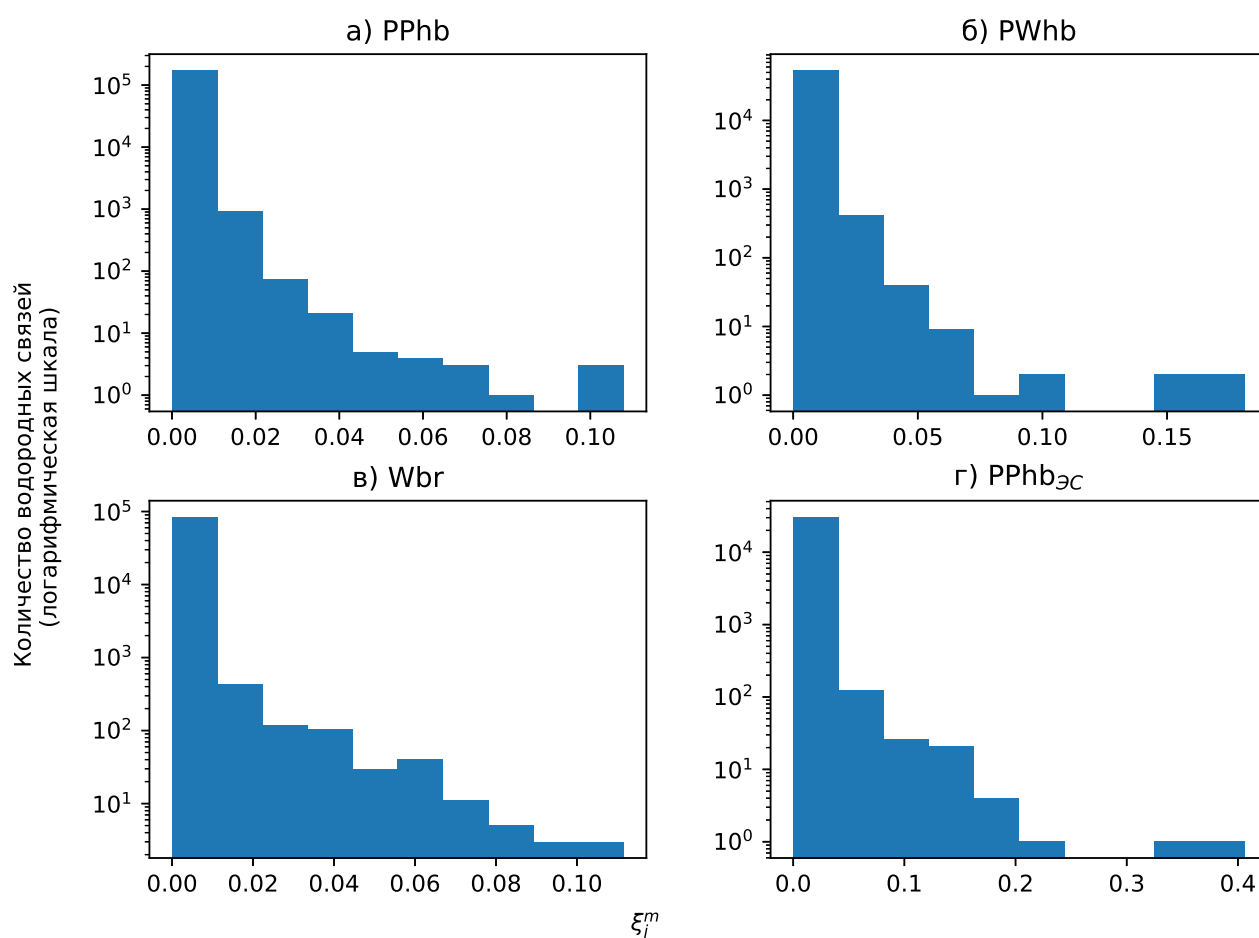


Рисунок 3.8 — Распределения вкладов ξ_j^m для водородных связей PPhb (а), PWhb (б), Wbr (в) и PPhb_{ЭC} (г).

Оказалось, что наибольший вклад имеет лишь небольшое количество водородных связей (см. рис. 3.8). При этом максимальный вклад для каждого типа связей различался: 0.1, 0.18, 0.11 и 0.4, соответственно для PPhb, PWhb, Wbr и PPhb_{ЭC}.

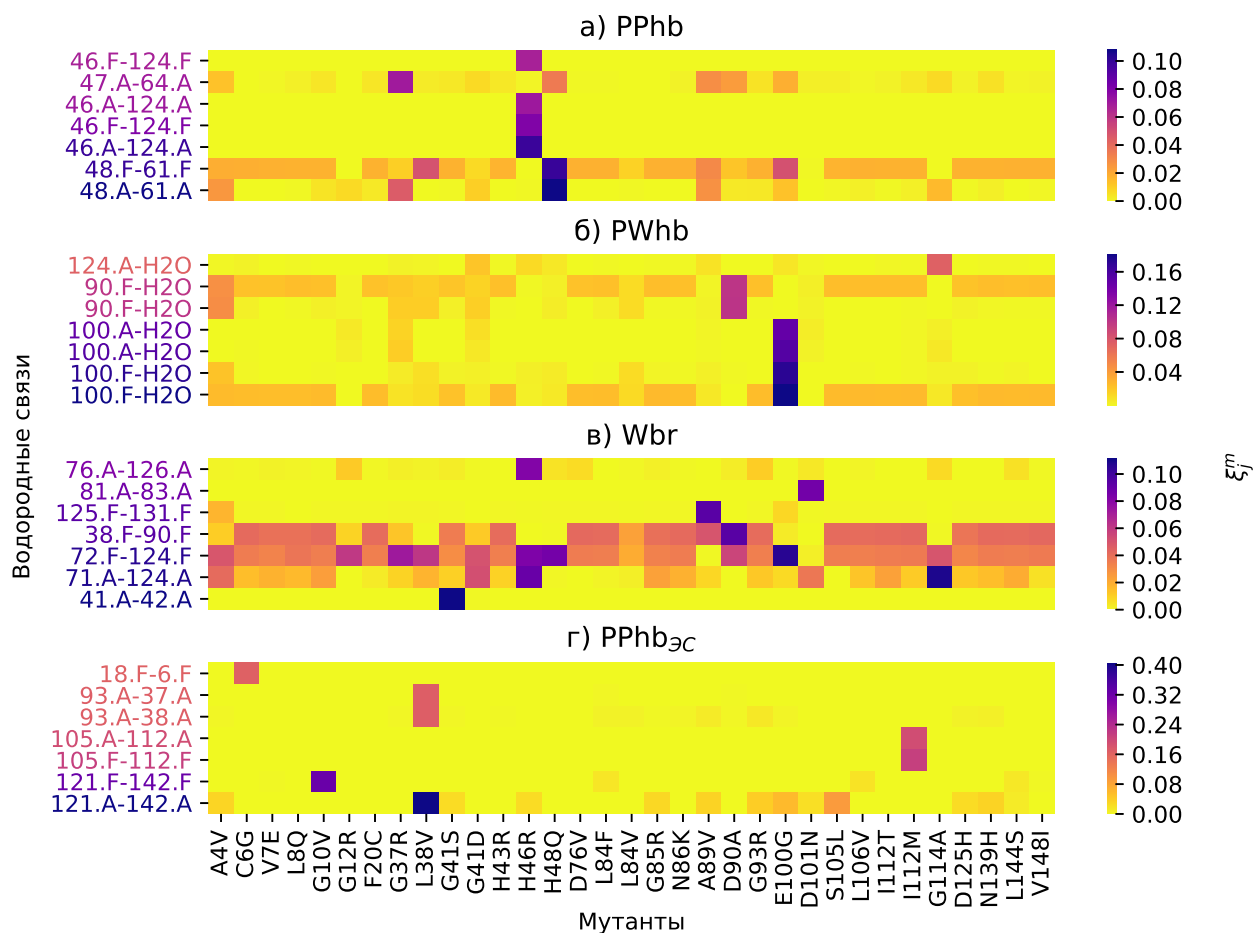


Рисунок 3.9 — Семь водородных связей PPhb (а), PWhb (б), Wbr (в) и PPhbЭС (г) с наивысшими вкладами ξ_j^m .

Среди водородных связей с наибольшими вкладами ξ_j^m (из 1 % правого хвоста распределения ξ_j^m) обнаружались водородные связи, сформированные между остатками в позициях: 6, 18, 37–38, 41–42, 46–48, 61, 64, 71–72, 76, 81, 83, 90, 93, 100, 105, 112, 121, 124–126, 131 и 142 (см. рис. 3.9). Данный набор позиций будет обозначаться как $\mathfrak{S}_{\text{дест}}$. Следует заметить, что остатки в позициях 46 и 124 чаще всего образуют водородные связи с наивысшим значением ξ_j^m . Наиболее сильное отличие стабильности водородных связей PPhb с участием остатков в позициях 46 и 48 от стабильности этих связей в белке дикого типа наблюдается для мутантов Гис46Арг и Гис48Глн. В противоположность этому, например, стабильность водных мостиков Wbr с участием остатков в позициях 38, 71–72, 90 и 124 заметно отличается от стабильности этих связей практически во всех изучаемых мутантах. Последний факт может указывать на важную роль данных остатков SOD1 в патогенезе БАС.

3.3.3 Кластеризация мутантов SOD1 по стабильности водородных связей

С помощью метода mean shift проведена кластеризация исследуемых мутантов SOD1 по стабильности водородных связей PPhb, PWhb, Wbr и PPhb_{ЭС}, образованных в этих мутантах по данным моделирования МД и ЭС.

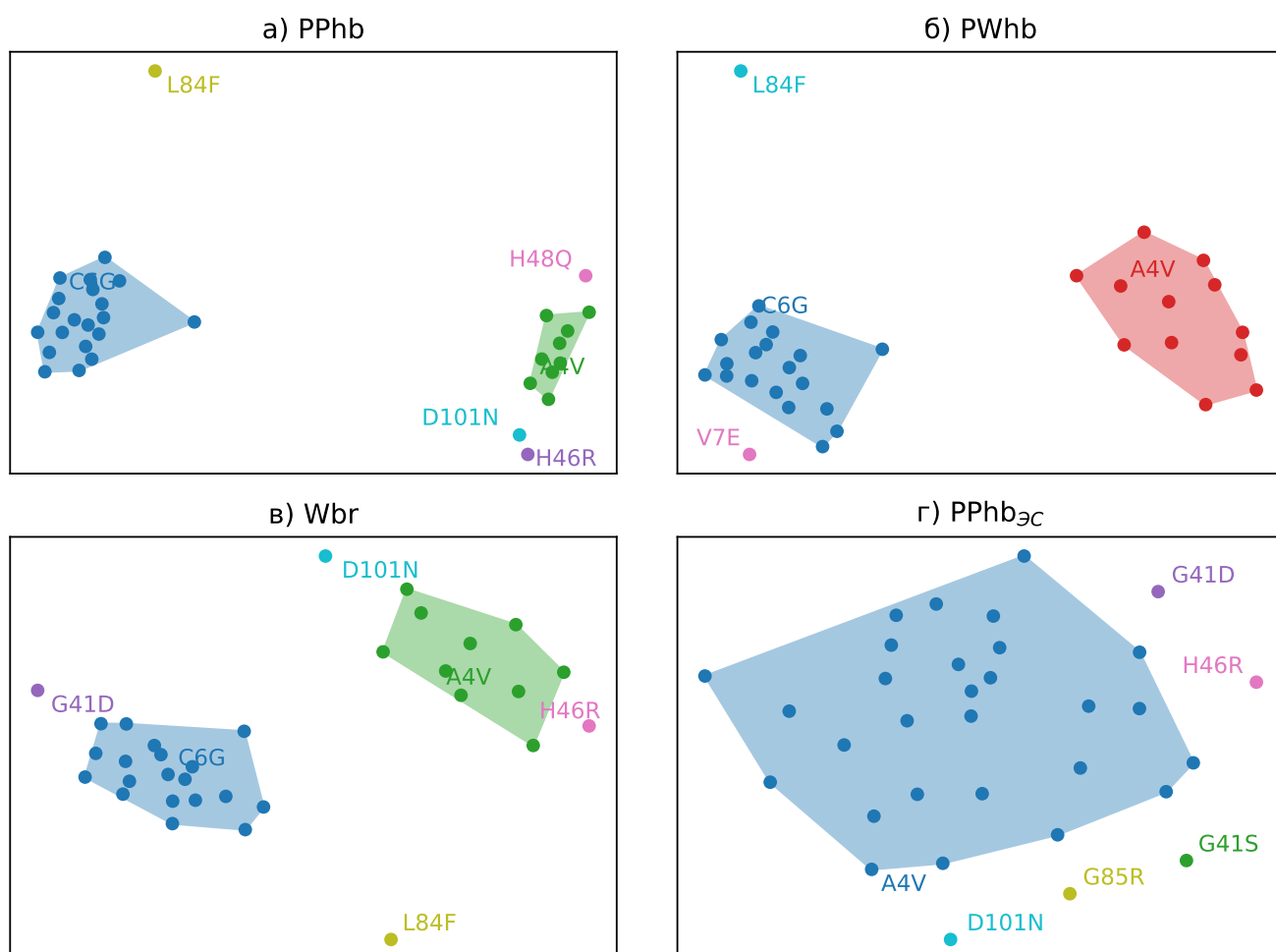


Рисунок 3.10 — Кластеризация мутантов SOD1 по стабильности всех обнаруженных водородных связей PPhb (а), PWhb (б), Wbr (в) и PPhb_{ЭС} (г).

В результате кластеризации мутантов SOD1 было обнаружено, что наибольшая часть мутантов попадает в один (PPhb_{ЭС}) или два (PPhb, PWhb, Wbr) крупнейших кластера (см. рис. 3.10). Тем не менее, образуются также одиночные кластеры – состоящие только из одного мутанта SOD1.

Мутанты SOD1, попавшие в крупнейшие кластеры

Для мутантов из крупнейших кластеров проанализированы позиции аминокислотных остатков, которые формируют водородные связи с наивысшими вкладами ξ_j^m (10 водородных связей с наивысшей медианой вкладов ξ_j^m по мутантам из одного кластера, см. также раздел 3.3.2). С помощью гипергеометрического распределения найдено значимое пересечение каждого из наборов таких позиций аминокислотных остатков с остатками, формирующими элементы вторичной структуры белка SOD1 (см. раздел 4.1.2). Оказалось, что аминокислотные остатки, формирующие водородные связи с наивысшими вкладами ξ_j^m в мутантах из крупнейших кластеров также участвуют в формировании следующих элементов вторичной структуры:

1. для водородных связей PPhb:
 - C9, E4, E7 (кластер, содержащий, в частности мутант Цис6Гли);
 - C6, E4 (кластер, содержащий, в частности мутант Ала4Вал);
2. для PWhb:
 - C2, C4, E6, H2, H3 (кластер с Цис6Гли);
 - C7, C10, H3 (кластер с Ала4Вал);
3. для Wbr:
 - C1, C10, E6 (кластер с Цис6Гли);
 - C10, C11 (кластер с Ала4Вал);
4. для PPhb_{ЭС}:
 - C7 (кластер с Ала4Вал);
 - C4, C10, C11 (кластер из Гли41Сер);
 - E4, H3 (кластер из Гли41Асп);
 - C4, C10, C11 (кластер из Гис46Арг);
 - C4, C7, C10, C12 (кластер из Гли85Арг);
 - C6, C7, C10, C11 (кластер из Асп101Асн).

Таким образом, видно, что в мутантах из различных кластеров водородные связи с наибольшим отличием стабильности от белка дикого типа формируются с участием различных элементов вторичной структуры. Многие водородные связи образованы аминокислотными остатками, располагающимися как в одном и том же элементе вторичной структуры, так и располагающимися в разных, удалённых друг от друга, элементах.

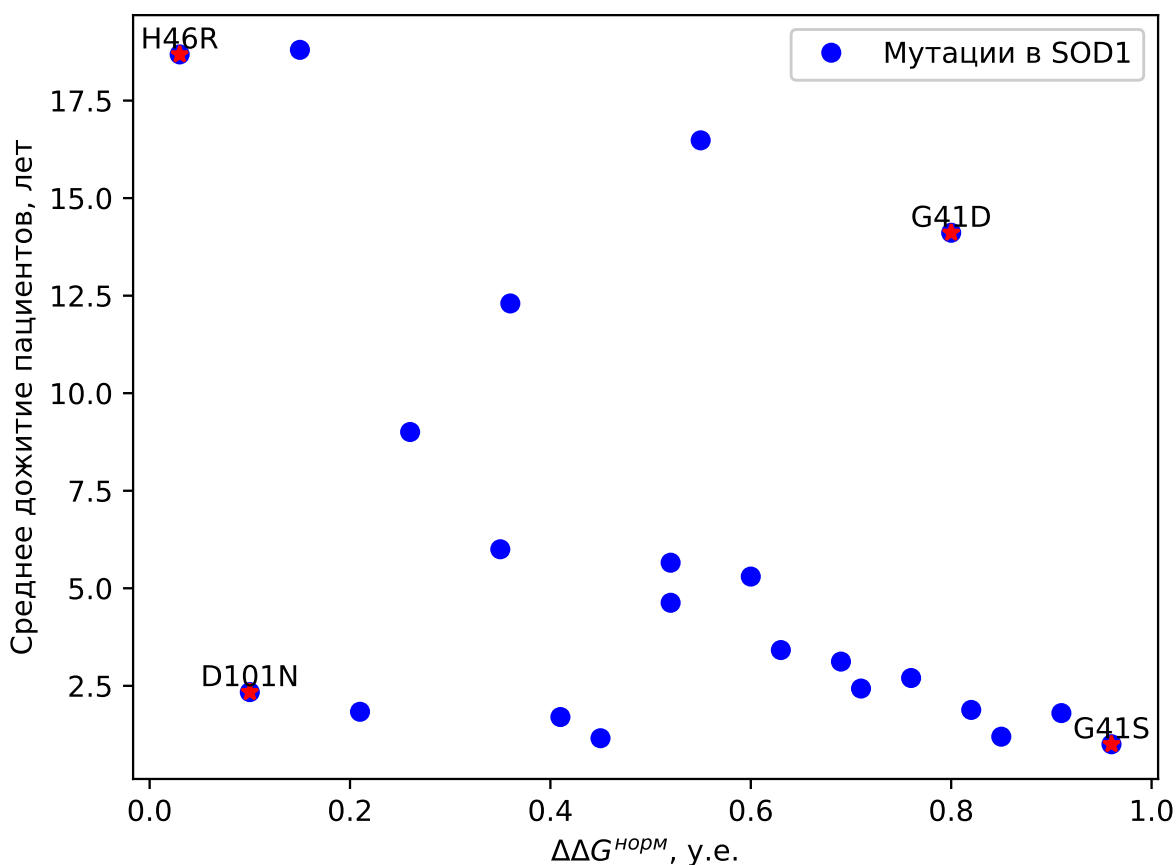
Анализ кластеризации показывает, что дестабилизация мутантных SOD1 может происходить в разных областях их структуры. Оказалось интересным, что в мутантах чаще всего дестабилизированными были следующие элементы вторичной структуры: H3, E4, C4, C6, C7, C10 и C11. Известно, что β -тяж E4 формирует часть сайта связывания иона меди, а петля C6 является цинк-связывающей. В то же время α -спираль H3, а также петли C10 и C11 составляют, собственно, электростатическую петлю. Петли C4 и C7 являются сближенными в пространственной структуре белка и, как считается, могут участвовать в образовании нестандартных взаимодействий между мутантными SOD1.

Мутанты SOD1, составляющие одиночные кластеры

Отдельно проанализированы мутанты SOD1, попавшие в одиночные кластеры. Среди них:

1. Гис46Арг, Гис48Гли, Лей84Фен и Асп101Асн (PPhb);
2. Вал7Глу и Лей84Фен (PWhb);
3. Гли41Асп, Гис46Арг, Лей84Фен и Асп101Асн (Wbr);
4. Гли41Асп, Гли41Сер, Гис46Арг, Гли85Арг и Асп101Асн (PPhb_{ЭС}).

Из рисунка 3.11 видно, что одиночные кластеры, чаще всего, составляли мутанты Гис46Арг, Лей84Фен и Асп101Асн. По-видимому, каждый из мутантов, попавших в одиночные кластеры при кластеризации мутантов SOD1 в МД и ЭС, имеет свою, не схожую с другими мутантами динамику. Известно, что, например, Асп101Асн имеет близкую к белку дикого типа термостабильность [10], однако пациенты, являющиеся носителями Асп101Асн, живут, в среднем, лишь чуть более двух лет после постановки диагноза [10]. Мутант Гли41Сер – напротив – имеет существенно сниженную термостабильность по сравнению с белком дикого типа, и пациенты с такой мутацией живут не более года [10; 209]. Патогенность мутации Гли41Асп значительно отличается от таковой для Гли41Сер. Пациенты с Гли41Асп, в среднем, живут около 14 лет [7]. При этом стабильность данного мутанта значительно снижена [10]. Мутант Гис46Арг отличается от других тем, что имеет незначительно сниженную стабильность по сравнению с белком дикого типа [10]. Пациенты с такой мутацией живут в среднем более 17 лет [7].



Подписями обозначены мутации, которые попали в отдельные кластеры при кластеризации мутантов по данным МД и ЭС. Значения термостабильности и дожития пациентов взяты из литературы [10].

Рисунок 3.11 — Зависимость между нормализованной термостабильностью мутантов SOD1 и их клиническими проявлениями, выраженными в форме дожития пациентов с БАС с соответствующими мутациями.

Таким образом, каждый из мутантов Гли41Асп, Гли41Сер, Гис46Арг и Асп101Асн попавших в отдельный кластер, имеет характерную комбинацию уровня снижения термостабильности в сравнении с белком дикого типа и дожития пациентов, являющихся носителями соответствующей мутации (см. рис. 3.11).

3.4 Заключение по главе 3

Методами МД и ЭС проведено моделирование структурных флуктуаций 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 PPhb_{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.

Кроме того выяснилось, что водородные связи в мутантах SOD1, как правило, менее стабильны по сравнению с таковыми в белке дикого типа (см. рис. 3.7). Это хорошо согласуется с данными о том, что мутантные SOD1, как правило, имеют сниженную стабильность и, в результате этого, приобретают способность образовывать агрегаты [10].

Среди водородных связей со значительным отклонением стабильности в мутантах SOD1 от таковой в белке дикого типа чаще других встречались остатки в позициях 46 и 124, и их стабильность в мутантах Гис46Арг и Гис48Глн наиболее сильно отличалась от стабильности этих связей в белке дикого типа. С другой стороны, стабильность водных мостиков, в которых участвовали остатки в позициях 38, 71–72, 90 и 124, заметно отличалась 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 [132; 210; 211]. Это 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 или два крупнейших кластера. Остальные образуют одиночные кластеры – состоящие только из одного мутанта SOD1. Оказалось, что в мутантах из крупнейших кластеров чаще всего с изменённой по сравнению с белком дикого типа стабильностью были водородные связи, образованные остатками из сайта связывания иона меди, цинк-связывающей и электростатической петель, а также остатками из областей структуры, участвующих в агрегации мутантных SOD1. Обнаружено, что каждый

из мутантов Гли41Асп, Гли41Сер, Гис46Арг и Асп101Асн, попавших в одиночный кластер, отличаются друг от друга и остальных мутантов, поскольку имеют характерную комбинацию уровня снижения термостабильности в сравнении с белком дикого типа и величины дожития пациентов, являющихся носителями соответствующей мутации.

Глава 4. Построение и исследование регрессионных моделей, связывающих дожитие пациентов с болезнью БАС со структурными характеристиками мутантных форм SOD1

В настоящей главе на основе этих конформационных свойств построены регрессионные модели, связывающие дожитие пациентов с БАС с мутациями в SOD1 со структурными характеристиками мутантных форм белка. Исследование основывалось на результатах, полученных в главе 3, в которой с помощью МД и ЭС были проанализированы конформационные свойства мутантов белка SOD1.

4.1 Регрессионные модели для предсказания дожития пациентов

4.1.1 Построение однофакторных регрессионных моделей

Оптимальная сумма абсолютных разностей стабильностей водородных связей

Для предсказания дожития пациентов, несущих определённую мутацию в SOD1, строилась таблица \bar{O} абсолютных различий между стабильностью водородных связей в мутантах и белке дикого типа (ДТ):

$$\bar{O}_i^m = \left| O_i^m - O_i^{\text{ДТ}} \right| \quad (4.1)$$

Далее производилась сортировка строк таблицы \bar{O} по убыванию абсолютного значения коэффициента корреляции Пирсона между \bar{O}_i^m и дожитием пациентов Θ^m (или Θ_{exp}^m), полученным из литературы (см. раздел 2.10). Рассчитывалась мера патогенности $\Omega_{\mathcal{K}}^m$ мутантов ($\mathcal{K} \subseteq [1, N]$), как разница между суммой стабильности водородных связей, положительно коррелирующих с дожитием пациентов и аналогичной суммой для отрицательно коррелирующих связей:

$$\Omega_{\mathcal{K}}^m = \sum_{k \in \mathcal{K}, R_k > 0} \bar{O}_k^m - \sum_{k \in \mathcal{K}, R_k < 0} \bar{O}_k^m \quad (4.2)$$

Здесь R_k – коэффициент корреляции строки $\{\bar{O}_k^m\}$ с $\{\Theta^m\}$ (или $\{\Theta_{exp}^m\}$), \mathcal{K} – подмножество номеров связей, удовлетворяющих условию: $|R_k| > R_{th}$. Для определения R_{th} применялась итерационная процедура, которая максимизировала абсолютное значение коэффициента корреляции между $\Omega_{\mathcal{K}}^m$ и Θ^m (см. ниже).

Мера патогенности водородной связи ($\Omega_{\mathcal{K}}^m$) может рассматриваться как величина, учитывающая абсолютную разницу (\bar{O}_i^m) между стабильностями водородных связей $\{i\}$, образованными в мутанте m белка SOD1 и стабильностями тех же водородных связей в белке дикого типа. Другими словами, чем больше отличается стабильность некоторой водородной связи в мутанте от таковой в белке дикого типа, тем больше значение $\Omega_{\mathcal{K}}^m$. Кроме того, если стабильность водородной связи i в мутантах положительно коррелирует с дожитием пациентов с соответствующими мутациями, то абсолютная разность \bar{O}_i^m включается в $\Omega_{\mathcal{K}}^m$ со положительным знаком. Таким же образом, если наблюдается отрицательная корреляция между стабильностью i и дожитием, то данная разность входит с отрицательным знаком. То есть мера патогенности $\Omega_{\mathcal{K}}^m$ подчёркивает степень, с которой группа водородных связей \mathcal{K} нарушает структуру белка.

Положительно коррелирующие с дожитием пациентов стабильности означают, что чем более стабильна водородная связь в мутанте, тем более долгое дожитие наблюдается у пациента. Аналогично, отрицательная корреляция означает, что чем большую стабильность имеет водородная связь в мутанте, тем меньшее дожитие наблюдается у пациента с данной мутацией. Можно предположить, что водородные связи с положительной корреляцией их стабильности демонстрируют стабилизирующее влияние на структуру белка, в то время как связи с отрицательной корреляцией стабильностей – дестабилизирующее. С этой точки зрения, величина $\Omega_{\mathcal{K}}^m$ отражает баланс между стабилизирующими и дестабилизирующими водородными связями.

На основе меры $\Omega_{\mathcal{K}}^m$, которая выступала в качестве предсказывающего фактора, строилась регрессионная модель. Формула 4.2 принимает во внимание предположение, что как дестабилизированные ($R_k < 0$), так и стабилизированные ($R_k > 0$) связи оказывают влияние на патогенность той или иной мутации в SOD1. Суммирование времён существования связей позволило снизить количество предикторов и построить регрессионную модель, связывающую коллективные изменения стабильности водородных связей с дожитием пациентов с БАС. Аналогичный подход суммирования значений предикторов для различных регрессионных моделей показал свою эффективность в работе [7].

Регрессионные уравнения для каждого из параметров (PPhb, PWhb, Wbr и PPhb_{ЭС}) имели следующий вид:

$$\hat{\Theta}^m = a_1 * \Omega_{\mathcal{R}}^m + a_0 \quad (4.3)$$

Для построения регрессионных моделей, связывающих каждое из конформационных свойств SOD1 (PPhb, PWhb, Wbr и PPhb_{ЭС}) с дожитием Θ (или Θ_{exp}) пациентов с БАС, требовалось подобрать пороговое значение R_{th} . Эта величина должна отсекает водородные связи, чья стабильность в мутантных белках коррелирует с меньшим, чем R_{th} , коэффициентом корреляции с Θ (или Θ_{exp}) (см. уравнение 4.2). Кроме того, требовалось также заполнить множество \mathcal{R} номерами оставшихся после такого отсечения водородных связей.

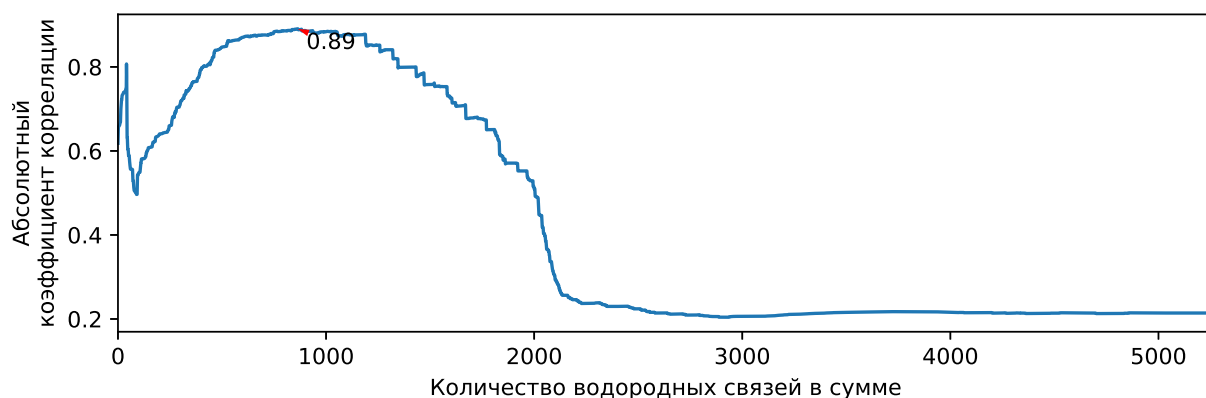


Рисунок 4.1 — Коэффициент корреляции между $\Omega_{\mathcal{R}}^m$ и Θ^m для водородных связей из группы PPhb.

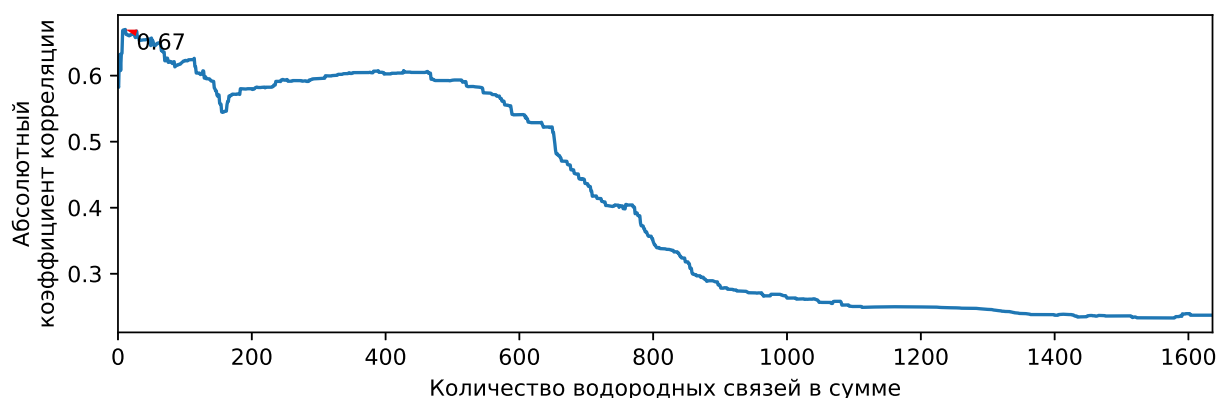


Рисунок 4.2 — Коэффициент корреляции между $\Omega_{\mathcal{R}}^m$ и Θ^m для водородных связей из группы PWhb.

Итерационная процедура подбора соответствующего R_{th} и \mathcal{R} заключалась в следующем.

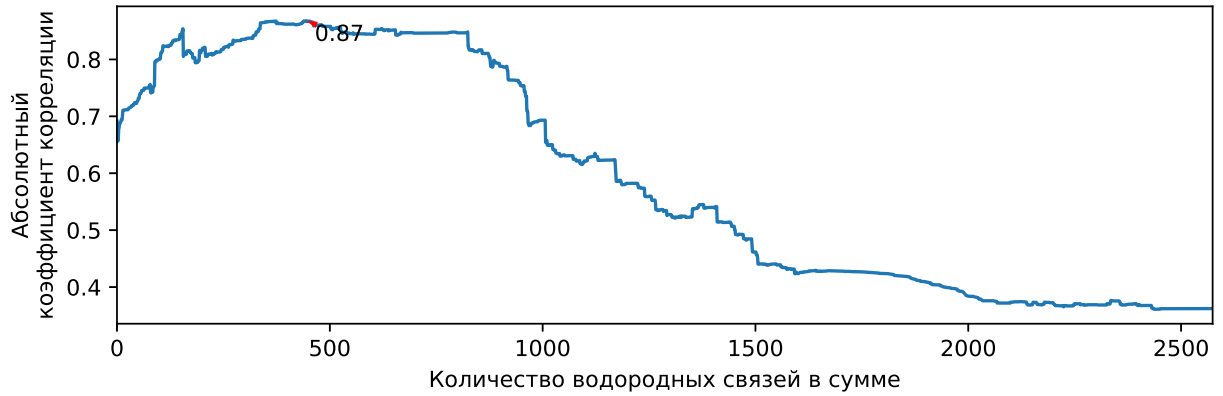


Рисунок 4.3 — Коэффициент корреляции между $\Omega_{\mathfrak{R}}^m$ и Θ^m для водородных связей из группы Wbr.

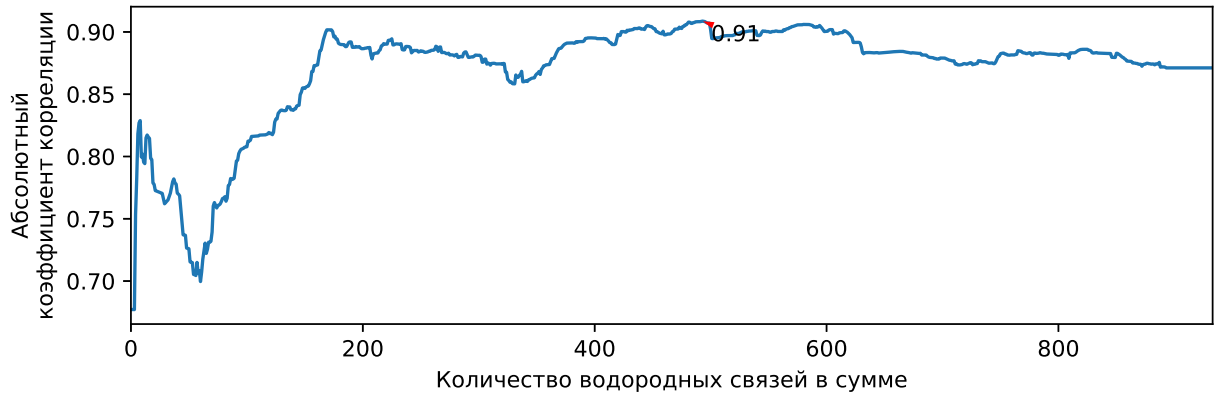


Рисунок 4.4 — Коэффициент корреляции между $\Omega_{\mathfrak{R}}^m$ и Θ^m для водородных связей из группы PPhbЭС.

1. В таблице \bar{O} строки (соответствующие конкретной водородной связи в различных мутантах) отсортированы по убыванию абсолютного значения коэффициента корреляции Пирсона между \bar{O}_i^m и дожитием пациентов Θ^m (или Θ_{exp}^m), а их номера наряду с порядком составляли вектор \mathfrak{J} .
2. Последовательно для каждого $K \in [1; N]$, на основе набора первых K компонент $\{k\}$, полученного из вектора \mathfrak{J} , составлялось множество \mathfrak{K}_K и вычислялось значение $\Omega_{\mathfrak{R}}^m K$.
3. Оценивался коэффициент корреляции R_K между $\Omega_{\mathfrak{R}}^m K$ и Θ^m (или Θ_{exp}^m) и строился график R_K в зависимости от K (для выборки Θ^m см. рис. 4.1–4.4).
4. В качестве \mathfrak{K} бралось такое \mathfrak{K}_K , коэффициент R_K для которого был максимален.

5. Пороговый коэффициент R_{th} в этом случае принимался равным коэффициенту корреляции между строкой $\overline{O}_{\mathcal{J}_K}^m$ и Θ^m (или Θ_{exp}^m).

Данный метод построения регрессионных моделей далее будет упоминаться как метод оптимальных сумм абсолютных разностей стабильности водородных связей (ОСАРС).

Поскольку в процессе расчёта величины $\Omega_{\mathcal{R}}^m$ для каждой водородной связи вычисляется коэффициент корреляции между её стабильностью в мутантах SOD1 и дожитием пациентов с данными мутациями, то возникает возможность отобрать часть наиболее хорошо коррелирующих связей для исследования того, какие аминокислотные остатки формируют эти примечательные водородные связи.

4.1.2 Построение многофакторных регрессионных моделей

Группировка водородных связей по элементам вторичной структуры SOD1

Был предложен также альтернативный метод построения регрессионной модели, предсказывающей дожитие пациентов с мутациями в SOD1, исходя из расположения аминокислотных остатков, формирующих водородные связи, в элементах вторичной структуры (модель SSRF). С помощью программы DSSP [212] была получена вторичная структура белка SOD1. В структуре для каждой из двух субъединиц выделяется 23 элемента:

1. α -спирали:
 - а) Н1 (остатки 57–60);
 - б) Н2 (остатки 108–111);
 - в) Н3 (остатки 132–135);
2. петли:
 - а) С1 (остатки 1–2);
 - б) С2 (остатки 10–14);
 - в) С3 (остатки 23–28);
 - г) С4 (остатки 37–40);
 - д) С5 (остатки 49–56);
 - е) С6 (остатки 61–82);

- ж) С7 (остатки 90–94);
- и) С8 (остатки 102–107);
- к) С9 (остатки 112–115);
- л) С10 (остатки 121–131);
- м) С11 (остатки 136–142);
- н) С12 (остатки 152–153);

3. β -тяжи:

- а) E1 (остатки 3–9);
- б) E2 (остатки 15–22);
- в) E3 (остатки 29–36);
- г) E4 (остатки 41–48);
- д) E5 (остатки 83–89);
- е) E6 (остатки 95–101);
- ж) E7 (остатки 116–120);
- и) E8 (остатки 143–151).

Все обнаруженные в исследуемых мутантах SOD1 водородные связи группировались в зависимости от принадлежности формирующих их аминокислотных остатков одному из элементов вторичной структуры, перечисленных выше. 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. Таким образом, каждому выделенному элементу вторичной структуры ставился в соответствие набор водородных связей, образованных с участием данного элемента.

Для каждой водородной связи, входящей в одну группу по формуле 2.2 рассчитывалась стабильность O . Полученные стабильности O_i^m суммировались для водородных связей из одной группы. В результате была получена таблица O_s^m стабильностей элементов вторичной структуры SOD1. Регрессионная модель, предсказывающая дожитие пациентов с мутациями в SOD1, получалась путём обучения множественной регрессии на основе метода random forest [213], в которую в качестве признаков входили значения стабильности O_s^m элементов вторичной структуры белка для каждого из изучаемых мутантов. Для построения регрессии было использовано 100 решающих деревьев.

Регрессионная модель SSRF, основанная на методе random forest позволяет получить оценки важности признаков для предсказания дожития пациентов. Метод random forest реализует данную возможность в виде усреднённой по всем деревьям ансамбля величины снижения среднеквадратичной ошибки предсказаний при перемешивании значений выбранного признака в сравнении с ошибкой предсказаний на основе исходной выборки. Поскольку признаками в данной модели является стабильность O_s^m водородных связей, образованных с участием выделенных выше элементов вторичной структуры SOD1, то эта информация может быть использована для определения важности элементов вторичной структуры SOD1 и аминокислотных остатков, образующих эти элементы.

Регрессионные модели, использующие метод random forest и ридж-регрессию

Помимо новых регрессионных моделей, предложенных в предыдущих подразделах, для предсказания дожития пациентов, несущих определённую мутацию в SOD1, использовались уже известные методы: random forest и гребневая (ридж-) регрессия [214]. Метод random forest было интересно применить потому, что в нём учитываются сложные нелинейные взаимные связи между используемыми предикторами и происходит неявная перекрёстная проверка [215]. Гребневая регрессия является более устойчивой к переобучению по сравнению с обычной линейной регрессией за счёт штрафов, накладываемых на величину коэффициентов регрессионного уравнения.

Регрессионная модель random forest (RF) строилась с использованием пакета scikit-learn 0.19.1 для Python 2.7. В качестве параметров выбраны следующие:

1. Количество деревьев (n_estimators): 100;
2. Мера качества разбиения (criterion): «mse»;
3. Максимальное количество предикторов (max_features): «auto»;
4. Максимальная глубина дерева (max_depth): неограниченна;
5. Минимальное количество элементов выборки для разбиения внутренней вершины (min_samples_split): 2;
6. Минимальное количество элементов выборки в вершине-листе (min_samples_leaf): 1;

7. Максимальное количество вершин-листов (`max_leaf_nodes`): неограниченно;
8. Минимальное снижение меры «impurity» (`min_impurity_decrease`): 0;
9. Использование `bootstrap` (`bootstrap`): используется;
10. Использование элементов выборки для оценки R^2 на тестовой выборке (`oob_score`): не используется;
11. Состояние датчика псевдослучайных чисел (`random_state`): 0;
12. Добавление новых деревьев к предыдущему решению (`warm_start`): не используется.

Гребневая регрессия (RR) также строилась с применением пакета `scikit-learn`. Для неё выбраны следующие параметры:

1. Глубина регуляризации (`alpha`): 1;
2. Центрирование данных (`fit_intercept`): используется;
3. Нормализация данных (`normalize`): не используется;
4. Максимальное количество итераций для метода сопряжённых градиентов (`max_iter`): выбирается автоматически;
5. Точность решения (`tol`): 0.0001;
6. Решатель (`solver`): «auto»;
7. Состояние датчика псевдослучайных чисел (`random_state`): 0.

Для индивидуальных регрессионных моделей на основе `random forest` и гребневой регрессии использовались факторы O_i^m , полученные для водородных связей между атомами белка (PPhb), атомами белка и атомами молекул воды (PWhb) и стабильность водных мостиков (Wbr) из траекторий МД, а также для водородных связей между атомами белка, из конформаций ЭС (PPhb_{ЭС}). В отличие от однофакторной простой регрессионной модели на основе метода OCAPC, в которой используется только один суммарный признак, обе регрессионные модели RF и RR являются множественными регрессиями, которые строились на основе значений стабильности всех обнаруженных в результате моделирования водородных связей одного типа. Специального отбора признаков не делалось.

Как и в случае с регрессионной моделью SSRF, модели RF и RR позволяют получить оценки важности признаков для предсказания дожития пациентов. Принцип определения важности признаков в рамках метода `random forest` был описан выше при определении модели SSRF. Метод гребневой регрессии, в свою очередь, позволяет получить коэффициенты уравнения регрессии для каждого из

признаков, которые могут использоваться в качестве меры важности этих признаков. Поскольку стабильность водородных связей – признаки – сходна по величине между всеми связями, то более высокие абсолютные значения коэффициентов при выбранном признаке будут говорить о более высокой важности данной водородной связи.

4.1.3 Комбинированные регрессионные модели

Чтобы учесть конформационные свойства, рассчитанные из моделирования МД и ЭС в одной регрессионной модели, было построено восемь множественных регрессионных моделей (MR) с использованием гребневой регрессии с параметрами по умолчанию, кроме параметра α , установленного в 1. В качестве факторов каждой множественной регрессии выбраны предсказанные по параметрам PPhb, PWhb, Wbr и PPhb_{ЭС} значения $\hat{\Theta}^m$ дожития пациентов одной моделью из: модели на основе метода OSCAR, SSRF, RF и RR. Другими словами, на основе обучающих выборок Θ и Θ_{exp} было построено по четыре четырёхфакторных регрессионных моделей. Уравнение каждой из них имело вид:

$$\hat{\Theta}^m = a_1 * \hat{\Theta}^m_{PPhb} + a_2 * \hat{\Theta}^m_{PWhb} + a_3 * \hat{\Theta}^m_{Wbr} + a_4 * \hat{\Theta}^m_{PPhb_{ЭС}} + a_0 \quad (4.4)$$

Здесь параметры a_0 – a_4 – коэффициенты уравнения гребневой регрессии. Факторы $\hat{\Theta}^m_{PPhb}$ и $\hat{\Theta}^m_{PPhb_{ЭС}}$ – предсказания дожития пациентов на основе стабильностей водородных связей, усреднённых по траекториям МД и по ансамблю конформаций из ЭС, соответственно.

4.1.4 Исследование точности регрессионных моделей

Проверочная регрессионная модель

В качестве проверочной регрессионной модели (Rhb) выступала линейная регрессия, использующая для предсказания стабильность O_i^m только одной водородной связи, выбранной из всех водородных связей случайным образом. При обучении модели целевой признак (наблюдаемое дожитие пациентов) случайным образом перемешивался. Построенная таким образом модель являлась оценкой сверху при анализе точности регрессионных моделей.

Перекры́стная проверка

Для статистической проверки точности регрессионных моделей OSCAR, SSRF, RF, RR и Rhb по предсказанию дожития пациентов был применён метод bootstrap. Для этого данные по дожитию пациентов, включающие в себя информацию об M мутантах (множество \mathfrak{M}) были разбиты на обучающую ($\frac{3}{5}$ элементов \mathfrak{M}) и тестовую выборку ($\frac{2}{5}$ элементов \mathfrak{M}). Было проведено 100 шагов bootstrap, на каждом из которых оценивалась величина стандартной ошибки S , как средний квадрат разности между полученными с помощью модели предсказаниями и опубликованным в литературе дожитием пациентов, носителей соответствующих мутаций. В качестве множества \mathfrak{M} выступало таковое для Θ или Θ_{exp} ; соответствующее M принимало значение 35 или 566.

Полученные распределения величины S для исходных моделей сравнивались с распределением S , рассчитанным для случайных моделей, с применением непараметрического критерия Манна-Уитни-Вилкоксона.

4.1.5 Аминокислотные остатки SOD1, формирующие водородные связи, важные для предсказания дожития пациентов

При расчёте величины $\Omega_{\mathcal{R}}^m$ в рамках метода OSCAPC для каждой водородной связи вычисляется коэффициент корреляции между её стабильностью \bar{O}_i^m (см. уравнение 4.1) в мутантах SOD1 и дожитием пациентов Θ^m (или Θ_{exp}^m) с данными мутациями. Полученный коэффициент корреляции можно использовать в качестве меры важности соответствующей водородной связи для структуры белка SOD1. Кроме того, можно отследить, какие аминокислотные остатки участвуют в образовании таких важных водородных связей. Таким образом, для каждого типа водородных связей из PPhb, PWhb, Wbr и PPhb_{ЭС} были проанализированы водородные связи, коэффициент корреляции R_k которых удовлетворял условию: $|R_k| > R_{th}$ (см. уравнение 4.2). Из всех обнаруженных водородных связей выбирались связи, коэффициент корреляции которых по модулю лежал в 1-% хвосте распределения R_k . Набор, в общем случае повторяющихся, позиций аминокислотных остатков, участвующих в образовании отобранных водородных связей для каждого из изучаемых четырёх типов далее будет упоминаться как \mathcal{T} . Всего для регрессионной модели OSCAPC выделено восемь различных наборов \mathcal{T}_{OSCAPC} – по одному на тип водородной связи в соответствии с наблюдаемыми значениями дожития Θ и столько же для Θ_{exp} . Остатки из разных субъединиц SOD1 считались разными и, в случае совпадения добавлялись в \mathcal{T}_{OSCAPC} дважды. Везде, где это потребуется, для \mathcal{T}_{OSCAPC} будет указываться конкретный тип водородных связей, остатки которого использованы для наполнения набора, а также будет указываться вариант наблюдаемых значений дожития пациентов с мутациями в SOD1, на основе которых строилась регрессионная модель OSCAPC. Аналогичные наборы важных аминокислотных остатков фиксировались для моделей SSRF, RF и RR.

Регрессионная модель SSRF, будучи построенной с использованием метода random forest, позволяет получить оценки важности признаков – элементов вторичной структуры SOD1 – для предсказания дожития пациентов, как описано двумя разделами выше. Элементы вторичной структуры, чья важность, как признаков, в рамках метода random forest лежала в 5-% хвосте распределения оценок важности 46 элементов вторичной структуры выделялись в набор важных

признаков. Аминокислотные остатки, образующие данные важные элементы, заносились в набор $\mathfrak{T}_{\text{SSRF}}$. Как и в случае OCAPC, всего выделено восемь наборов $\mathfrak{T}_{\text{SSRF}}$.

Регрессионная модель RF также использует метод random forest. Среди признаков – водородных связей – отбирались наиболее значимые для предсказания дожития пациентов – находящиеся в 1-% хвосте распределения важности этих признаков. Набор позиций аминокислотных остатков, формирующих отобранные водородные связи обозначен как \mathfrak{T}_{RF} . Всего сформировано восемь наборов \mathfrak{T}_{RF} : для четырёх типов водородных связей и двух наборов наблюдаемых значений дожития пациентов.

При построении модели RR выбраны признаки – водородные связи –, абсолютные коэффициенты уравнения гребневой регрессии которых лежат в 1-% хвосте распределения этих коэффициентов. Далее набор позиций аминокислотных остатков, формирующих выбранные водородные связи будет упоминаться как \mathfrak{T}_{RR} . Как и в предыдущих трёх случаях, всего фиксировалось восемь наборов \mathfrak{T}_{RR} .

Все полученные 32 набора \mathfrak{T} также комбинировались в один – $\widehat{\mathfrak{T}}$ – который включал в себя все, в том числе и повторяющиеся более одного раза, позиции важных аминокислотных остатков \mathfrak{T} .

Позиции остатков из всех 32 наборов важных остатков \mathfrak{T} , а также из $\widehat{\mathfrak{T}}$, сопоставлялись с позициями из наборов $\mathfrak{S}_{\text{agr}}^9$, $\mathfrak{S}_{\text{дест}}$, $\mathfrak{S}_{\text{гфоб}}^9$ (см. разделы 2.7 и 2.8). Для этого для каждого набора важных аминокислотных остатков рассчитывалась мера $N(x)$ – нормированная доля количества $F(x)$ этих остатков, находящихся в заданной позиции x в первичной структуре мономера SOD1:

$$N(x) = \frac{F(x)}{\sum_{x=1}^{153} F(x) \times \max\left(\frac{F(x)}{\sum_{x=1}^{153} F(x)}\right)} \quad (4.5)$$

Значимость пересечения номеров аминокислотных остатков из пары наборов A и B определялась с помощью гипергеометрического распределения с помощью модуля scipy для Python. Параметры распределения задавались следующим образом:

1. M – размер генеральной совокупности – устанавливался в 153 (количество аминокислотных остатков в мономере SOD1);
2. n – количество особых объектов в генеральной совокупности – приравнивалось к количеству остатков в B ;

3. N – размер выборки – устанавливалось в значение, равное количеству остатков в A ;
4. k – количество особых объектов в выборке – приравнивалось к количеству пересекающихся позиций остатков в A и B .

Проводилось также исследование того, с какими позициями остатков из ключевых элементов структуры SOD1 есть значимое пересечение наборов \mathcal{T} . Для этого выделялись остатки в позициях:

1. 57, 146 – дисульфидная связь (коротко «SS»);
2. 46, 48, 63, 120 – сайт связывания иона меди (коротко «Cu»);
3. 63, 71, 80, 83 – сайт связывания иона цинка (коротко «Zn»);
4. 121–142 – электростатическая петля (коротко «E»);
5. 102–115 – «greek-key»-петля (коротко «G»);
6. 49–62 – дисульфидная петля (коротко «S»);
7. 63–85 – цинк-связывающая петля (коротко «Z»);
8. 5, 7, 17, 50–54, 113–115, 148, 150–153 – интерфейс между субъединицами (коротко «I»).

Набор позиций, соответствующих дисульфидной связи, сайтам связывания ионов и интерфейсу между субъединицами, далее будет называться $\mathcal{S}_{\text{крит}}$.

Кроме того, для каждого набора из 35 сформированных рассчитывался вектор \mathcal{J} длины 23 с компонентами, соответствующими количеству остатков из набора, формирующих тот или иной элемент вторичной структуры SOD1 из 23 выделенных ранее элементов (см. стр. 69). Данные векторы \mathcal{J} , построенные для каждого набора \mathcal{T} , а также $\widehat{\mathcal{T}}$, сравнивались с векторами \mathcal{J} наборов $\mathcal{S}_{\text{агр}}^9$ и $\mathcal{S}_{\text{глоб}}^9$ с помощью коэффициента корреляции Пирсона между компонентами пары заданных векторов, а также уровня значимости этого коэффициента корреляции.

4.2 Исследование регрессионных моделей

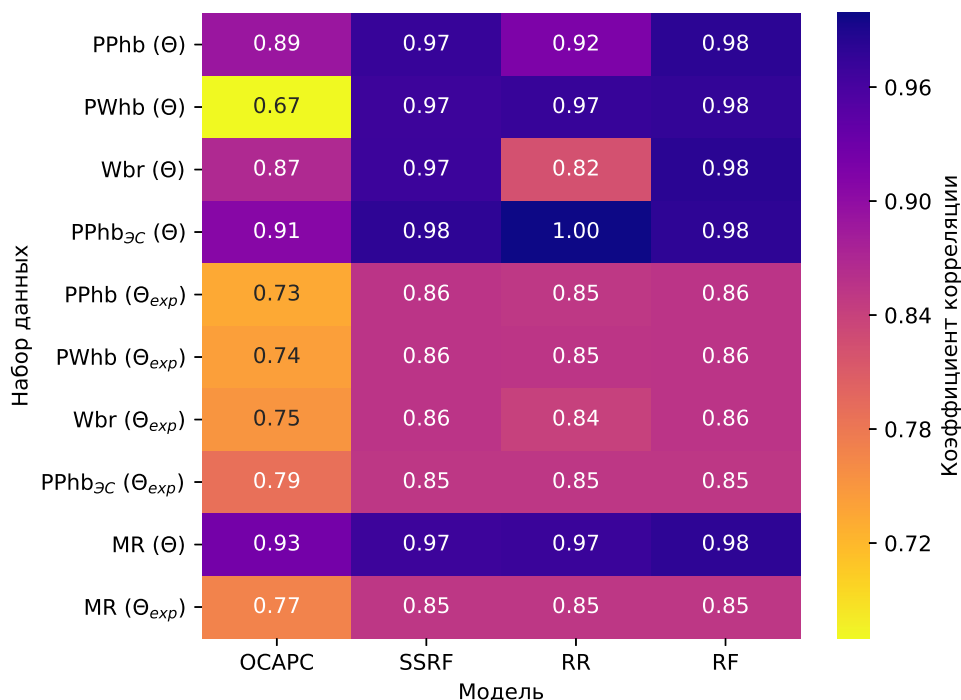
4.2.1 Анализ точности регрессионных моделей

Для построения моделей, предсказывающих дожитие пациентов, несущих определённую патологическую мутацию, исследовались структурные и динамические характеристики белка. В фазовой траектории МД отслеживалась стабильность трёх типов водородных связей: внутри белка (PPhb), между белком и окружающими молекулами воды (PWhb), а также между аминокислотными остатками белка, через взаимодействие с молекулами воды – водных мостиков (Wbr). В ансамбле конформаций ЭС анализировалась стабильность водородных связей только между атомами белка (PPhb_{ЭС}).

В качестве наблюдаемых значений дожития пациентов с БАС, являющихся носителями мутаций в SOD1 выступали значения Θ и Θ_{exp} . При обучении регрессионных моделей на основе расширенной выборки Θ_{exp} данные о стабильности водородных связей O_i^m подлежали соответствующему расширению. Это делалось таким образом, чтобы значения O_i^m , используемые в качестве факторов для предсказания, для каждого мутанта m повторялись n раз. Здесь n – количество пациентов с мутацией m по данным из работы [7] (см. также раздел 2.10). Таким образом, регрессионные модели строились на основе восьми наборов данных: значения O_i^m из PPhb, PWhb, Wbr и PPhb_{ЭС}, которым ставились в соответствие значения Θ^m , а также расширенные O_i^m из этих же четырёх наборов данных, которым ставились в соответствие значения Θ_{exp}^m . Всего в имеющейся выборке Θ содержалась информация о конформационных свойствах и дожитии пациентов с 33 мутациями в SOD1; её расширенный вариант Θ_{exp} содержал дожитие 566 пациентов.

В качестве предсказывающих регрессионных моделей использовались однофакторные модели OCAPC и Rhb, а также многофакторные модели SSRF, RF, RR и четырёхфакторные модели MR, построенные, согласно описанию в разделе 4.1.

Коэффициенты корреляции для каждой из построенных регрессионных моделей приведены на диаграмме, представленной на рис. 4.5. Из рисунка видно, что регрессионные модели с наивысшим коэффициентом корреляции получены



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

Рисунок 4.5 — Значения коэффициентов корреляции между конформационными свойствами мутантов белка SOD1 и дожитием пациентов с БАС с соответствующими мутациями.

на основе выборок, не подвергавшихся расширению, то есть соответствующих наблюдаемым значениям Θ . Оказалось, что наибольший коэффициент корреляции имеют модели SSRF, RR и RF ($R > 0.97$, $p \ll 0.01$). При этом стабильность водородных связей из группы PPhb_{ЭC} наиболее сильно коррелировала с Θ по сравнению с остальными группами связей ($R > 0.91$, $p \ll 0.01$). Четырёхфакторные регрессионные модели, построенные на основе всех четырёх групп PPhb, PWhb, Wbr и PPhb_{ЭC}, также имели одни из самых высоких коэффициентов корреляции ($R > 0.93$, $p \ll 0.01$). На основе расширенных выборок стабильностей водородных связей, соответствующих Θ_{exp} , регрессионные модели получились с ожидаемо более низкими, но, при этом, более значимыми коэффициентами корреляции ($R \sim 0.85$, $p < 10^{-100}$).

Как видно из рис. 4.6, при построении регрессионных моделей с использованием выборки Θ_{exp} , предсказанные значения дожития попадают в 95 % интервал наблюдаемых значений дожития из выборки Θ . Этот факт свидетельствует в пользу того, что теоретическая выборка Θ_{exp} , сформированная с

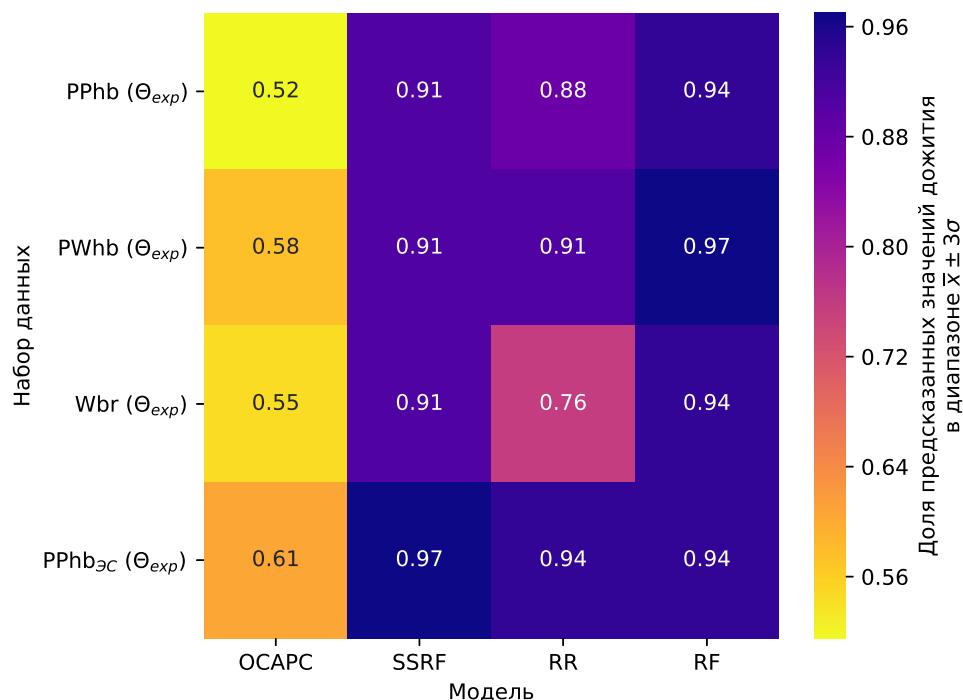


Рисунок 4.6 — Доля предсказанных значений дожития в диапазоне наблюдавшегося среднего значения дожития (из Θ) $\pm 3\sigma$.

помощью использованного в данной работе подхода, отражает действительные средние величины дожития пациентов с мутациями в SOD1.

С использованием метода bootstrap в 100 шагах была исследована среднеквадратичная ошибка предсказаний S для каждой из построенных регрессионных моделей (см. рис. 4.7). Моделями с наименьшей медианой среднеквадратичной ошибки предсказаний дожития пациентов оказались модели на основе метода random forest (SSRF, RF, $S = 4$ года), которые были построены для группы связей PPhbЭС, расширенного в соответствии с Θ_{exp} . Четырёхфакторные модели с использованием этих же методов SSRF и RF имели практически идентичную лучшим одиночным моделям ошибку предсказаний ($S = 4.02$ года). Примечательно, что в противоположность диаграмме коэффициентов корреляции (рис. 4.5), модели на основе расширенных выборок для Θ_{exp} продемонстрировали заметно меньшие ошибки предсказаний (~ 4 лет в сравнении с > 5 лет).

Для каждой модели из OCAPC, SSRF, RR и RF по критерию Манна-Уитни-Вилкоксона была рассчитана достоверность отличия среднего среднеквадратичной ошибки S от такового для проверочной модели Rhb (см. рис. 4.8). Выяснилось, что не смотря на то, что эти регрессионные модели на выборке Θ продемонстрировали высокий ($R \sim 0.9$) коэффициент корреляции с дожитием

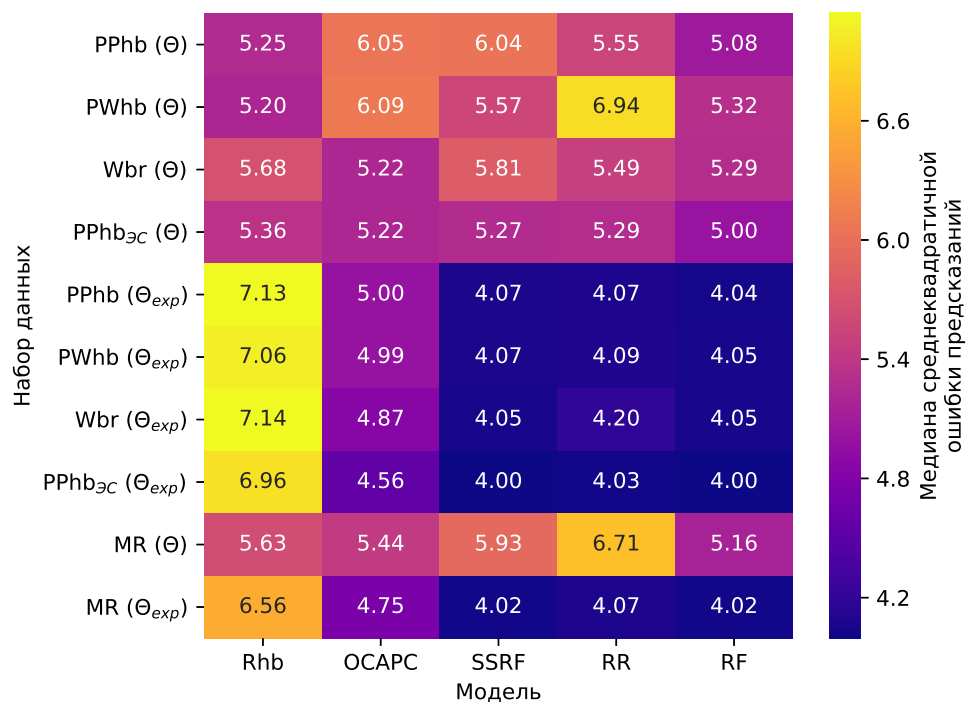
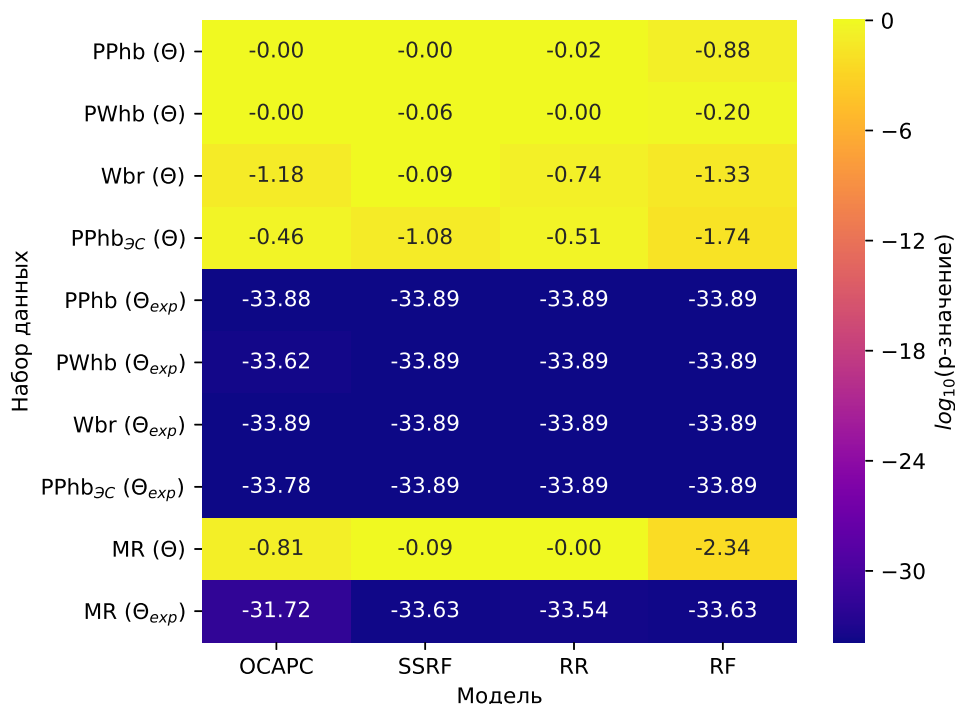


Рисунок 4.7 — Значения медианы среднеквадратичной ошибки предсказаний дожития пациентов с изучаемыми мутациями в SOD1, полученной в 100 шагах bootstrap.

пациентов, средние ошибки предсказаний большинства из них были статистически неотличимы от средних ошибок модели Rhb. Исключением являются модели RF, построенные на основе стабильностей Wbr и PPhb_{ЭС}, и модель MR_{RF}, построенная на основе Θ , которые имели значимое ($p < 0.05$) отличие средней ошибки предсказаний от таковой для для проверочной модели Rhb. Все регрессионные модели, построенные на наборах PPhb, PWhb, Wbr и PPhb_{ЭС} (соответствующих Θ_{exp}), имели значимо сниженную ошибку предсказаний по сравнению с моделью Rhb.

Таким образом, одной из наилучших регрессионных моделей в отношении ошибки предсказаний, отличия её от ошибки самой неудачной модели и коэффициента корреляции с дожитием пациентов является модель RF. Графики зависимости между предсказанным с помощью модели RF_{PPhb_{ЭС}} и наблюдаемым дожитием пациентов представлены на рис. 4.9 (а, б).



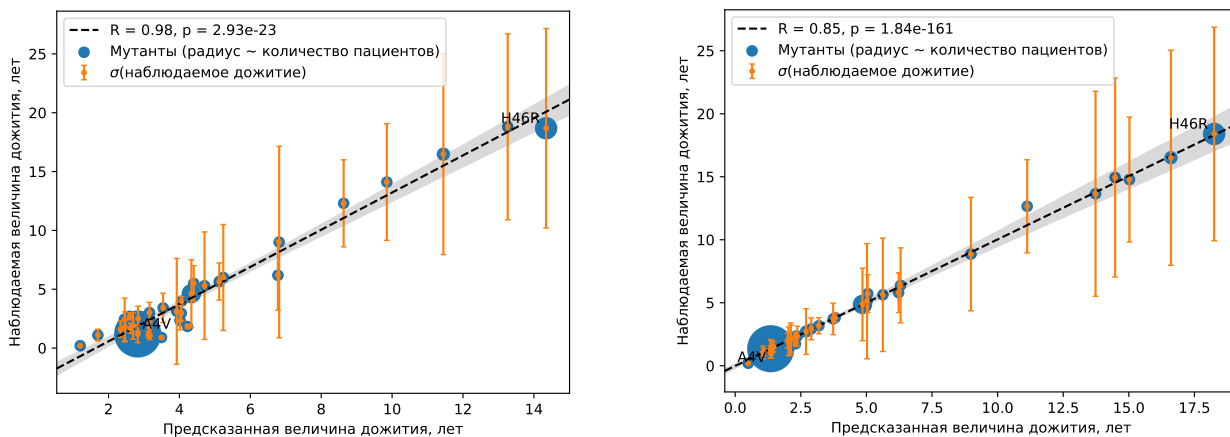
Значения логарифмов ниже -1.3 соответствуют р-значениям < 0.05 .

Рисунок 4.8 — Десятичный логарифм р-значения отличий среднего среднеквадратичной ошибки S , которую совершают регрессионные модели ОСАРС, SSRF, RF и RR при предсказании дожития пациентов, от среднего ошибки проверочной модели Rhb в соответствии с односторонним критерием Манна-Уитни-Вилкоксона.

4.2.2 Конформационные свойства белка SOD1, важные для предсказания дожития пациентов

Расположение важных аминокислотных остатков в первичной структуре белка

В соответствии с описанием в разделе 4.1.5 (см. стр. 75), при построении регрессионных моделей были сформированы наборы важных аминокислотных остатков \mathfrak{T} . Так при обучении модели ОСАРС на данных о дожитии пациентов Θ было выделено 56, 7, 22 и 9 уникальных аминокислотных остатков $\mathfrak{T}_{\text{ОСАРС}}$ (для PPhb, PWhb, Wbr и PPhb_{ЭС}, соответственно), а при обучении этой модели на Θ_{exp} – 8, 9, 27 и 11. При построении регрессионной модели SSRF было выделено 9, 15, 9 и 11 аминокислотных остатков $\mathfrak{T}_{\text{SSRF}}(\Theta)$, а также 5, 14, 9 и 30 $\mathfrak{T}_{\text{SSRF}}(\Theta_{exp})$.

а) Θ б) Θ_{exp}

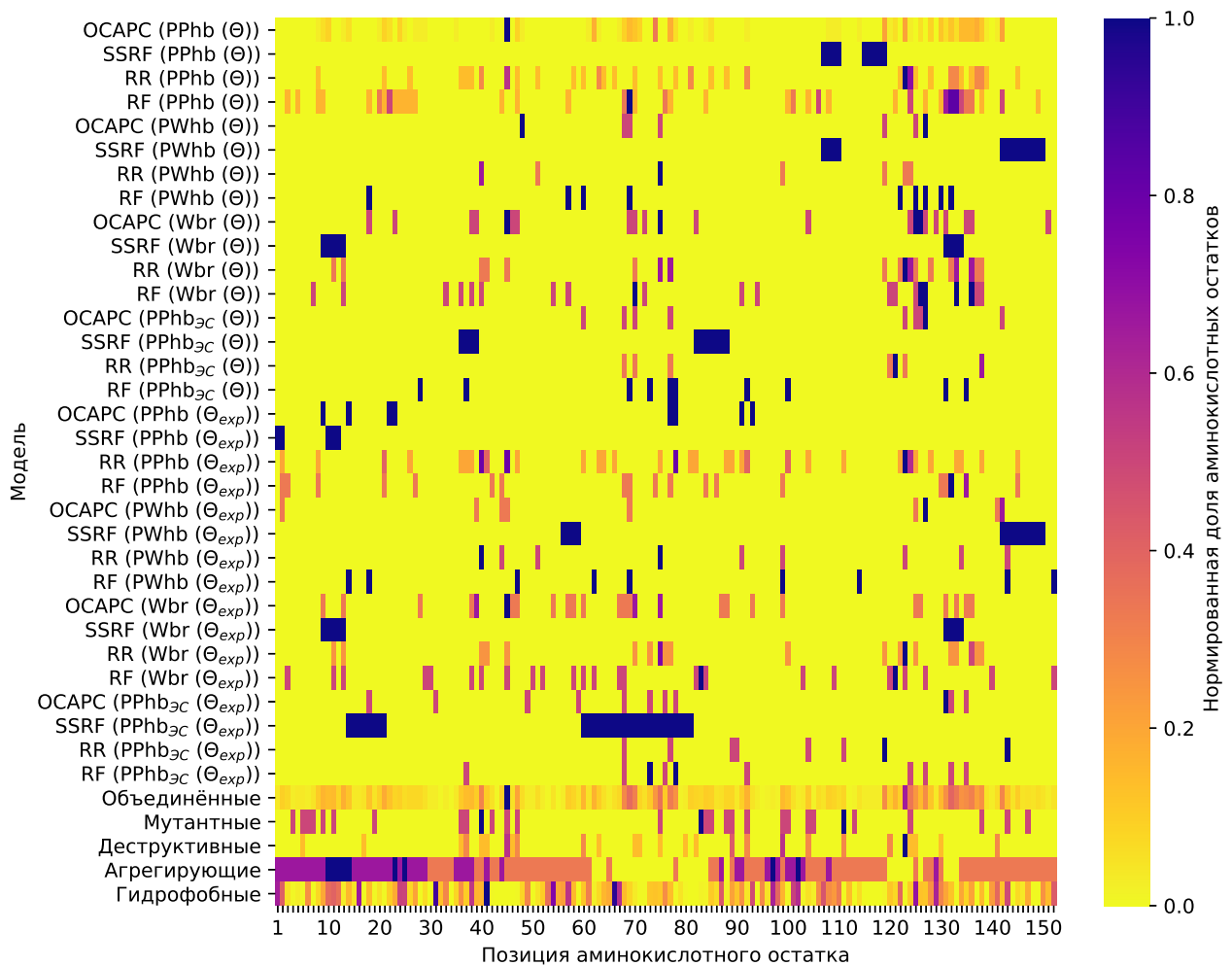
На графиках отмечены точки, соответствующие мутациям (Ала4Вал и Гис46Арг), для которых известно наибольшее количество пациентов. Радиусы кругов отражают количество пациентов с соответствующей мутацией.

Рисунок 4.9 — Зависимость между наблюдаемым дожитием пациентов и предсказанным с помощью модели $RF_{pphb_{\Theta}}$, построенной на основе выборок Θ (а) и Θ_{exp} (б).

Обучение модели RF на Θ позволило выделить 40, 9, 21 и 10 не повторяющихся остатков \mathfrak{T}_{RF} , а обучение её на Θ_{exp} — 19, 9, 26 и 10. При построении модели RR выбрано 35, 7, 19, 7 (Θ) и 35, 9, 20 и 8 (Θ_{exp}) важных остатков \mathfrak{T}_{RR} .

Для понимания того, в каких областях находятся аминокислотные остатки из наборов \mathfrak{T} , играющие важную роль для структуры белка, был проведён анализ расположения этих остатков в первичной структуре SOD1 (см. рис. 4.10). На представленной тепловой карте отражены значения $N(x)$ (см. уравнение 4.5) — нормированная доля количества остатков из наборов \mathfrak{T} , $\hat{\mathfrak{T}}$, а также из наборов $\mathfrak{S}_{мут}$ (см. раздел 2.10), $\mathfrak{S}_{дест}$ (см. раздел 3.3.2) и \mathfrak{S}_{arg}^9 (см. раздел 2.7, находящихся в заданной позиции x в первичной структуре мономера SOD1, а также меру SAHR (см. раздел 2.8). Из рисунка видно, что остатки в позициях 67–78 и 120–140 в аминокислотной последовательности SOD1 содержатся в достаточно большой доле наборов \mathfrak{T} . При этом часть остатков в позициях 120–140 пересекаются с таковыми для набора \mathfrak{S}_{arg}^9 и набора $\mathfrak{S}_{фоб}^9$, позиции которого выбраны в соответствии с распределением SAHR.

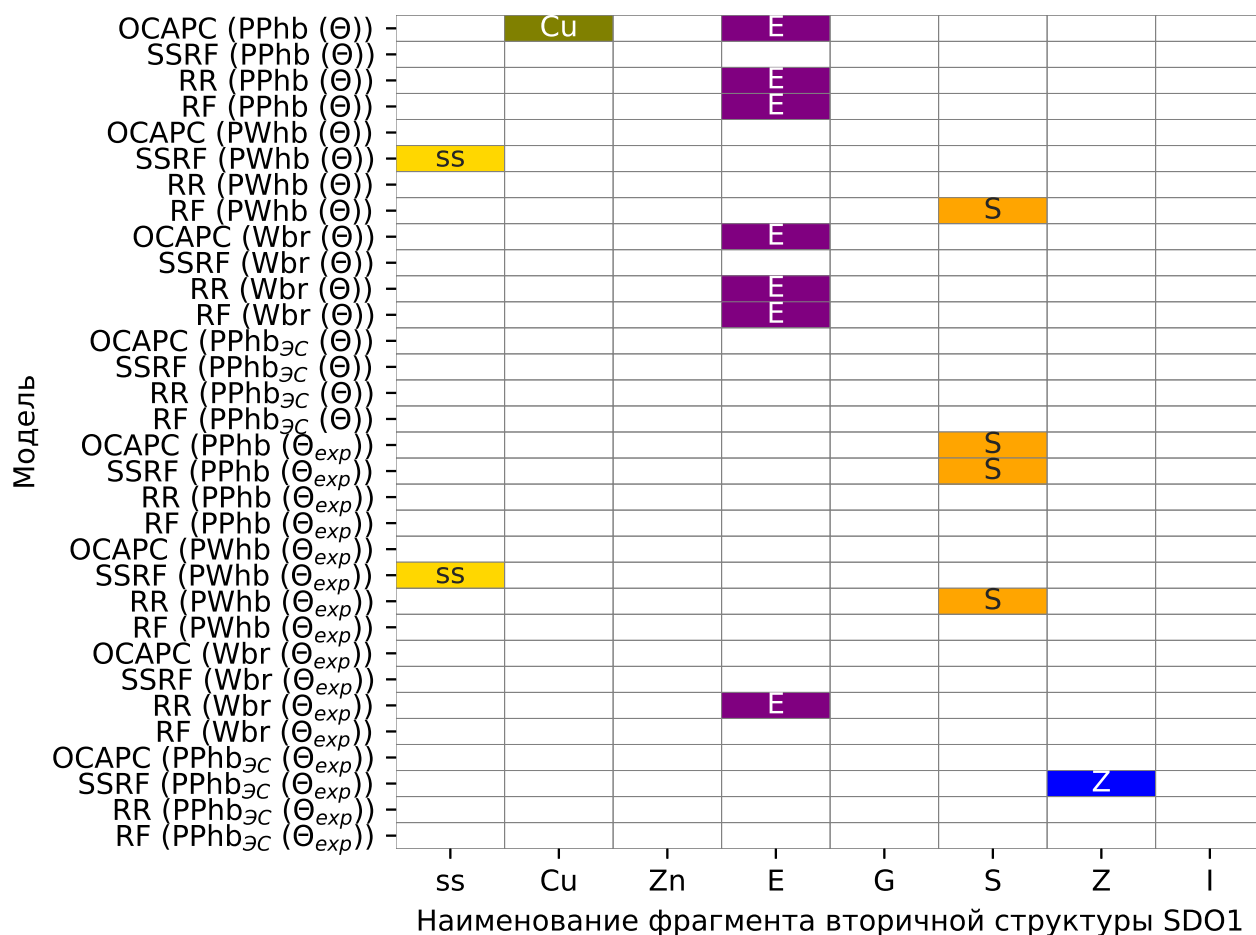
Наборы \mathfrak{T} были охарактеризованы в отношении их значимого ($p < 0.05$, с поправкой на множественное сравнение) пересечения с позициями критически



По оси абсцисс отложены позиции аминокислотных остатков в первичной структуре мономера белка SOD1. По оси ординат – остатки из наборов \mathcal{T} для каждой из регрессионных моделей (а также – указано в скобках – типов водородных связей и наборов наблюдаемых значений дожития пациентов Θ или Θ_{exp}), и остатки из наборов: $\hat{\mathcal{T}}$ («Объединённые»), $\mathcal{S}_{\text{мут}}$ («Мутантные»), $\mathcal{S}_{\text{дест}}$ («Деструктивные») и $\mathcal{S}_{\text{агр}}^9$ («Агрегирующие»). График «Гидрофобные» отражает значения меры SAHR (см. уравнение 2.3).

Рисунок 4.10 — Тепловая карта, отражающая расположение в первичной структуре SOD1 аминокислотных остатков \mathcal{T} , $\hat{\mathcal{T}}$, $\mathcal{S}_{\text{мут}}$, $\mathcal{S}_{\text{дест}}$, $\mathcal{S}_{\text{агр}}^9$, а также меру SAHR.

важных для структуры SOD1 аминокислотных остатков (см. рис. 4.11). Обнаружено, что остатки в позициях \mathcal{T} чаще всего участвовали в формировании дисульфидной связи (значимо пересекались с $\mathcal{T}_{\text{SSRF}}$), сайта связывания иона меди (значимо пересекались с $\mathcal{T}_{\text{OCAPC}}$), электростатической петли (значимо пересекались с $\mathcal{T}_{\text{OCAPC}}$, \mathcal{T}_{RR} и \mathcal{T}_{RF}), дисульфидной петли (значимо пересекались с наборами



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.

Рисунок 4.11 — Диаграмма, отражающая элементы структуры SOD1, позиции остатков которых значимо пересекаются с остатками из \mathcal{T} .

$\mathcal{T}_{\text{OCAPC}}$, $\mathcal{T}_{\text{SSRF}}$, \mathcal{T}_{RR} и \mathcal{T}_{RF}), а также цинк-связывающей петли (значимо пересекались с $\mathcal{T}_{\text{SSRF}}$).

Статистическая проверка пересечений наборов позиций важных аминокислотных остатков

Более чёткое обоснование того какие из наборов \mathfrak{T} значимо пересекаются с наборами $\mathfrak{S}_{\text{мут}}$, $\mathfrak{S}_{\text{дест}}$, $\mathfrak{S}_{\text{агр}}^9$ и $\mathfrak{S}_{\text{гфоб}}^9$ сделано с использованием гипергеометрического распределения и значимости коэффициента корреляции между компонентами пары векторов \mathfrak{J} , построенных для соответствующей пары наборов важных аминокислотных остатков. Пересечение \mathfrak{T} с позициями остатков, в которых имелись мутации ($\mathfrak{S}_{\text{мут}}$), важно отслеживать для того, чтобы отделить аллостерическое влияние мутаций от прямого. Пересечение наборов \mathfrak{T} с позициями из $\mathfrak{S}_{\text{дест}}$ может также дать дополнительную информацию.

Таблица 2 — Параметры гипергеометрического распределения и уровень значимости пересечения наборов \mathfrak{T} и $\mathfrak{S}_{\text{агр}}^9$.

Набор \mathfrak{T} (тип связей (Θ или Θ_{exp}))	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 $\mathfrak{S}_{\text{агр}}^9$, как указано в разделе 4.1.5 на стр. 75. Столбец «p-значение» содержит значимость пересечения позиций остатков из заданной пары наборов \mathfrak{T} и $\mathfrak{S}_{\text{агр}}^9$. Столбец «p-значение (попр.)» содержит p-значения с поправкой на множественное сравнение.

Обнаружено, что наборы $\mathfrak{T}_{\text{SSRF}}$ для связей Wbr и наблюдаемых значений Θ и Θ_{exp} , а также для связей PPhb и Θ_{exp} значимо ($p < 0.001$) пересекаются с набором $\mathfrak{S}_{\text{агр}}^9$ (см. таблицу 2). Среди пересекающихся позиций: 11–14, составляющие петлю C2 во вторичной структуре белка.

Стабильность водных мостиков Wbr, образованных с участием остатков в позициях 11–14 оказалась значимо ($p < 0.001$) ниже в мутантах по сравнению с белком дикого типа (см. рис. 4.12). Однако стабильность водородных связей PPhb,

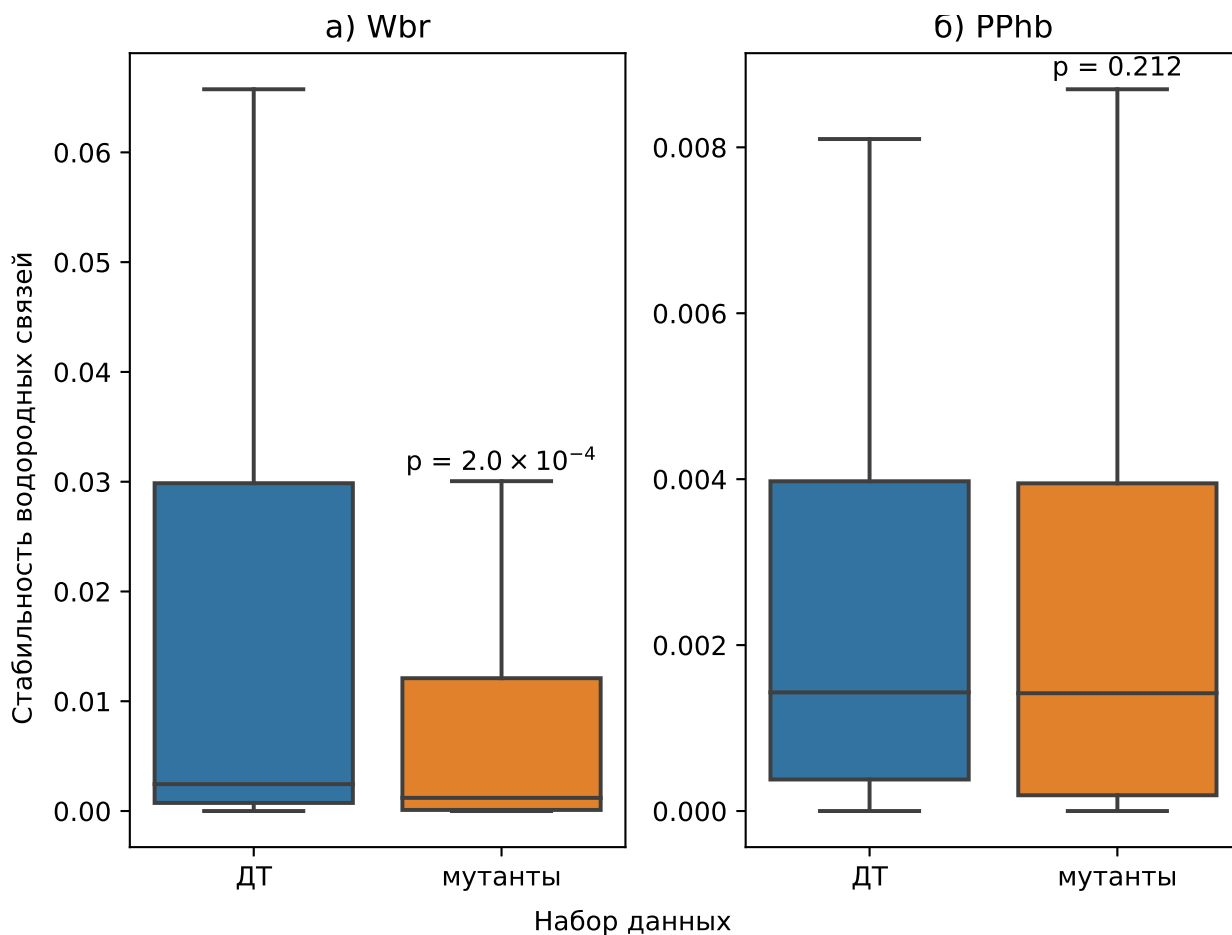


График слева (а) отражает стабильность водородных связей Wbr, справа (б) — PPhb. Значение «р» над графиком стабильностей связей в мутантах является значимостью отличия этих стабильностей от таковых в белке дикого типа по одностороннему критерию Манна-Уитни-Вилкоксона.

Рисунок 4.12 — Стабильность водородных связей Wbr (а) и PPhb (б), сформированных остатками в позициях 11–14, в SOD1 дикого типа и его мутантах.

образованных с участием данных аминокислотных остатков, в мутантах не отличалась от таковой в SOD1 дикого типа.

Исследование корреляции между векторами \mathcal{J} наборов \mathcal{T} и $\mathcal{S}_{\text{гфоб}}^9$ выявило, что ряд векторов \mathcal{J} для наборов важных аминокислотных остатков, найденных при построении каждой из четырёх регрессионных моделей, имел значимый коэффициент корреляции с таковым для $\mathcal{S}_{\text{гфоб}}^9$ ($R > 0.58$, $p < 0.05$; см. таблицу 3). Другими словами, элементы вторичной структуры, содержащие остатки SOD1, формирующие водородные связи, которые, в свою очередь имеют наибольшую важность для предсказания дожития пациентов, часто содержат также гидрофобные остатки, легко доступные для молекул растворителя.

Таблица 3 — Корреляция между компонентами векторов \mathfrak{J} для набора $\mathfrak{S}_{\text{гфоб}}^9$ и для наборов \mathfrak{T} .

Набор \mathfrak{T} (тип связей (Θ или Θ_{exp}))	R	p-значение	p-значение (попр.)
\mathfrak{T}_{RR} (PPhb (Θ))	0.58	3.7×10^{-3}	0.048
$\mathfrak{T}_{\text{OCAPC}}$ (PWhb (Θ))	0.70	1.8×10^{-4}	0.009
\mathfrak{T}_{RR} (PWhb (Θ))	0.66	6.3×10^{-4}	0.016
\mathfrak{T}_{RF} (PWhb (Θ))	0.59	3.0×10^{-3}	0.043
$\mathfrak{T}_{\text{OCAPC}}$ (Wbr (Θ))	0.62	1.6×10^{-3}	0.025
$\mathfrak{T}_{\text{OCAPC}}$ (PPhb $_{\text{ЭС}}$ (Θ))	0.74	6.1×10^{-5}	0.005
\mathfrak{T}_{RR} (PPhb $_{\text{ЭС}}$ (Θ))	0.66	5.4×10^{-4}	0.016
\mathfrak{T}_{RR} (Wbr (Θ_{exp}))	0.63	1.3×10^{-3}	0.025
\mathfrak{T}_{RF} (Wbr (Θ_{exp}))	0.62	1.6×10^{-3}	0.025
\mathfrak{T}_{RF} (PPhb $_{\text{ЭС}}$ (Θ_{exp}))	0.66	6.6×10^{-4}	0.016

Примечание: Столбец «R» содержит коэффициент корреляции между компонентами векторов \mathfrak{J} для набора $\mathfrak{S}_{\text{гфоб}}^9$ и для наборов \mathfrak{T} . Столбец «p-значение» содержит значимость коэффициента корреляции. Столбец «p-значение (попр.)» содержит p-значения с поправкой на множественное сравнение.

Оказалось интересным, что сами по себе позиции остатков из наборов $\mathfrak{S}_{\text{арг}}^9$ и $\mathfrak{S}_{\text{гфоб}}^9$ также пересекаются (см. для иллюстрации рис. 4.13 ниже). Так позиции из 20-% хвоста распределения $\mathfrak{S}_{\text{арг}}$ (см. рис. 2.1) и позиции из 20-% хвоста распределения величины SAHR (см. уравнение 2.3):

1. $\mathfrak{S}_{\text{арг}}^{20\%}$: 1–30, 36–39, 42, 45, 88, 91–92, 97–99, 101–104, 109, 130;
2. SAHR $^{20\%}$: 1, 17, 25–26, 32, 39, 42, 55, 58, 67, 68, 98, 102–103, 126, 130;
3. Пересечение: 1, 17, 25–26, 39, 42, 98, 102–103, 130.

– значимо ($p = 0.0013$) пересекаются, как видно из соответствующих параметров гипергеометрического распределения ($k = 10$, $M = 153$, $n = 16$, $N = 48$).

Этот факт может свидетельствовать в пользу того, что аминокислотные остатки, участвующие в нестандартных контактах между SOD1, чаще всего являются доступными растворителю гидрофобными остатками.

Таблица 4 — Параметры гипергеометрического распределения и уровень значимости пересечения наборов \mathfrak{T} и $\mathfrak{S}_{\text{дест}}$ с $\mathfrak{S}_{\text{мут}}$.

Набор \mathfrak{T} (тип связей (Θ или Θ_{exp}))	M	n	N	k	p-значение	p-значение (попр.)
\mathfrak{T}_{RR} (PWhb (Θ))	153	35	7	5	5.9×10^{-4}	0.018
\mathfrak{T}_{RR} (PPhb (Θ_{exp}))	153	35	35	15	5.1×10^{-4}	0.018
$\mathfrak{S}_{\text{дест}}$	153	35	27	14	4.1×10^{-5}	0.002

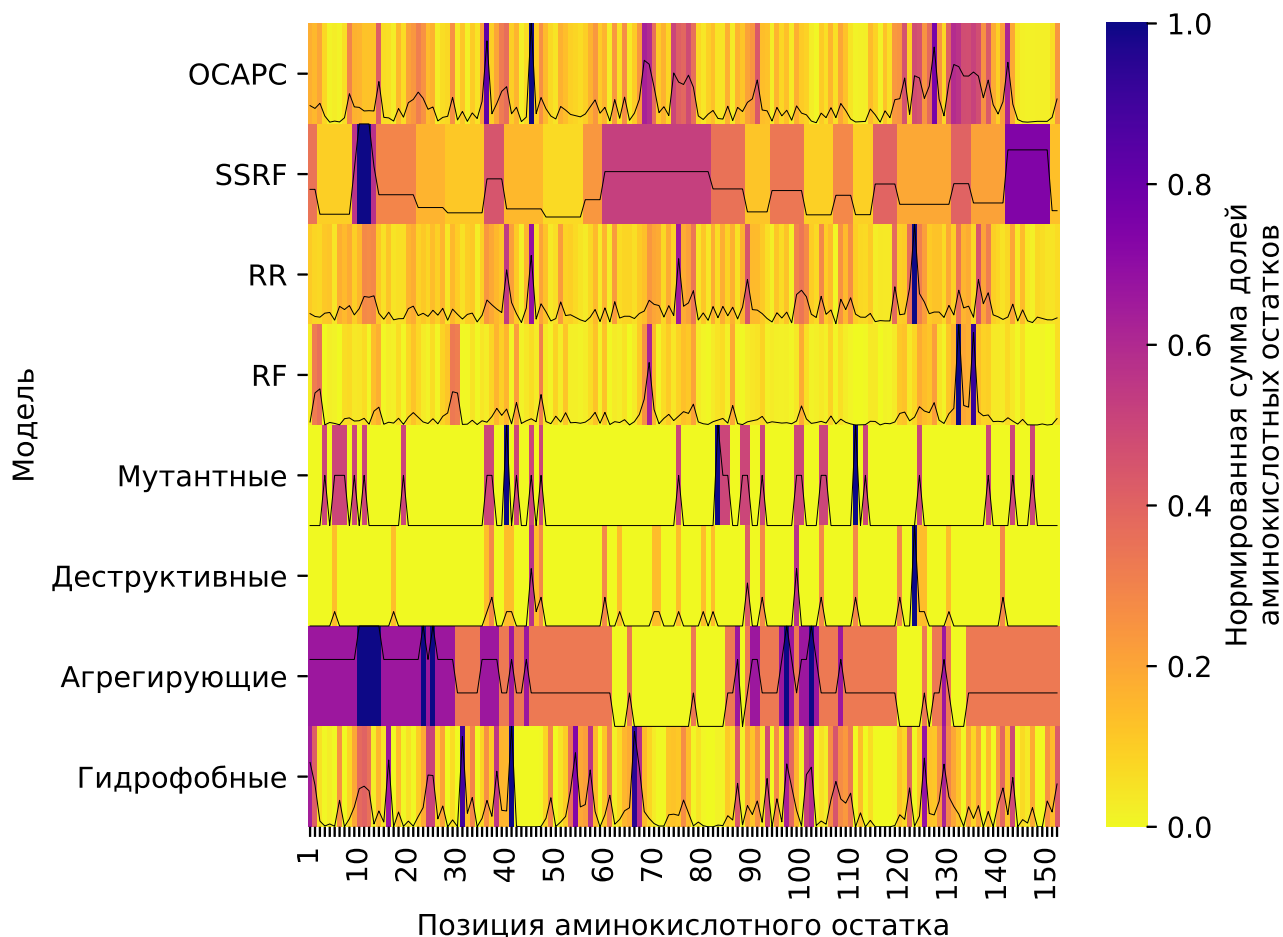
Примечание: Столбцы «M», «n», «N» и «K» соответствуют параметрам гипергеометрического распределения для сопоставления набора \mathfrak{T} из первого столбца таблицы с набором $\mathfrak{S}_{\text{мут}}$. Столбец «p-значение» содержит значимость пересечения позиций остатков из заданной пары наборов \mathfrak{T}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

Сопоставив позиции важных аминокислотных остатков с мутантными позициями, выяснилось, что остатки \mathfrak{T}_{RR} (PWhb (Θ) и PPhb (Θ_{exp})) значимо ($p < 0.02$) пересекаются с $\mathfrak{S}_{\text{мут}}$ (см. таблицу 4). Кроме того, оказалось, что $\mathfrak{S}_{\text{дест}}$ также значимо ($p = 0.002$) пересекаются с аминокислотными остатками, в которых присутствовали мутации в SOD1.

Просуммировав долю количества важных остатков ($N(x)$, см. уравнение 4.5), по всем наборам $\mathfrak{T}_{\text{OCAPC}}$, $\mathfrak{T}_{\text{SSRF}}$, \mathfrak{T}_{RR} и \mathfrak{T}_{RF} , то есть для всех типов водородных связей и наборов наблюдаемых величин дожития пациентов (Θ и Θ_{exp}), для каждой из регрессионных моделей были получены позиции важных остатков, чья оценка важности взята из 10-% хвоста распределения данных оценок (см. также рис. 4.13 и 4.14):

1. $\mathfrak{T}_{\text{OCAPC}}$: 37, 46, 69–70, 128, 132–133, 143;
2. $\mathfrak{T}_{\text{SSRF}}$: 11–13 (петля C2), 143–151 (β -тяж E8);
3. \mathfrak{T}_{RR} : 41, 46, 76, 90, 120, 123–124, 137;
4. \mathfrak{T}_{RF} : 2–3, 30–31, 48, 70, 133, 136.

Видно, что аминокислотные остатки, имеющие наибольшую важность в регрессионных моделях, частично пересекаются по позициям: 46 (встречаются в



В качестве значений для построения тепловой карты использована величина $N(x)$ (см. уравнение 4.5), просуммированная по наборам \mathcal{T} , сгруппированным по регрессионным моделям. Графики напротив наименования моделей дублируют данные значения.

Рисунок 4.13 — Тепловая карта, отражающая расположение в первичной структуре SOD1 важных аминокислотных остатков $\mathcal{T}_{\text{OCAPC}}$, $\mathcal{T}_{\text{SSRF}}$, \mathcal{T}_{RR} , \mathcal{T}_{RF} , $\mathcal{G}_{\text{мут}}$, $\mathcal{G}_{\text{дест}}$, $\mathcal{G}_{\text{агр}}^9$ и меру SAHR в позиции каждого остатка белка.

$\mathcal{T}_{\text{OCAPC}}$ и \mathcal{T}_{RR} , 70 и 133 ($\mathcal{T}_{\text{OCAPC}}$ и \mathcal{T}_{RF}), а также 143 ($\mathcal{T}_{\text{OCAPC}}$ и $\mathcal{T}_{\text{SSRF}}$). Известно, что остаток Гис120 (из \mathcal{T}_{RR}) связывает ион меди, а Арг143 участвует в каталитической активности SOD1 [138]. Остаток Гис71, рядом с которым расположен Лиз70, участвует в связывании иона цинка, который стабилизирует структуру SOD1 [136; 216]. Также известно, что остатки 38–40, контактирующие с Гли41 (из \mathcal{T}_{RR}), входят в β -бочонок [217]. При этом Лей38 имеет ван-дер-Ваальсовый контакт с Гис43, который, в свою очередь, связан с каталитически важным остатком Арг143. Остатки Гис46 и Гис48 (\mathcal{T}_{RF}) участвует в связывании иона меди [218].



Красным цветом выделены остатки в позициях 2–3, 30–31, 48, 70, 133 и 136. Рисунок 4.14 — Пространственное расположение аминокислотных остатков, образующих важные для предсказания дожития пациентов водородные связи в белке SOD1, при обучении регрессионной модели RF.

Таким образом, заметная доля остатков, выделенных как важные при обучении регрессионных моделей, может играть также важную роль в поддержании структуры SOD1 и выполнении им каталитической функции.

4.2.3 Связь конформационных свойств мутантов SOD1 и агрегации

В литературе описаны аминокислотные остатки, которые могут участвовать в образовании нестандартных межатомных взаимодействиях с участием белка SOD1 и, тем самым, приводить к агрегации SOD1, влекущей за собой заболевание БАС (см. раздел 2.7). Так, в работе [131] описано взаимодействие между субъединицами SOD1, приводящее к формированию амилоидов. Авторы показали, что

подобное взаимодействие возможно между мутантами, лишёнными ионов металла в активном сайте («аро»-форма). Ими описано, как линейное (в виде нитей), так и спиральное (в виде микротрубочек) взаимодействия субъединиц. В случае линейных нитей, белки агрегируют через участок β -бочонка (остатками 45, 87–88, 97–99) и электростатическую петлю (остатки 125–131). В случае спиральных образований – через остатки 78–81, 101, 103 обоих мономеров.

Этими же авторами [132] при изучении мутантной формы Гис46Арг белка SOD1 выделялись три области, участвующие в образовании агрегатов: электростатическая петля (остатки с 121 по 144), дисульфидная петля (остатки 49–62) и петля, содержащая сайт связывания иона цинка (остатки 63–84). Электростатическая петля участвует в каталитической активности белка (также в [18]). Дисульфидная–связана дисульфидной связью (через остатки 57 и 146) с β -тяжем 8 (143–148) β -бочонка. Помимо этого, существуют взаимодействия и между указанными петлями через водородные связи с остатком 124. Сам по себе остаток 124 образует водородную связь с остатком 71, который связывает атом цинка, и с остатком 46, который связывает ион меди. Такая комплексная связь авторами называется «вторичным мостом» по аналогии с «первичным мостом» – остатком 63, связывающим оба металла. Другими словами, в белке существует двойная связь между элементами электростатической и цинк-связывающей петлей, которая стабилизирует сайты связывания ионов металлов. В результате моделирования обнаружено, что дестабилизация многих остатков в электростатической петле связана с изменением дожития пациентов с мутациями в SOD1.

В этой же работе [132] авторы приводят сведения о дополнительных взаимодействиях, которые возникают в связи с мутацией Гис46Арг. Спиральные образования возникают из-за контактов между димерами через остатки 11, 13 и 36. Образования в виде зигзага возникают через остатки β -бочонка и электростатической петли, описанные выше [131]. Образования в виде зигзага и линейных нитей могут контактировать друг с другом через остатки 109, 12-13, 37-39, 13-15, 36-39, 91-92. В построенных в настоящей работе регрессионных моделях учтено большинство из перечисленных остатков, как важные. В работе других авторов [133] также обнаружены возможные контакты между мутантными SOD1, образующими «зигзаг» через остатки 24, 26, 66, 86, 88, 98, 102-104, 109 и 128–131.

Известно, что β -тяжи 3 и 4 (а именно – остатки 21–53) белка SOD1 дикого типа, лишённого ионов металлов, а также мутантов А4V и G93R частично разупорядочены в условиях физиологических значений температуры [130]. Кроме того,

для лишённых ионов металлов SOD1 дикого типа и мутантов A4V, H48Q и G93R область 117-144 дестабилизирована даже при 10 °С. Авторы также указывают на две области, которые являются наиболее стабильными. Первая – сформированная остатками 7–20, вторая – остатками 104–116. Последние две области, по мнению авторов, играют роль опорных точек в укладке SOD1.

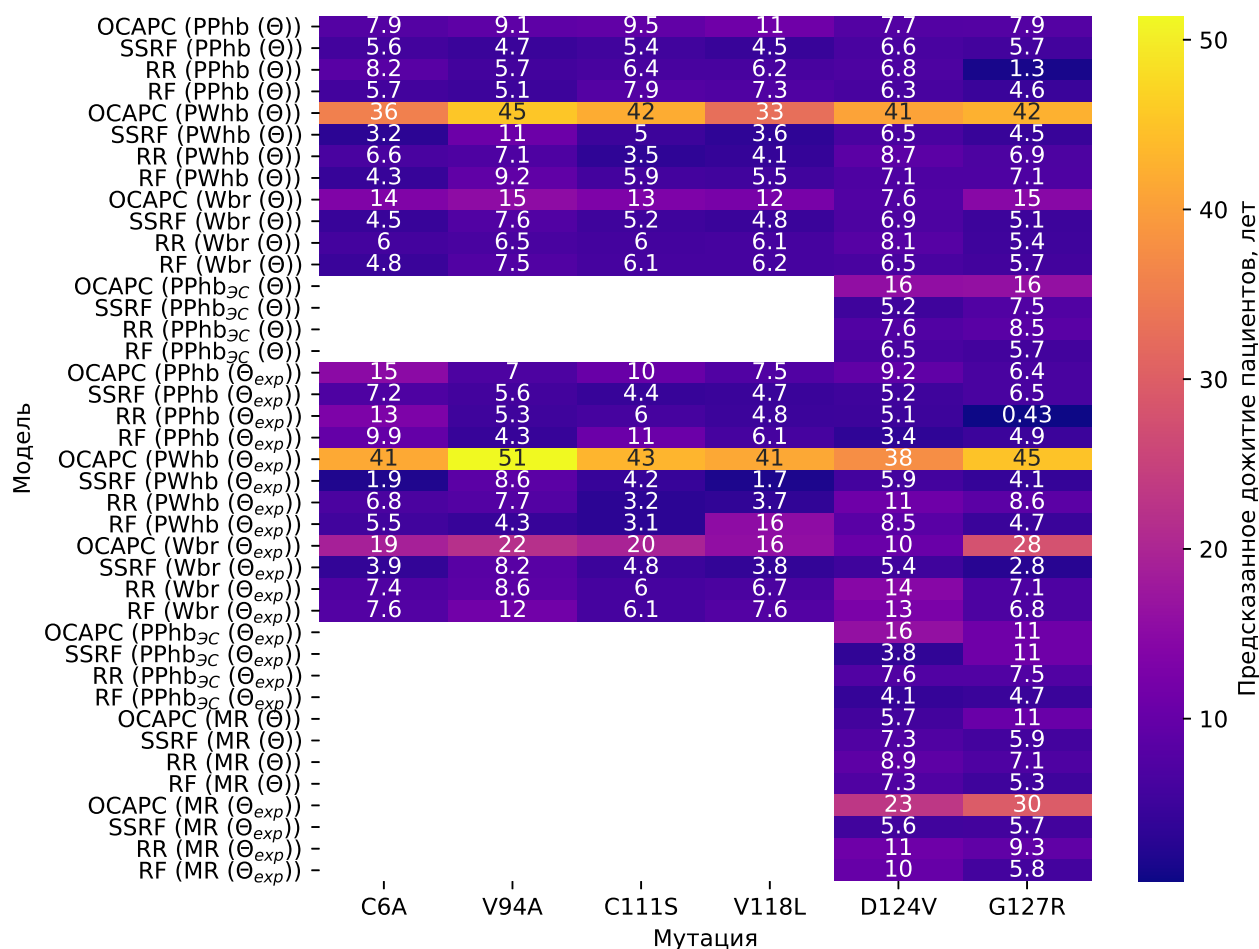
В работе [134] выделяются три района (1–30, 90–120 и 135–153), имеющих особое значение для агрегации мутантных белков SOD1. В работе [219] выделен также сайт области димеризации (Вал7-Гли147-Вал148).

Таким образом, можно видеть, что чаще всего в качестве аминокислотных остатков, участвующих в образовании агрегатов SOD1 выступают остатки в функционально важных позициях в структуре белка: остатки, образующие дисульфидную связь, влияющую на стабильность мономера, цинк связывающую и электростатическую петли, задействованные в ферментативной активности, а также непосредственно сами сайты связывания ионов металлов. При построении регрессионных моделей для предсказания дожития пациентов с мутациями в SOD1, страдающих БАС, с помощью статистических методов обнаружены водородные связи, имеющие наибольшее влияние на точность предсказаний. Позиции аминокислотных остатков в аминокислотной последовательности, участвующих в образовании данных важных водородных связей, как выяснилось при построении регрессионной модели SSRF, достоверно пересекаются с позициями остатков из контактов между агрегирующими SOD1. В это пересечение попали остатки петли C2: 11–14. Регрессионная модель RF, показавшая наибольшую точность предсказаний, опиралась на значения стабильности водородных связей, в формировании которых участвуют остатки в позициях: 2–3, 30–31, 48, 70, 133, 136 – большинство из которых отмечено, как формирующие нестандартные межатомные взаимодействия в мутантах SOD1.

4.2.4 Предсказание дожития пациентов с мутациями в SOD1, для которых сведения в литературе неполны

Построенные регрессионные модели были применены для предсказания дожития пациентов с тремя точечными мутациями Вал118Лей, Асп124Вал и Гли127Арг, данные о которых в литературе неполны. Было известно, что один

пациент – носитель мутации Вал118Лей – жил около восьми лет [220]. О пациенте с мутацией Асп124Вал известно, что он жил более двух лет [7]. Для мутации Гли127Арг в работе [221] были приведены сведения об одном пациенте, дожитие которого составило семь месяцев. Данные мутации не вносились в обучающую выборку при построении регрессий. Предсказания также были сделаны ещё для двух мутаций, в которых искусственно заменялись свободные цистеины (Цис6Ала, Цис111Сер), и для которых нет литературных данных о дожитии пациентов [222]. Эти предсказания делались в предположении, что данные две мутации в SOD1 могут вызывать БАС. Дополнительно была промоделирована ранее неизвестная мутация Вал94Ала.



Ячейки, соответствующие характеристикам водородных связей PPhbЭС, пропущены из-за того, что в рамках ЭС мутации Цис6Ала, Вал94Ала, Цис111Сер и Вал118Лей не моделировались.

Рисунок 4.15 — Тепловая карта, отражающая предсказания дожития пациентов с мутациями Цис6Ала, Вал94Ала, Цис111Сер, Вал118Лей, Асп124Вал и Гли127Арг в SOD1.

Оказалось, что регрессионные модели OSCAR дают заметно завышенные предсказания дожития для всех шести выбранных мутаций практически на всех наборах данных (см. рис. 4.15). Если исключить из рассмотрения завышенные показания OSCAR, то на основе оставшихся регрессионных моделей предсказания будут выглядеть следующим образом (лет; справа от \pm указано стандартное отклонение; см. также рис. 4.16):

1. Цис6Ала 5.86 ± 2.5
2. Вал94Ала 7.32 ± 2.08
3. Цис111Сер 5.64 ± 1.72
4. Вал118Лей 5.17 ± 2.79
5. Асп124Вал 6.86 ± 2.50
6. Гли127Арг 5.7 ± 2.16

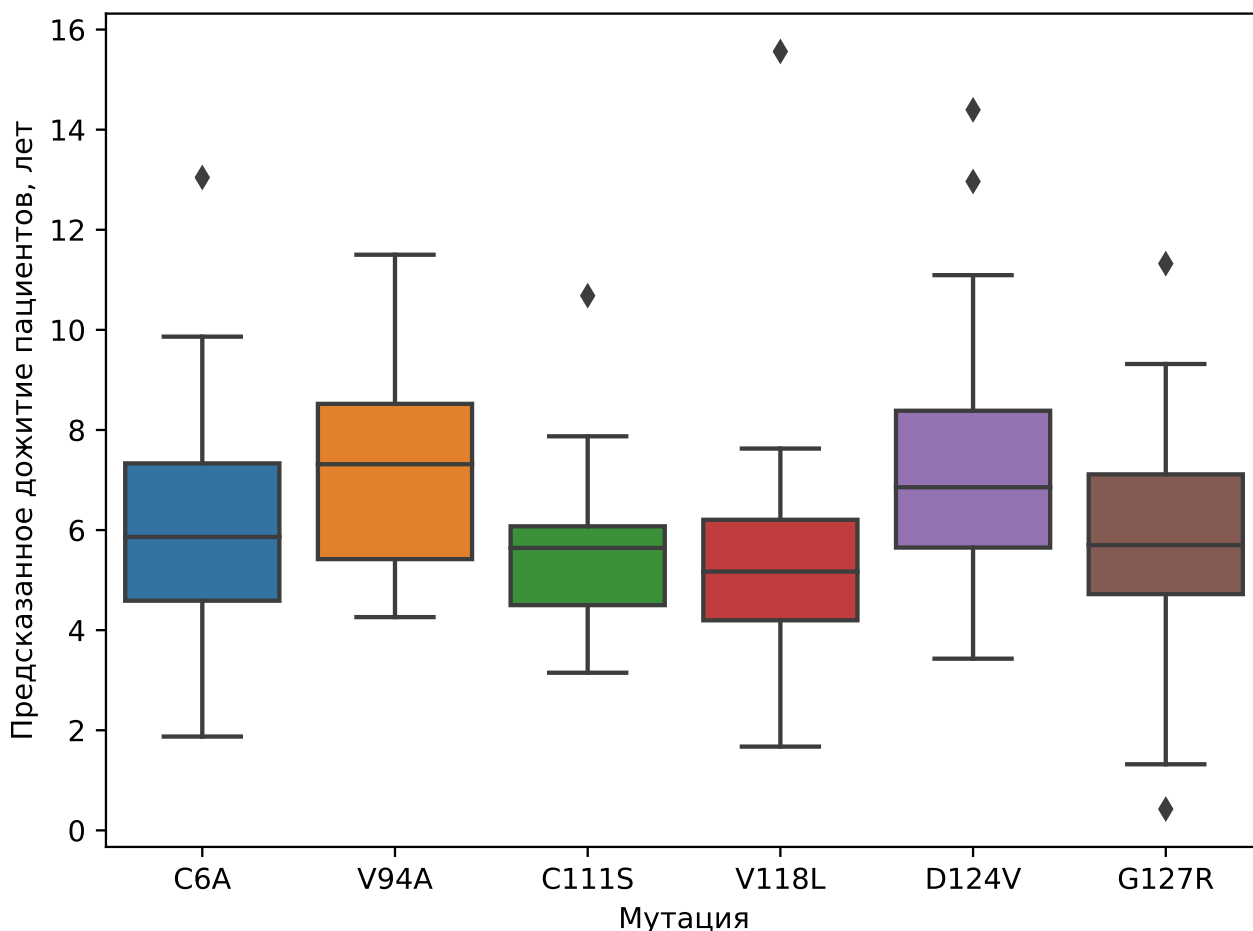


Рисунок 4.16 — Предсказания дожития пациентов с мутациями Цис6Ала, Вал94Ала, Цис111Сер, Вал118Лей, Асп124Вал и Гли127Арг в SOD1, усреднённые по регрессионным моделям SSRF, RR и RF.

Как видно из рисунков 4.15 и 4.16, предсказания дожития пациентов с БАС с мутацией Вал118Лей в SOD1 по данным моделей варьировались от 1.7 (SSRF, PWhb, Θ_{exp}) до 16 (RF на тех же данных), в среднем составляя 5.17 лет. В целом, предсказания на основе построенных регрессионных моделей согласуются с наблюдаемой величиной дожития (порядка 8 лет). Предсказания для мутации Асп124Вал варьировались от 3.8 (SSRF, PPhb_{ЭС}, Θ_{exp}) до 14 лет (RR, Wbr, Θ_{exp}), со средним равным 6.86 лет, что согласуется с наблюдениями (более двух лет). Наконец, предсказания для Гли127Арг варьировались от 0.43 или чуть более пяти месяцев (RR, PPhb, Θ_{exp}) до 9.3 (четырёхфакторная MR_{RR} на основе Θ_{exp}), имея среднее на уровне 5.69 лет. Здесь только одна регрессионная модель (RR_{PPhb}, Θ_{exp}) предсказала невысокое дожитие пациентов с данной мутацией, что могло бы считаться согласованием с наблюдаемым дожитием (около семи месяцев).

Интересным оказалось, что аминокислотные остатки Асп90 и Вал94 формируют водородные связи с Лей38, замены в котором связаны с БАС. Район белка, составленный остатками 90–94 связывает β -тяж 5 (остатки 83–89) и 6 (остатки 95–101), которые, как считается, участвуют в образовании амилоидов [133]. Известно, что при поиске функционально важных аминокислотных остатков часто используется аланиновый скрининг, при котором все аминокислотные остатки изучаемого белка поочередно заменяются на аланин [223]. Как ожидается, такие мутации в наименьшей степени нарушают структуру белка, но оказывают влияние на его функцию при заменах функционально важных аминокислот. В связи с этим, в настоящей работе была промоделирована замена на аланин валина в позиции 94, который в белке дикого типа может формировать водородные связи с остатками Асп90 и Лей38, мутации в которых ассоциированы с БАС. Дожитие пациентов, носителей мутации Вал94Ала, не найдено в литературных источниках. Однако, согласно предсказаниям, сделанным в настоящей работе, данная мутация может вызывать у пациентов заболевание БАС.

4.3 Заключение по главе 4

Разработана методика расчёта и исследования конформационных свойств мутантов белка, основанная на оценке характеристик водородных связей в этих

мутантах с использованием моделирования МД и ЭС. В рамках данной методики учитывается также расположение водородных связей в элементах вторичной структуры белка.

Как на основе исходной выборки наблюдаемых значений дожития пациентов, так и на основе выборки значений дожития, сгенерированной с помощью нормального распределения дожития с определёнными по экспериментальным данным средним и стандартным отклонением по каждой анализируемой мутации, 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 OCAPC). Однофакторные модели были отдельно построены для каждой группы водородных связей (PPhb, PPhb_{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). Всего для четырёх групп водородных связей на основе обучающих выборок Θ и Θ_{exp} построено 24 многофакторных моделей. Результаты предсказаний однофакторных и многофакторных моделей были интегрированы в рамках специальных четырёхфакторных регрессионных моделей MR, основанных на гребневой регрессии. 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}).

Для оценки точности построенных моделей применён метод bootstrap и статистический критерий Краскела-Уоллиса, проверяемый как для результатов предсказания исходных регрессионных моделей, так и для проверочных моделей (Rhb), построенных на основе случайного выбора предикторов. При построении регрессионных моделей фиксировались водородные связи (или элементы вторичной структуры), которые обладали наибольшей важностью для предсказания дожития пациентов. Наиболее точной регрессионной моделью оказалась модель RF ($R = 0.98$, $p < 10^{-22}$ при обучении на Θ , $R = 0.85$, $p < 10^{-160}$ при обучении на Θ_{exp}), построенная с помощью метода random forest, которая в ходе перекрёстной

проверки продемонстрировала наименьшую среднеквадратичную ошибку предсказаний, равную 4 года.

При построении регрессионной модели SSRF была рассчитана важность факторов – совокупная стабильность водородных связей, формирующихся с участием элементов вторичной структуры белка, – используемых для предсказания дожития пациентов с мутациями в SOD1. Позиции аминокислотных остатков в аминокислотной последовательности, участвующих в образовании данных элементов, как оказалось, достоверно пересекаются с позициями остатков из контактов между агрегирующими SOD1. С другой стороны, при обучении регрессионных моделей выяснилось, что элементы вторичной структуры, содержащие остатки SOD1, формирующие водородные связи, которые, в свою очередь, имеют наибольшую важность для предсказания дожития пациентов, часто содержат также гидрофобные остатки, легко доступные для молекул растворителя. Это может означать, что патогенные мутации в SOD1 вызывают агрегацию через такие растворимые гидрофобные остатки.

Чаще всего в образовании важных при построении регрессионных моделей водородных связей участвуют аминокислотные остатки в позициях, соответствующих остаткам, формирующим дисульфидную связь, сайт связывания иона меди, электростатическую, дисульфидную и цинк-связывающую петли. Результаты анализа, проведённого в данной главе, указывают также на то, что аминокислотные остатки, участвующие в нестандартных контактах между мутантными SOD1, в большом количестве случаев являются доступными растворителю гидрофобными остатками. Кроме того, обнаружено, что мутации в SOD1 влияют на структуру белка напрямую, дестабилизируя, в первую очередь, остатки в позициях, в которых данные мутации имелись.

С помощью построенных регрессионных моделей предсказано дожитие пациентов с мутацией Вал94Ала, в случае если они будут носителями данной мутации, которое составляет 7.3 ± 2.1 .

Глава 5. Исследование крупномасштабной динамики мутантов SOD1

В главах 3 и 4 на основе результатов моделирования МД и ЭС были исследованы конформационные свойства мутантов белка SOD1. В настоящей главе исследовались особенности крупномасштабных флуктуаций белка SOD1 дикого типа и его мутантов, полученных из моделей ЭС.

5.1 Механическая жёсткость структуры белка SOD1

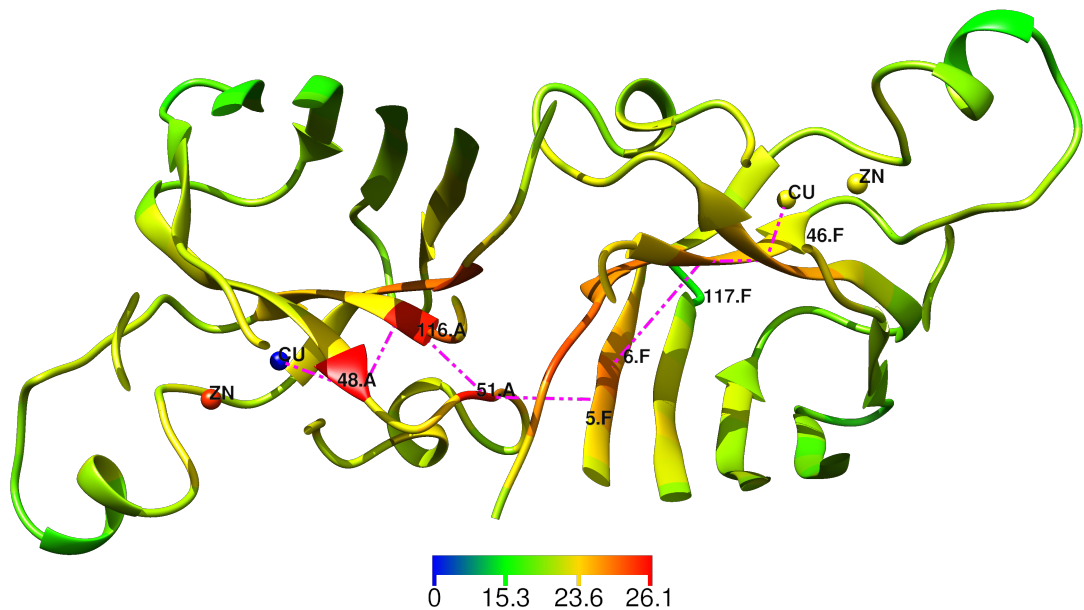
5.1.1 Позиции в белке с наибольшей средней жёсткостью

С помощью эластичной модели белка SOD1 дикого типа было промоделировано механическое воздействие внешней силы на все пары (i, j) аминокислотных остатков белка, и рассчитана соответствующая механическая жёсткость κ_{ij} . Для каждой позиции по формуле 2.1 рассчитывалась средняя жёсткость \bar{K}_i . Средняя жёсткость характеризует степень зависимости движений некоторого аминокислотного остатка от других остатков в белке или, другими словами, «механическую сопротивляемость отдельных остатков деформациям, в целом» [183]. Оказалось, что среди позиций с наибольшей средней жёсткостью (позиции были взяты из крайнего правого 7 % хвоста распределения \bar{K}_i .) находятся следующие: 45–47, 114–118, 148–150 в субъединице А, а также 45–47, 114, 117, 148–149 в субъединице F. Далее эти позиции будут обозначаться $\mathfrak{S}_{\text{мжѐст}}$. Остатки в данных позициях являются частью сайтов связывания иона меди, интерфейса между субъединицами, «greek-key» петли, а также β -тяжей 4, 7 и 8.

5.1.2 Жёсткость между активными центрами белка

Далее был проведён анализ жёсткости $\kappa_{Cu,i}$, где Cu – ион меди в пространственной структуре SOD1. Наибольшую жёсткость $\kappa_{Cu,i}$, согласно крайнему

правому 15 % хвосту распределения $\kappa_{Cu,i}$, имели позиции 48, 50–53, 115–116, 118–119, 146–148 из субъединицы А и 5–8, 45–47, 117–118, 146–151 из субъединицы F (далее эти позиции упоминаются как \mathcal{S}_{Cu-Cu}). Оказалось, что данные позиции расположились между сайтами связывания ионов меди в обеих мономерях SOD1, формируя компактный фрагмент пространственной структуры белка, который может выступать в качестве интерфейса взаимодействия между этими активными сайтами мономеров (см. рис. 5.1).



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

Рисунок 5.1 — Структура белка SOD1 с отмеченной цветом механической жёсткостью $\kappa_{Cu,i}$ между ионом меди и аминокислотными остатками белка

В данном интерфейсе 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.

Аналогичная картина наблюдалась также для жёсткости между ионом меди субъединицы F и остальными остатками в SOD1. Наибольшую жёсткость $\kappa_{Cu,i}$ имели позиции 5–7, 46–47, 117, 146–151 из субъединицы А, а также 48, 51–52, 115–116, 118–119, 146–148 из субъединицы F. Предполагаемый путь передачи механических колебаний от сайта связывания иона меди субъединицы F к сайту

связывания иона меди субъединицы А оказался зеркально отражённым относительно центра димера: Cu.F-48.F-116.F-51.F-5.A-6.A-117.A-46.A-Cu.A.

Таким образом, оказалось, что движения обоих ионов меди в SOD1 взаимно скоррелированы и, по-видимому, эти движения происходят посредством остатков, которые, в свою очередь, движутся скоординировано, как компактный фрагмент, составляющий интерфейс взаимодействия между ионами меди (см. рис. 5.1). Данный фрагмент включает остатки из интерфейса между субъединицами, который, так или иначе связан с патогенезом БАС [16; 134; 143; 148]. Кроме того, обнаруженные взаимодействия активных центров могут влиять на области SOD1, ответственные за ферментативную функцию [2; 137; 141; 142], стабильность белка [136; 137; 141; 142] и димеризацию SOD1, которая, как полагают, связана, как с агрегацией [16; 142; 143; 148], так и с потерей белком функции [224].

5.1.3 Отличие жёсткости к важным остаткам от средней жёсткости в белке

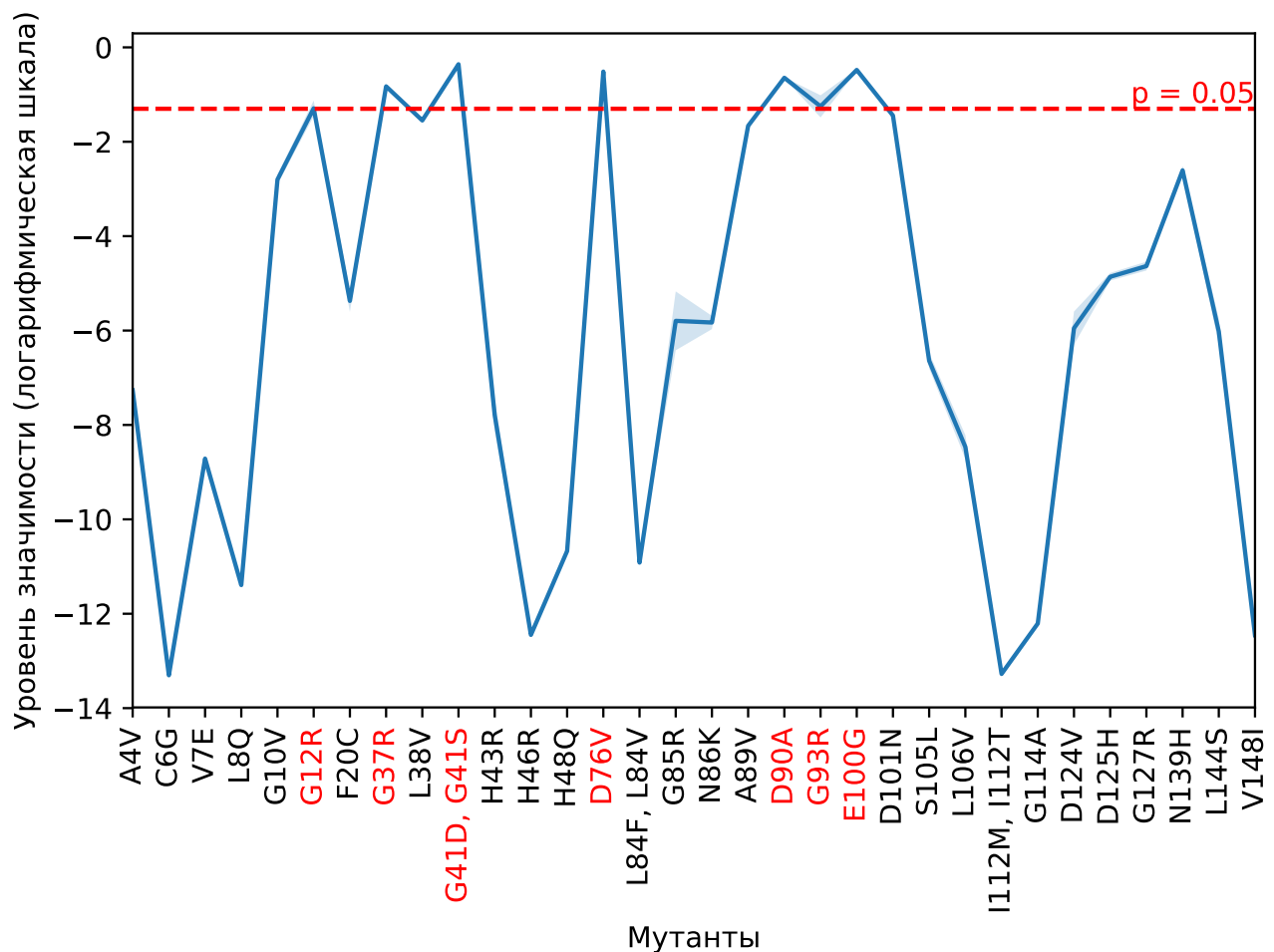
Для анализа механической жёсткости для всех пар позиций m_k , в одной из которых есть мутация, были рассчитаны следующие величины:

$$\bar{\kappa}_{m_k}^s = \frac{1}{26} \sum_{l=1}^{26} \kappa_{m_k, s_l} \quad (5.1)$$

Здесь $\bar{\kappa}_{m_k}^s$ – средние значения механической жёсткости, $m_k \in \mathfrak{S}_{\text{мут}}$ ($k \in [1, 35]$) в субъединицах А и F белка SOD1 дикого типа (PDBID: 2V0A), $s_l \in \mathfrak{S}_{\text{крит}}$ (всего 26 остатков, $l \in [1, 26]$)

В качестве важных остатков рассматривались остатки, содержащиеся в сайтах связывания ионов металлов, интерфейсе между субъединицами и остатки, образующие дисульфидную связь (см. разделы 1.5 и 4.1.5). С помощью одностороннего непараметрического критерия Манна-Уитни-Вилкоксона сравнивались средние величины значений $\bar{\kappa}_{m_k}^s$ и κ_{ij} – механической жёсткости между i и j позициями в белке ($i \in [1; 153], j \in [1; 153], i \neq j$).

График статистической значимости отличий этих двух наборов, рассчитанной для каждой из изучаемых мутаций SOD1, представлена на рисунке 5.2. Оказалось, что жёсткость $\bar{\kappa}_{m_k}^s$ значимо выше ($p < 0.05$) от κ_{ij} во всех позициях



Метки на оси абсцисс отмечены красным цветом для тех позиций, значимость отличия средних для которых выше уровня 0.05.

Рисунок 5.2 — Статистическая значимость отличия средних величин наборов $\bar{\kappa}_{m_k}^s$ и κ_{ij} для соответствующих 32 мутантных позиций в SOD1.

анализируемых мутаций SOD1, кроме позиций 12, 37, 41, 76, 90, 93 и 100. Известно, что остатки в позициях 37 и 76 находятся в гибких петлях, что может быть причиной сниженной жёсткости данных позиций [16; 225].

5.1.4 Связь механической жёсткости и патогенности мутаций в SOD1

Следующим шагом анализа механической жёсткости был расчёт корреляции дожития пациентов с мутациями в SOD1 с жёсткостью $\kappa_{m_k,i}$, где m_k соответствуют изучаемым мутантным позициям в белке, а i — всем позициям в

нём. Наибольший коэффициент корреляции ($R = -0.41$, $p = 0.0217$) наблюдался для позиции 46 (см. рис. 5.3).

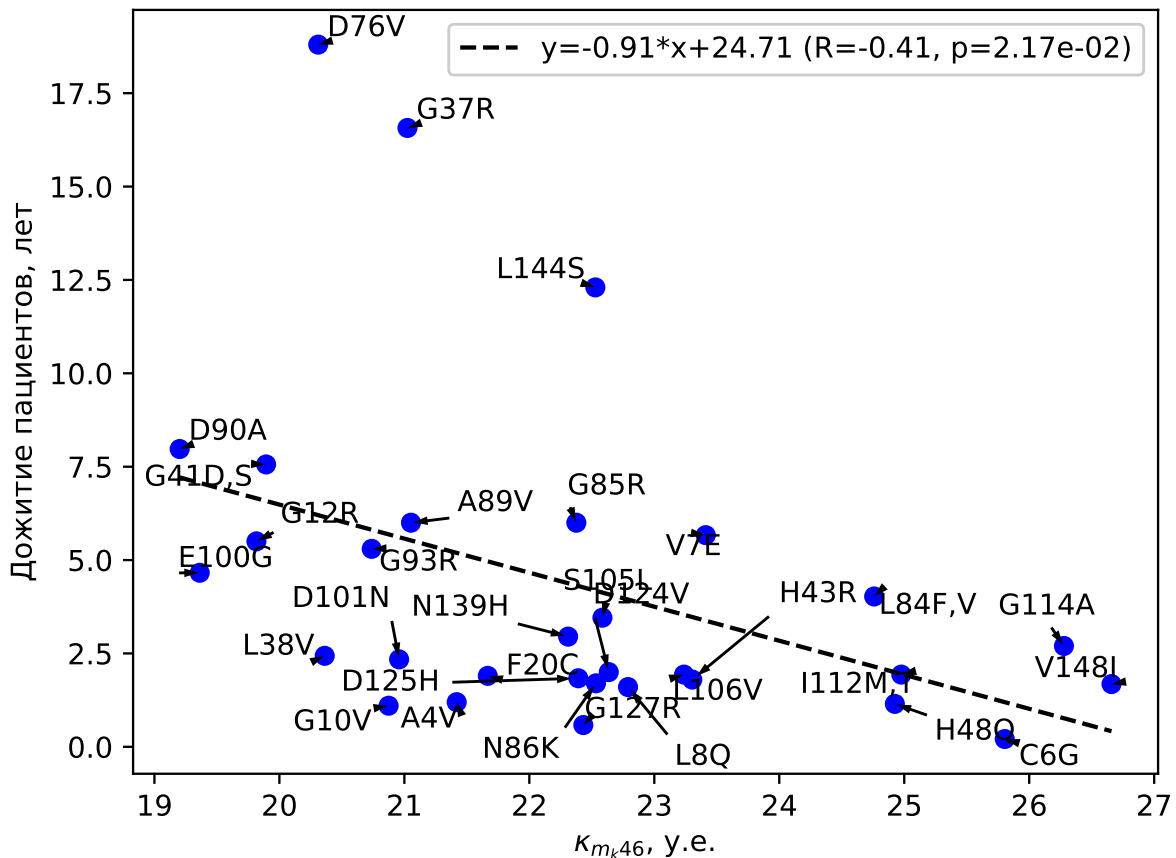


Рисунок 5.3 — График зависимости между механической жёсткостью для пар (m_k , 46) и дожитием пациентов с мутациями в SOD1 в соответствующих позициях m_k .

Стоит отметить, что остаток в позиции 46 участвует в формировании сайта связывания иона меди. Из рис. 5.3 видно, что из общей закономерности выпадают мутантные позиции 37, 76 и 144. Пациенты с мутациями в данных позициях имели более высокое дожитие, чем ожидалось из корреляционной зависимости. Интересно, что позиции 37 и 76 выделялись также при изучении механической жёсткости между мутантными позициями и набором важных для структуры SOD1 остатков (см. раздел 5.1.3).

Важность связывания иона меди в поддержании структуры SOD1 уже упоминалась в литературе [2; 137; 141; 142]. Показано, что роль иона меди в поддержании кинетической стабильности выше, чем иона цинка [142]. Известно, что в ряде патогенных мутантов SOD1 отсутствует ион меди, что также подтверждает

важность его для поддержания структуры белка [141]. Показано, что способность мутантных SOD1 удерживать ион меди связана с их агрегацией и патогенностью [137]. В недавней работе SOD1 отводится важная роль в поддержании гомеостаза меди в клетке [2]. В частности, обсуждался вопрос о том, что потеря способности связывать ионы меди мутантами данного белка ведёт к накоплению этих ионов в клетке, что является токсичным для клетки [2]. Также ранее выдвигалась гипотеза о том, что дестабилизация сайтов связывания ионов металлов, а также дестабилизация интерфейса между субъединицами является важным фактором развития заболевания БАС [53; 137].

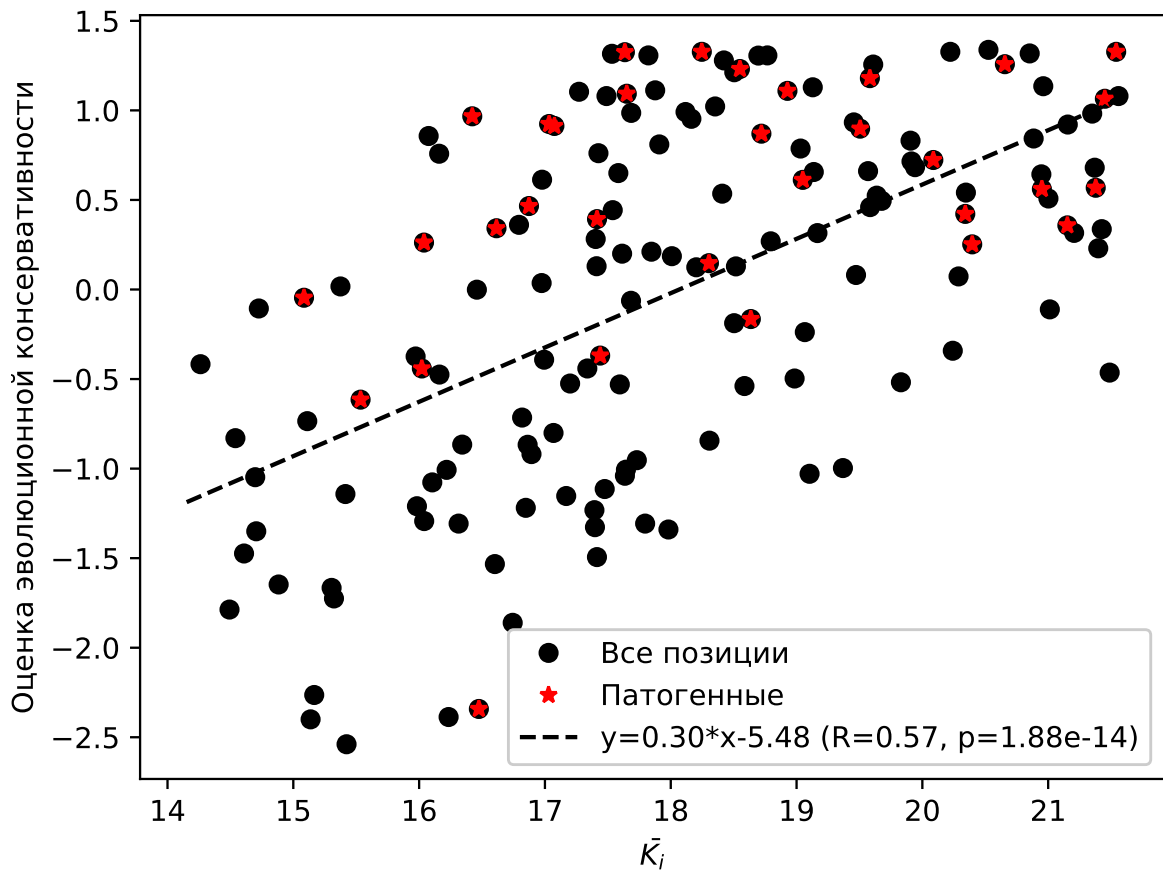
5.1.5 Зависимость между механической жёсткостью и эволюционной консервативностью аминокислотных остатков

Наряду с механической жёсткостью исследовалась также эволюционная консервативность аминокислотных остатков в SOD1 (позиции с наивысшей консервативностью обозначены как $\mathcal{S}_{\text{эконс}}$).

Была построена зависимость между эволюционной консервативностью отдельно взятого аминокислотного остатка в белке и его средней механической жёсткостью \bar{K}_i (см. рис. 5.4). Оказалось, что между упомянутыми двумя признаками наблюдается достоверная положительная корреляция ($R = 0.57$, $p = 1.88 * 10^{-14}$). Таким образом, позиции с повышенной механической жёсткостью имеют повышенную эволюционную консервативность, что указывает на их функциональную и/или структурную важность для SOD1. Аналогичное предположение для белка nesprin-1 α уже делалось ранее [226].

5.2 Важные области в структуре SOD1

Был проведён анализ эволюционной консервативности аминокислотных остатков SOD1 из наборов \mathcal{T}_{RF} , $\mathcal{S}_{\text{мут}}$, $\mathcal{S}_{\text{дест}}$, $\mathcal{S}_{\text{мжест}}$, $\mathcal{S}_{\text{Cu-Cu}}$, $\mathcal{S}_{\text{агр}}^9$ и $\mathcal{S}_{\text{гфоб}}^9$ (см. рис. 5.5).

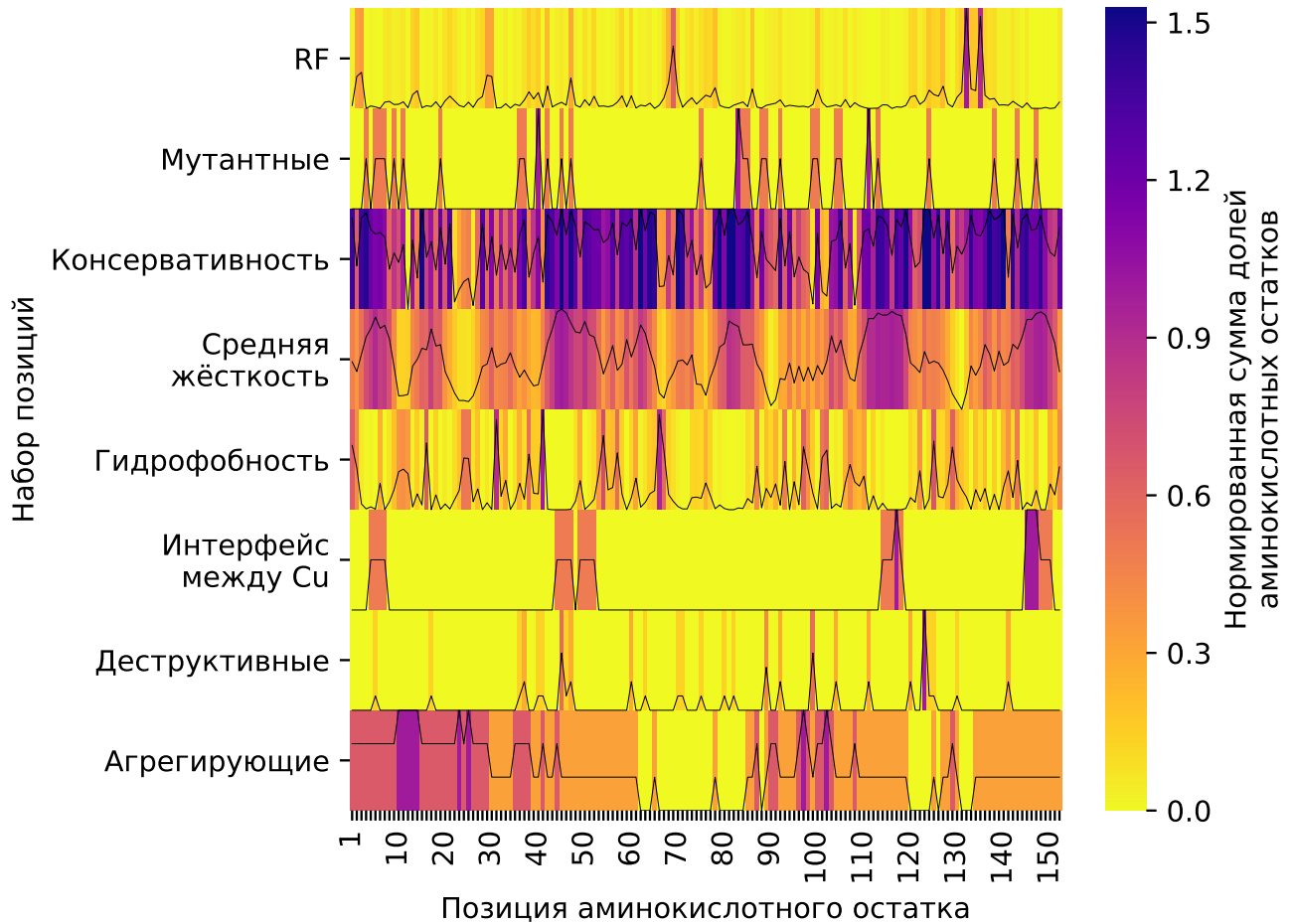


Звёздочками помечены позиции в белке, соответствующие изучаемым 35 патогенным мутантам.

Рисунок 5.4 — Зависимость между средней механической жёсткостью и эволюционной консервативностью аминокислотных остатков SOD1

Попарное пересечение позиций с наивысшими характеристиками из упомянутых наборов \mathfrak{S} с использованием гипергеометрического распределения показало, что позиции с наибольшей эволюционной консервативностью пересекались с позициями из интерфейса между активными центрами SOD1 ($p < 1.0 \times 10^{-6}$, пересечение: 45–47, 117–118, 148–150).

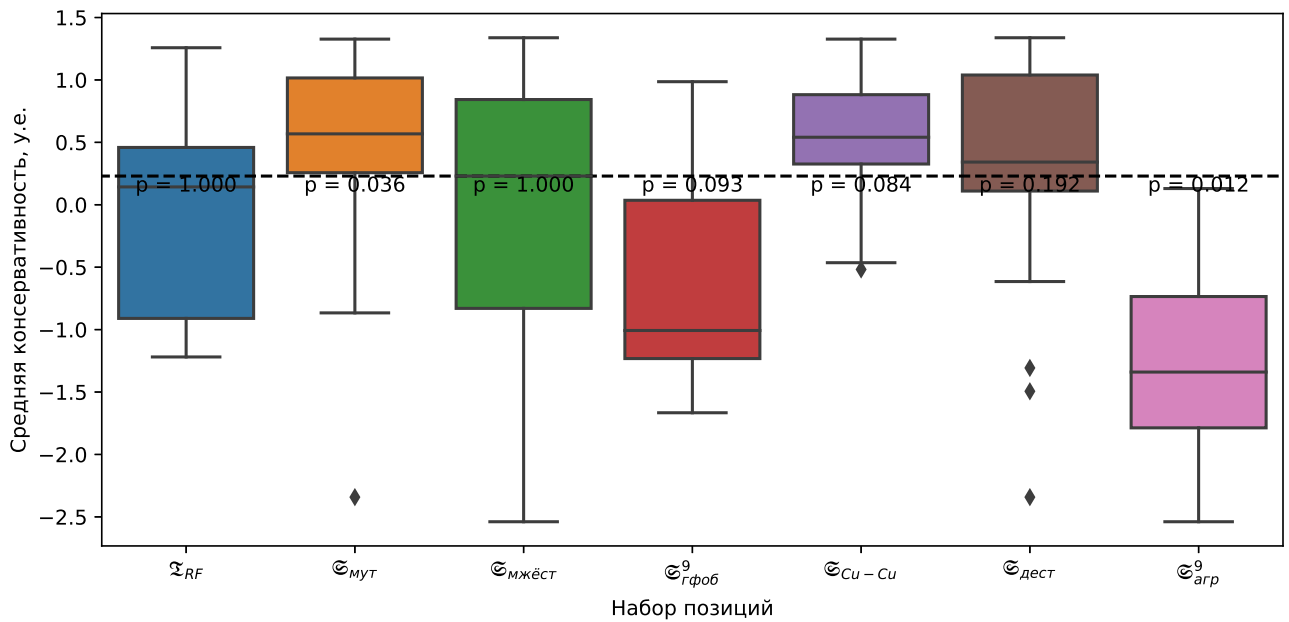
С использованием критерия Манна-Уитни-Вилкоксона показано, что средняя эволюционная консервативность остатков из наборов \mathfrak{T}_{RF} , $\mathfrak{S}_{мж\acute{e}ст}$, $\mathfrak{S}_{гфоб}^9$, \mathfrak{S}_{Cu-Cu} и $\mathfrak{S}_{дест}$ не имела значимого отличия от средней среди всех остатков белка. В то же время среднее значение эволюционной консервативности остатков $\mathfrak{S}_{мут}$ было значимо ($p < 0.05$) выше, а среднее значение эволюционной консервативности



«RF» соответствуют \mathfrak{T}_{RF} , «Мутантные» – $\mathfrak{S}_{\text{мут}}$, «Консервативность» получалась как описано в разделе 2.5, «Средняя жёсткость» – \overline{K}_i , «Гидрофобность» – мера SAHR (см. уравнение 2.3), «Интерфейс между Cu» – $\mathfrak{S}_{\text{Cu-Cu}}$, «Деструктивные» – $\mathfrak{S}_{\text{дест}}$, «Агрегирующие» – $\mathfrak{S}_{\text{агр}}^9$.

Рисунок 5.5 — Тепловая карта, отражающая расположение в первичной структуре SOD1 важных аминокислотных остатков \mathfrak{T} и \mathfrak{S} , а также их расположение относительно позиций в последовательности с наибольшей консервативностью.

остатков $\mathfrak{S}_{\text{агр}}^9$ – значимо ($p < 0.05$) ниже средней эволюционной консервативности всех остатков SOD1 (см. рис. 5.6). Таким образом, влияние мутаций в SOD1 на эволюционно-консервативные остатки может быть важным фактором БАС.



Горизонтальным пунктиром отмечена медиана оценок эволюционной консервативности всех остатков SOD1. Значения p под графиками соответствует уровню значимости отличия средних величин консервативностей соответствующего набора позиций от таковой для всех позиций по критерию Манна-Уитни-Вилкоксона.

Рисунок 5.6 — Средняя эволюционная консервативность аминокислотных остатков SOD1 из наборов \mathcal{T}_{RF} , $\mathcal{E}_{мут}$, $\mathcal{E}_{дест}$, $\mathcal{E}_{мжест}$, \mathcal{E}_{Cu-Cu} , $\mathcal{E}_{агр}^9$ и $\mathcal{E}_{Гфоб}^9$.

5.3 Заключение по главе 5

Исследована механическая жёсткость структуры SOD1. Найдены позиции, обладающие наибольшей средней жёсткостью. Остатки в данных позициях являются частью сайтов связывания иона меди, интерфейса между субъединицами, «greek-key» петли, а также β -тяжей 4, 7 и 8. Обнаружены аминокислотные остатки, образующие 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. Также оказалось, 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).

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. Можно заключить, что остатки с мутациями расположены преимущественно в эволюционно консервативных позициях, в то время как остатки из области контактов между агрегатами SOD1 не затрагивают эволюционно консервативных позиций.

Заключение

Разработана методология оценки патогенных свойств мутантных белков, способных к образованию агрегатов, на основе анализа сети водородных связей с помощью методов молекулярного моделирования (МД и ЭС). С использованием этой методологии построены регрессионные модели, связывающие конформационные свойства мутантов SOD1, носители которых проявляют признаки БАС, с дожитием таких пациентов. Показано, что дожитие пациентов с момента диагностики признаков заболевания достоверно коррелирует с конформационными особенностями мутантных белков SOD1 этих пациентов.

В рамках настоящей работы было сделано предположение, что как стабилизация, так и дестабилизация структуры белка влияет на увеличение вероятности нахождения его в метастабильном «патогенном» состоянии. Следовательно, эти факторы могут влиять и на подверженность SOD1 агрегации. В частности, можно предположить, что стабилизация структуры происходит за счет появления новых водородных связей в структуре мутанта, отсутствующих в белке дикого типа. В свою очередь, основным фактором дестабилизации структуры является разрушение водородных связей. Вновь возникшие связи могут стабилизировать белок в «патогенной» конформации, увеличивая его способность образовывать внутриклеточные агрегаты. На поверхности свободной энергии в пространстве конформационных состояний белка такие мутации могут привести к понижению локального минимума энергии, соответствующего «патогенной» конформации белка. Разрушение водородных связей в результате мутаций может приводить к снижению потенциального барьера на поверхности свободной энергии между нативным и «патогенным» состоянием белка, что повышает вероятность перехода между этими двумя состояниями.

В данной работе с использованием моделирования МД и ЭС, были рассчитаны показатели стабильности водородных связей и водных мостиков в SOD1, на основе которых были построены регрессионные модели, связывающие конформационные характеристики мутантов SOD1 с дожитием пациентов – носителей данных мутаций. Было показано, что водородные связи, важные для предсказания дожития, в основном образуются остатками из интерфейса между агрегатами

SOD1. Среди других важных факторов, которые могли влиять на конформационные свойства мутантов, можно выделить водородные связи, образуемые остатками, формирующими дисульфидную связь, а также располагающимися в сайте связывания иона меди, электростатической, дисульфидной и цинк-связывающей петлях.

Результат анализа механической жёсткости позволил сделать предположение о потенциальном интерфейсе, который может служить путём передачи механических колебаний от сайта связывания иона меди одной субъединицы к сайту связывания иона меди другой субъединицы. 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. Следует также отметить, что в интерфейс между сайтами связывания ионов металлов входили остатки белка SOD1, образующие деструктивные водородные связи – имеющие значительное расхождение по стабильности у мутантов SOD1 и белка дикого типа.

Также было показано, что существует достоверная положительная корреляция ($R = 0.57$, $p = 2 \times 10^{-14}$) между средней механической жёсткостью позиций в белке и показателем 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). В число позиций с повышенной консервативностью вошли позиции остатков, образующих важные для предсказания дожития пациентов водородные связи, формирующих деструктивные водородные связи, обладающих наивысшей средней механической жёсткостью, а также располагающихся в интерфейсе между активными центрами мономеров SOD1. Кроме того, преимущественно консервативными оказались позиции, в которых имелись мутации.

Таким образом, можно предположить, что изменение стабильности водородных связей в мутантах SOD1 может являться одним из важнейших механизмов, ведущих к конформационным изменениям мутантов SOD1, в результате которых может происходить нарушение их стабильности и образование агрегатов. Показанная хорошая корреляция между стабильности водородных связей и

дожитием пациентов свидетельствует в пользу того, что изменение стабильности водородных связей можно рассматривать как фактор патогенности мутантов и тяжести заболевания 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.

В заключение автор выражает благодарность и большую признательность научному руководителю Иванисенко В.А. за поддержку, помощь, обсуждение результатов и научное руководство. Также автор благодарит Колчанова Н.А. и Фомина Э.С., как первых научных руководителей, которые направляли автора в ходе обучения в аспирантуре. Отдельная признательность выражается Иванисенко Н.В. за предоставленную возможность использовать пакет программ AMBER, параметры силового поля для сайта связывания иона цинка в модели белка SOD1 и участие в устном обсуждении результатов. Автор благодарит:

- коллег Закияна С.М. и Медведева С.П. за введение в проблему изучения бокового амиотрофического склероза и мутаций в белке SOD1,
- Лашина С.А. и Афонникова Д.А. за ценные рекомендации относительно подходов к написанию и оформлению диссертации,
- отдельно Афонникова Д.А. за идею о доступных растворителю гидрофобных остатках,
- Подкольного Н.Л. и Ефимова В.М. за помощь и ценные советы в области статистической обработки данных,
- отдельно Ефимова В.М. и Деменкова П.С. за замечания, комментарии и рекомендации относительно текста диссертации,
- рецензентов на этапе подготовки работы,
- авторов шаблона *Russian-Phd-LaTeX-Dissertation-Template* за помощь в оформлении диссертации.

Автор благодарит также тех многих людей, кто помог сделать настоящую работу возможной.

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 А) и остатки Вал5, Цис6, Лей117, Гис46 (субъединица F).

Список сокращений и условных обозначений

БАС	боковой амиотрофический склероз
ДНК	дезоксирибонуклеиновая кислота
МД	молекулярная динамика (метод)
ОСАРС	оптимальная сумма абсолютных разностей стабильности [водородных связей] (регрессионная модель)
РНК	рибонуклеиновая кислота
РНП	рибонуклеопротеид
ЭС	эластичные сети (метод)
PPhb	группа внутрибелковых водородных связей
PWhb	группа водородных связей между белком и молекулами воды
Wbr	группа водородных связей – водных мостиков
PPhb_{ЭС}	группа внутрибелковых водородных связей, полученная в результате моделирования методом ЭС
RF	семейство множественных регрессионных моделей на основе метода random forest
Rhb	семейство множественных регрессионных моделей со случайным выбором предиктора
RR	семейство множественных регрессионных моделей на основе гребневой регрессии
SSRF	семейство множественных регрессионных моделей на основе сумм стабильностей водородных связей, сгруппированных по элементам вторичной структуры SOD1
MR	семейство четырёхфакторных регрессионных моделей на основе гребневой регрессии, факторами для которых выступают величины дожития, предсказанные с помощью одной из регрессионных моделей (ОСАРС, RF, RR или SSRF), построенных на базе стабильностей четырёх групп водородных связей (PPhb, PWhb, Wbr или PPhb _{ЭС}).
SAHR	доступные растворителю гидрофобные остатки (с англ. solvent accessible hydrophobic residues)
Σ	общее обозначение позиций аминокислотных остатков в последовательности SOD1, образующих водородные связи, имеющие наибольшую важность при построении регрессионных моделей
$\Sigma_{\text{ОСАРС}}$	Σ , важные при построении модели на основе метода ОСАРС

$\mathcal{T}_{\text{SSRF}}$	\mathcal{T} , важные при построении модели SSRF
\mathcal{T}_{RR}	\mathcal{T} , важные при построении модели RR
\mathcal{T}_{RF}	\mathcal{T} , важные при построении модели RF
$\widehat{\mathcal{T}}$	сумма частот встречаемости \mathcal{T} для всех четырёх регрессионных моделей выше
\mathcal{S}	общее обозначение особых позиций аминокислотных остатков в последовательности SOD1
$\mathcal{S}_{\text{мут}}$	\mathcal{S} , в которых имелись изучаемые мутации в SOD1, связанные с БАС
$\mathcal{S}_{\text{дест}}$	\mathcal{S} с остатками, образующими деструктивные водородные связи
$\mathcal{S}_{\text{мжѐст}}$	\mathcal{S} с остатками с наибольшей средней механической жѐсткостью
$\mathcal{S}_{\text{Cu-Cu}}$	\mathcal{S} и интерфейса между активными центрами субъединиц
$\mathcal{S}_{\text{эконс}}$	наиболее эволюционно консервативные \mathcal{S}
$\mathcal{S}_{\text{агр}}$	\mathcal{S} , содержащие остатки, участвующие в образовании агрегатов SOD1
$\mathcal{S}_{\text{агр}}^9$	\mathcal{S} , содержащие остатки, наиболее часто упоминаемые в литературе, как участвующие в образовании агрегатов SOD1
$\mathcal{S}_{\text{гфоб}}^9$	\mathcal{S} , содержащие остатки SOD1 с наибольшей гидрофобностью и, в то же время, наибольшей доступностью молекулам растворителя
$\mathcal{S}_{\text{крит}}$	\mathcal{S} , содержащие остатки SOD1, формирующие дисульфидную связь, сайты связывания ионов и интерфейс между субъединицами

Словарь терминов

Амилоидные фибриллы – длинные, неразветвлённые структуры диаметром несколько нанометров, образованные перевитыми между собой белковыми волокнами, в которых β -слои ориентированы перпендикулярно оси фибриллы [227].

Аморфный агрегат – структура, состоящая из ассоциированных белковых молекул, образованная под действием межмолекулярных взаимодействий гидрофобных участков полипептидной цепи, экспонированных растворителю [227].

Пенетрантность – частота или вероятность проявления аллеля определенного гена у разных особей родственной группы организмов [228].

Список литературы

1. Amyotrophic lateral sclerosis / M. C. Kiernan [et al.] // *The Lancet*. — 2011. — Vol. 377, no. 9769. — P. 942–955.
2. Tokuda E., Furukawa Y. Copper Homeostasis as a Therapeutic Target in Amyotrophic Lateral Sclerosis with SOD1 Mutations // *International Journal of Molecular Sciences*. — 2016. — Vol. 17, no. 5. — P. 636.
3. Real-time analysis and direct observations of different superoxide dismutase (SOD1) molecules bindings to aggregates in temporal evolution step / S. Hong [et al.] // *Colloids and Surfaces B: Biointerfaces*. — 2013. — Vol. 101. — P. 266–271.
4. Sugaya K., Nakano I. Prognostic role of “prion-like propagation” in SOD1-linked familial ALS: an alternative view // *Frontiers in Cellular Neuroscience*. — 2014. — Vol. 8.
5. Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. / P. B. Stathopoulos [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. — 2003. — Vol. 100, no. 12. — P. 7021–6.
6. Rapid disease progression correlates with instability of mutant SOD1 in familial ALS. / T. Sato [et al.] // *Neurology*. — 2005. — Vol. 65, no. 12. — P. 1954–7.
7. Protein aggregation and protein instability govern familial amyotrophic lateral sclerosis patient survival. / Q. Wang [et al.] // *PLoS biology*. — 2008. — Vol. 6, no. 7. — e170.
8. Chiti F., Dobson C. M. Amyloid formation by globular proteins under native conditions. // *Nature chemical biology*. — 2009. — Vol. 5, no. 1. — P. 15–22.
9. Variation in aggregation propensities among ALS-associated variants of SOD1: correlation to human disease. / M. Prudencio [et al.] // *Human molecular genetics*. — 2009. — Vol. 18, no. 17. — P. 3217–26.
10. SOD1 mutations targeting surface hydrogen bonds promote amyotrophic lateral sclerosis without reducing apo-state stability / R. Byström [et al.] // *The Journal of biological chemistry*. — 2010. — Vol. 285, no. 25. — P. 19544–52.

11. Computing disease-linked SOD1 mutations: deciphering protein stability and patient-phenotype relations / V. Kumar [et al.] // *Scientific Reports*. — 2017. — Vol. 7, no. 1. — P. 4678.
12. Cornilescu G., Hu J. S., Bax A. Identification of the hydrogen bonding network in a protein by scalar couplings // *Journal of the American Chemical Society*. — 1999. — Vol. 121, no. 12. — P. 2949–2950.
13. Fernández A., Scheraga H. A. Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. // *Proceedings of the National Academy of Sciences of the United States of America*. — 2003. — Vol. 100, no. 1. — P. 113–8.
14. Vogt G., Woell S., Argos P. Protein thermal stability, hydrogen bonds, and ion pairs // *Journal of Molecular Biology*. — 1997. — Vol. 269, no. 4. — P. 631–643.
15. Xu D., Tsai C. J., Nussinov R. Hydrogen bonds and salt bridges across protein-protein interfaces // *Protein Engineering Design and Selection*. — 1997. — Vol. 10, no. 9. — P. 999–1012.
16. Khare S. D., Dokholyan N. V. Common dynamical signatures of familial amyotrophic lateral sclerosis-associated structurally diverse Cu, Zn superoxide dismutase mutants. // *Proceedings of the National Academy of Sciences of the United States of America*. — 2006. — Vol. 103, no. 9. — P. 3147–52.
17. Alder B. J., Wainwright T. E. Studies in Molecular Dynamics. I. General Method // *The Journal of Chemical Physics*. — 1959. — Vol. 31, no. 2. — P. 459.
18. Keerthana S. P., Kolandaivel P. Structural investigation on the electrostatic loop of native and mutated SOD1 and their interaction with therapeutic compounds // *RSC Adv*. — 2015. — Vol. 5, no. 44. — P. 34630–34644.
19. 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.
20. Papoian G. A., Ulander J., Wolynes P. G. Role of water mediated interactions in protein-protein recognition landscapes. // *Journal of the American Chemical Society*. — 2003. — Vol. 125, no. 30. — P. 9170–8.

21. Local water bridges and protein conformational stability / M. Petukhov [et al.] // *Protein science : a publication of the Protein Society*. — 1999. — Vol. 8, no. 10. — P. 1982–1989.
22. 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.
23. Tirion M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. // *Physical review letters*. — 1996. — Vol. 77, no. 9. — P. 1905–1908.
24. 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.
25. 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.
26. Genetic analysis and SOD1 mutation screening in Iranian amyotrophic lateral sclerosis patients. / A. Alavi [et al.] // *Neurobiology of aging*. — 2013. — Vol. 34, no. 5. — 1516.e1–8.
27. Sixteen novel mutations in the Cu/Zn superoxide dismutase gene in amyotrophic lateral sclerosis: a decade of discoveries, defects and disputes. / P. M. Andersen [et al.] // *Amyotrophic lateral sclerosis and other motor neuron disorders : official publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases*. — 2003. — Vol. 4, no. 2. — P. 62–73.
28. Nitric oxide-mediated oxidative damage and the progressive demise of motor neurons in ALS. / D. A. Drechsel [et al.] // *Neurotoxicity research*. — 2012. — Vol. 22, no. 4. — P. 251–64.
29. Astrocytes from familial and sporadic ALS patients are toxic to motor neurons. / A. M. Haidet-Phillips [et al.] // *Nature biotechnology*. — 2011. — Vol. 29, no. 9. — P. 824–8.

30. Localization of a toxic form of superoxide dismutase 1 protein to pathologically affected tissues in familial ALS. / T. E. Brotherton [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. — 2012. — Vol. 109, no. 14. — P. 5505–10.
31. Brown R. H. Amyotrophic lateral sclerosis. Insights from genetics. // *Archives of neurology*. — 1997. — Vol. 54, no. 10. — P. 1246–50.
32. Amyotrophic lateral sclerosis / M. A. van Es [et al.] // *The Lancet*. — 2017. — Vol. 390, no. 10107. — P. 2084–2098.
33. Rowland L. P. How Amyotrophic Lateral Sclerosis Got Its Name // *Archives of Neurology*. — 2001. — Vol. 58, no. 3. — P. 512–515.
34. Marangi G., Traynor B. J. Genetic causes of amyotrophic lateral sclerosis: new genetic analysis methodologies entailing new opportunities and challenges. // *Brain research*. — 2015. — Vol. 1607. — P. 75–93.
35. Clinical characteristics of familial amyotrophic lateral sclerosis with Cu/Zn superoxide dismutase gene mutations. / K. Abe [et al.] // *Journal of the neurological sciences*. — 1996. — Vol. 136, no. 1–2. — P. 108–16.
36. SOD1 gene mutations in ALS patients from British Columbia, Canada: clinical features, neurophysiology and ethical issues in management. / A. Eisen [et al.] // *Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. — 2008. — Vol. 9, no. 2. — P. 108–19.
37. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. / A. E. Renton [et al.] // *Neuron*. — 2011. — Vol. 72, no. 2. — P. 257–68.
38. Belzil V. V., Rouleau G. A. Familial ALS: less common than we think? // *Journal of neurology, neurosurgery, and psychiatry*. — 2012. — Vol. 83, no. 12. — P. 1133.
39. Taylor J. P., Brown R. H., Cleveland D. W. Decoding ALS: from genes to mechanism. // *Nature*. — 2016. — Vol. 539, no. 7628. — P. 197–206.
40. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. / E. T. Cirulli [et al.] // *Science (New York, N.Y.)* — 2015. — Vol. 347, no. 6229. — P. 1436–41.

41. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. / M. DeJesus-Hernandez [et al.] // *Neuron*. — 2011. — Vol. 72, no. 2. — P. 245–56.
42. ALSod: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics / O. Abel [et al.] // *Human Mutation*. — 2012. — Vol. 33, no. 9. — P. 1345–1351.
43. Stage at which riluzole treatment prolongs survival in patients with amyotrophic lateral sclerosis: a retrospective analysis of data from a dose-ranging study / T. Fang [et al.] // *The Lancet Neurology*. — 2018.
44. Intraneuronal protein aggregation as a trigger for inflammation and neurodegeneration in the aging brain / A. Currais [et al.] // *FASEB Journal*. — 2017. — Vol. 31, no. 1. — P. 5–10.
45. Analysis of amyotrophic lateral sclerosis as a multistep process: A population-based modelling study / A. Al-Chalabi [et al.] // *The Lancet Neurology*. — 2014. — Vol. 13, no. 11. — P. 1108–1113.
46. Miller R., Mitchell J. Riluzole for amyotrophic lateral sclerosis (ALS)/ motor neuron disease (MND) (Review) // *The Cochrane Library*. — 2012. — Vol. 1, no. 3.
47. Yoshino H., Kimura A. Investigation of the therapeutic effects of edaravone, a free radical scavenger, on amyotrophic lateral sclerosis (Phase II study) // *Amyotrophic Lateral Sclerosis*. — 2006. — Vol. 7, no. 4. — P. 247–251.
48. Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial / K. Abe [et al.] // *The Lancet Neurology*. — 2017. — Vol. 16, no. 7. — P. 505–512.
49. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. / H. Laaksovirta [et al.] // *The Lancet. Neurology*. — 2010. — Vol. 9, no. 10. — P. 978–985.
50. Lagier-Tourenne C., Cleveland D. W. Rethinking ALS: The FUS about TDP-43 // *Cell*. — 2009. — Vol. 136, no. 6. — P. 1001–1004.
51. Wild-type and mutant SOD1 share an aberrant conformation and a common pathogenic pathway in ALS. / D. A. Bosco [et al.] // *Nature neuroscience*. — 2010. — Vol. 13, no. 11. — P. 1396–403.

52. Aggregation-triggering segments of SOD1 fibril formation support a common pathway for familial and sporadic ALS. / M. I. Ivanova [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. — 2014. — Vol. 111, no. 1. — P. 197–201.
53. Familial amyotrophic lateral sclerosis-associated mutations decrease the thermal stability of distinctly metallated species of human copper/zinc superoxide dismutase. / J. A. Rodriguez [et al.] // *The Journal of biological chemistry*. — 2002. — Vol. 277, no. 18. — P. 15932–7.
54. Metal-deficient aggregates and diminished copper found in cells expressing SOD1 mutations that cause ALS. / M. W. Bourassa [et al.] // *Frontiers in aging neuroscience*. — 2014. — Vol. 6. — P. 110.
55. Ross C. A., Poirier M. A. Protein aggregation and neurodegenerative disease. // *Nature medicine*. — 2004. — Vol. 10 Suppl. — S10–7.
56. Moloney E. B., Winter F. de, Verhaagen J. ALS as a distal axonopathy: molecular mechanisms affecting neuromuscular junction stability in the presymptomatic stages of the disease // *Frontiers in Neuroscience*. — 2014. — Vol. 8. — P. 252.
57. Das A., Plotkin S. S. Mechanical probes of SOD1 predict systematic trends in metal and dimer affinity of ALS-associated mutants // *Journal of Molecular Biology*. — 2013. — Vol. 425, no. 5. — P. 850–874.
58. Aggregation and motor neuron toxicity of an ALS-linked SOD1 mutant independent from wild-type SOD1. / L. I. Bruijn [et al.] // *Science (New York, N.Y.)* — 1998. — Vol. 281, no. 5384. — P. 1851–4.
59. Conversion to the amyotrophic lateral sclerosis phenotype is associated with intermolecular linked insoluble aggregates of SOD1 in mitochondria. / H.-X. Deng [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. — 2006. — Vol. 103, no. 18. — P. 7142–7.
60. Misfolded SOD1 is not a primary component of sporadic ALS / S. Da Cruz [et al.] // *Acta Neuropathologica*. — 2017. — P. 1–15.
61. Post{-}translational Modifications Promote Formation of SOD1 Oligomers With Potential Toxicity in ALS / N. Dokholyan [et al.] // *FASEB J*. — 2015. — Vol. 29, 1_Supplement. — P. 564.1–.

62. Distinct conformers of transmissible misfolded SOD1 distinguish human SOD1-FALS from other forms of familial and sporadic ALS / J. I. Ayers [et al.] // *Acta Neuropathologica*. — 2016. — Vol. 132, no. 6. — P. 827–840.
63. Amyotrophic lateral sclerosis-associated SOD1 mutant proteins bind and aggregate with Bcl-2 in spinal cord mitochondria. / P. Pasinelli [et al.] // *Neuron*. — 2004. — Vol. 43, no. 1. — P. 19–30.
64. Genetic variation of superoxide dismutases in patients with primary open-angle glaucoma / D. Celojevic [и др.] // *Ophthalmic Genetics*. — 2014. — Т. 35, № 2. — С. 79—84.
65. High frequency of TARDBP gene mutations in italian patients with amyotrophic lateral sclerosis / L. Corrado [et al.] // *Human Mutation*. — 2009. — Vol. 30, no. 4. — P. 688–694.
66. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. / E. Kabashi [et al.] // *Nature genetics*. — 2008. — Vol. 40, no. 5. — P. 572–4.
67. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis / V. M. Van Deerlin [et al.] // *The Lancet Neurology*. — 2008. — Vol. 7, no. 5. — P. 409–416.
68. Mutations in Familial and Sporadic Amyotrophic Lateral Sclerosis / J. Sreedharan [et al.] // *Science*. — 2008. — Vol. 319, no. 2008. — P. 1668–1672.
69. Two German Kindreds With Familial Amyotrophic Lateral Sclerosis Due to TARDBP Mutations / P. Kühnlein [et al.] // *Archives of Neurology*. — 2008. — Vol. 65, no. 9. — P. 2031–2041.
70. Cassel J. A., Reitz A. B. Ubiquilin-2 (UBQLN2) binds with high affinity to the C-terminal region of TDP-43 and modulates TDP-43 levels in H4 cells: Characterization of inhibition by nucleic acids and 4-aminoquinolines // *Biochimica et Biophysica Acta - Proteins and Proteomics*. — 2013. — Vol. 1834, no. 6. — P. 964–971.
71. Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis / T. J. Kwiatkowski [et al.] // *Science*. — 2009. — Vol. 323, no. 5918. — P. 1205–1208.

72. Molecular Determinants and Genetic Modifiers of Aggregation and Toxicity for the ALS Disease Protein FUS/TLS / Z. Sun [et al.] // *PLoS Biology*. — 2011. — Vol. 9, no. 4. — e1000614.
73. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs / C. Lagier-Tourenne [et al.] // *Nature neuroscience*. — 2012. — Vol. 15, no. 11. — P. 1488–1497.
74. Stepwise acquirement of hallmark neuropathology in FUS-ALS iPSC models depends on mutation type and neuronal aging / J. Japtok [et al.] // *Neurobiology of Disease*. — 2015. — Vol. 82. — P. 420–429.
75. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation / A. Patel [et al.] // *Cell*. — 2015. — Vol. 162, no. 5. — P. 1066–1077.
76. ALS-associated mutant FUS induces selective motor neuron degeneration through toxic gain of function / A. Sharma [et al.] // *Nature Communications*. — 2016. — Vol. 7, no. 10465.
77. Coady T. H., Manley J. L. ALS mutations in TLS/FUS disrupt target gene expression. // *Genes & development*. — 2015. — Vol. 29, no. 16. — P. 1696–706.
78. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function / T. Murakami [et al.] // *Neuron*. — 2015. — Vol. 88, no. 4. — P. 678–690.
79. Motor neuron intrinsic and extrinsic mechanisms contribute to the pathogenesis of FUS-associated amyotrophic lateral sclerosis / J. Scekcic-Zahirovic [et al.] // *Acta Neuropathologica*. — 2017. — P. 1–20.
80. Mutations of optineurin in amyotrophic lateral sclerosis / H. Maruyama [et al.] // *Nature*. — 2010. — Vol. 465, no. 7295. — P. 223–226.
81. Wong Y. C., Holzbaur E. L. F. Optineurin is an autophagy receptor for damaged mitochondria in parkin-mediated mitophagy that is disrupted by an ALS-linked mutation. // *Proceedings of the National Academy of Sciences of the United States of America*. — 2014. — Vol. 111, no. 42. — E4439–48.
82. “Dominant-Negative Effect” Mechanism in OPTNE478G-Linked Amyotrophic Lateral Sclerosis (S43.004) / Y. Shi [et al.] // *Neurology*. — 2016. — Vol. 86, 16 Supplement.

83. Slowicka K., Vereecke L., Loo G. van Cellular Functions of Optineurin in Health and Disease // Trends in Immunology. — 2016. — Vol. 37, no. 9. — P. 621–633.
84. Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS / J. O. Johnson [et al.] // Neuron. — 2010. — Vol. 68, no. 5. — P. 857–864.
85. Eukaryotic Stress Granules Are Cleared by Autophagy and Cdc48/VCP Function / J. R. Buchan [et al.] // Cell. — 2013. — Vol. 153, no. 7. — P. 1461–1474.
86. Алейникова Т. Л. Биохимия. — Москва : ГЭОТАР-МЕД, 2003. — С. 779.
87. Privalov P. L. Intermediate States in Protein Folding // Journal of Molecular Biology. — 1996. — Vol. 258, no. 5. — P. 707–725.
88. Рубин А. Конформационная энергия и пространственная организация биополимеров // Биофизика. — Москва : Изд-во МГУ, 2004. — Гл. IX. С. 198—254.
89. Williams R. J. P. The conformation properties of proteins in solution // Biological Reviews. — 1979. — Vol. 54, no. 4. — P. 389–437.
90. Кантор Ч., Шиммел П. Биофизическая химия. Том 1. — Москва : Мир, 1984. — С. 336.
91. Рубин А. Различные типы взаимодействий в макромолекулах // Биофизика. — Москва : Изд-во МГУ, 2004. — Гл. VIII. С. 183—197.
92. The RING 2.0 web server for high quality residue interaction networks / D. Piovosan [et al.] // Nucleic Acids Research. — 2016. — Vol. 44, W1. — W367–W374.
93. Bruce Martin R. Free energies and equilibria of peptide bond hydrolysis and formation // Biopolymers. — 1998. — Vol. 45, no. 5. — P. 351–353.
94. Radzicka A., Wolfenden R. Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases // Journal of the American Chemical Society. — 1996. — Vol. 118, no. 26. — P. 6105–6109.
95. Макшакова О. Н., Ермакова Е. А., Чачков Д. В. Теоретическое исследование влияния молекул карбоновых кислот на кислотный гидролиз пептидной связи // Вестник Казанского технологического университета. — 2013. — Т. 16, № 10. — С. 41—44.
96. Москва В. Водородная связь в органической химии // Соросовский образовательный журнал. — 1999. — № 2. — С. 58—64.

97. Bosshard H. R., Marti D. N., Jelesarov I. Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings // *Journal of Molecular Recognition*. — 2004. — Vol. 17, no. 1. — P. 1–16.
98. Коваль И. В. Химия дисульфидов // *Успехи химии*. — 1994. — Т. 63, № 9. — С. 776—792.
99. Dougherty D. A. Cation- π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp // *Science*. — 1996. — Vol. 271, no. 5246. — P. 163–168.
100. Dougherty D. A. The cation- π interaction // *Accounts of Chemical Research*. — 2013. — Vol. 46, no. 4. — P. 885–893.
101. Hunter C. A., Sanders J. K. M. The nature of π - π interactions // *Journal of the American Chemical Society*. — 1990. — Vol. 112, no. 14. — P. 5525–5534.
102. McGaughey G. B., Gagné M., Rappé A. K. π -Stacking interactions. Alive and well in proteins // *Journal of Biological Chemistry*. — 1998. — Vol. 273, no. 25. — P. 15458–15463.
103. Martinez C. R., Iverson B. L. Rethinking the term “ π -stacking” // *Chemical Science*. — 2012. — Vol. 3, no. 7. — P. 2191.
104. Sinnokrot M. O., Valeev E. F., Sherrill C. D. Estimates of the ab initio limit for π - π interactions: The benzene dimer // *Journal of the American Chemical Society*. — 2002. — Vol. 124, no. 36. — P. 10887–10893.
105. Рубин А. Динамические свойства глобулярных белков // *Биофизика*. — Москва : Изд-во МГУ, 2004. — Гл. X. С. 255—297.
106. Influence of denatured and intermediate states of folding on protein aggregation. / N. L. Fawzi [et al.] // *Protein science : a publication of the Protein Society*. — 2005. — Vol. 14, no. 4. — P. 993–1003.
107. Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations / L. Orellana [et al.] // *Nature Communications*. — 2016. — Vol. 7. — P. 12575.
108. Defining the Nature of Thermal Intermediate in 3 State Folding Proteins: Apoflavodoxin, a Study Case / R. García-Fandiño [et al.] // *PLoS Computational Biology*. — 2012. — Vol. 8, no. 8.

109. Structure of an Intermediate State in Protein Folding and Aggregation / P. Neudecker [et al.] // *Science*. — 2012. — Vol. 336, no. 6079. — P. 362–366.
110. Perutz M. F. Mechanisms of cooperativity and allosteric regulation in proteins // *Quarterly Reviews of Biophysics*. — 1989. — Vol. 22, no. 02. — P. 139–237.
111. Kern D., Zuiderweg E. R. The role of dynamics in allosteric regulation // *Current Opinion in Structural Biology*. — 2003. — Vol. 13, no. 6. — P. 748–757.
112. Dokholyan N. V. Controlling Allosteric Networks in Proteins // *Chemical Reviews*. — 2016. — Vol. 116, no. 11. — P. 6463–6487.
113. Meyer E. E., Rosenberg K. J., Israelachvili J. Recent progress in understanding hydrophobic interactions // *Proceedings of the National Academy of Sciences*. — 2006. — Vol. 103, no. 43. — P. 15739–15746.
114. Contribution of hydrophobic interactions to protein stability. / J. T. Kellis [et al.] // *Nature*. — 1988.
115. Contribution of hydrophobic interactions to protein stability / C. N. Pace [et al.] // *Journal of Molecular Biology*. — 2011.
116. Wickner S., Maurizi M. R., Gottesman S. Posttranslational quality control: Folding, refolding, and degrading proteins. — 1999.
117. Fink A. L. Protein aggregation: Folding aggregates, inclusion bodies and amyloid. — 1998.
118. Dobson C. M. Protein folding and misfolding // *Nature*. — 2003. — Vol. 426, no. 6968. — P. 884–890.
119. Phase Separation: Linking Cellular Compartmentalization to Disease / A. Aguzzi [et al.] // *Trends in Cell Biology*. — 2016. — Vol. 26, no. 7. — P. 547–558.
120. Kopito R. R. Aggresomes, inclusion bodies and protein aggregation. — 2000.
121. Hipp M. S., Park S.-H., Hartl F. U. Proteostasis impairment in protein-misfolding and -aggregation diseases // *Trends in Cell Biology*. — 2014. — Vol. 24, no. 9. — P. 506–514.
122. Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution / R. van der Lee [et al.] // *Cell Reports*. — 2014. — Vol. 8, no. 6. — P. 1832–1844.

123. Madan Babu M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease // *Biochemical Society Transactions*. — 2016. — Vol. 44, no. 5. — P. 1185–1200.
124. Gsponer J., Babu M. M. Cellular Strategies for Regulating Functional and Non-functional Protein Aggregation // *Cell Reports*. — 2012. — Vol. 2, no. 5. — P. 1425–1437.
125. Protein aggregation in amyotrophic lateral sclerosis. / A. M. Blokhuis [et al.] // *Acta neuropathologica*. — 2013. — Vol. 125, no. 6. — P. 777–94.
126. Bence N. F. Impairment of the Ubiquitin-Proteasome System by Protein Aggregation // *Science*. — 2001. — Vol. 292, no. 5521. — P. 1552–1555.
127. Evolution of CuZn superoxide dismutase and the Greek Key β -barrel structural motif / E. D. Getzoff [et al.] // *Proteins: Structure, Function, and Genetics*. — 1989. — Vol. 5, no. 4. — P. 322–336.
128. An Efficient Null Model for Conformational Fluctuations in Proteins / T. Harder [et al.] // *Structure*. — 2012. — Vol. 20, no. 6. — P. 1028–1039.
129. Structural and biophysical properties of metal-free pathogenic SOD1 mutants A4V and G93A / A. Galaledeen [et al.] // *Archives of Biochemistry and Biophysics*. — 2009. — Vol. 492, no. 1–2. — P. 40–47.
130. Metal-free superoxide dismutase-1 and three different amyotrophic lateral sclerosis variants share a similar partially unfolded beta-barrel at physiological temperature. / A. Durazo [et al.] // *The Journal of biological chemistry*. — 2009. — Vol. 284, no. 49. — P. 34382–9.
131. Amyloid-like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS. / J. S. Elam [et al.] // *Nature structural biology*. — 2003. — Vol. 10, no. 6. — P. 461–7.
132. Structural consequences of the familial amyotrophic lateral sclerosis SOD1 mutant His46Arg. / S. Antonyuk [et al.] // *Protein science : a publication of the Protein Society*. — 2005. — Vol. 14, no. 5. — P. 1201–13.
133. Structural and dynamic aspects related to oligomerization of apo SOD1 and its mutants. / L. Banci [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. — 2009. — Vol. 106, no. 17. — P. 6980–5.

134. Mutation-dependent polymorphism of Cu,Zn-superoxide dismutase aggregates in the familial form of amyotrophic lateral sclerosis. / Y. Furukawa [et al.] // *The Journal of biological chemistry*. — 2010. — Vol. 285, no. 29. — P. 22221–31.
135. Structural similarity of wild-type and ALS-mutant superoxide dismutase-1 fibrils using limited proteolysis and atomic force microscopy. / P. K. Chan [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. — 2013. — Vol. 110, no. 27. — P. 10934–9.
136. Ding F., Dokholyan N. V. Dynamical roles of metal ions and the disulfide bond in Cu, Zn superoxide dismutase folding and aggregation. // *Proceedings of the National Academy of Sciences of the United States of America*. — 2008. — Vol. 105, no. 50. — P. 19696–701.
137. Aggregation propensities of superoxide dismutase G93 hotspot mutants mirror ALS clinical phenotypes. / A. J. Pratt [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. — 2014. — Vol. 111, no. 43. — E4568–76.
138. Comparative structural and conformational studies on H43R and W32F mutants of copper-zinc superoxide dismutase by molecular dynamics simulation. / G. Muneeswaran [et al.] // *Biophysical chemistry*. — 2014. — Vol. 185. — P. 70–8.
139. Disulfide bond mediates aggregation, toxicity, and ubiquitylation of familial amyotrophic lateral sclerosis-linked mutant SOD1. / J.-i. Niwa [et al.] // *The Journal of biological chemistry*. — 2007. — Vol. 282, no. 38. — P. 28087–95.
140. Mutant SOD1 causes motor neuron disease independent of copper chaperone-mediated copper loading / J. R. Subramaniam [et al.] // *Nature Neuroscience*. — 2002. — Vol. 5, no. 4. — P. 301–307.
141. Ip P., Mulligan V. K., Chakrabartty A. ALS-Causing SOD1 Mutations Promote Production of Copper-Deficient Misfolded Species // *Journal of Molecular Biology*. — 2011. — Vol. 409, no. 5. — P. 839–852.
142. Lynch S. M., Colón W. Dominant role of copper in the kinetic stability of Cu/Zn superoxide dismutase // *Biochemical and Biophysical Research Communications*. — 2006. — Vol. 340, no. 2. — P. 457–461.

143. Amyloidogenic Regions and Interaction Surfaces Overlap in Globular Proteins Related to Conformational Diseases / V. Castillo [et al.] // PLoS Computational Biology. — 2009. — Vol. 5, no. 8. — e1000476.
144. Complete loss of post-translational modifications triggers fibrillar aggregation of SOD1 in the familial form of amyotrophic lateral sclerosis. / Y. Furukawa [et al.] // The Journal of biological chemistry. — 2008. — Vol. 283, no. 35. — P. 24167–76.
145. Destabilization of apoprotein is insufficient to explain Cu,Zn-superoxide dismutase-linked ALS pathogenesis. / J. A. Rodriguez [et al.] // Proceedings of the National Academy of Sciences of the United States of America. — 2005. — Vol. 102, no. 30. — P. 10516–21.
146. Decreased metallation and activity in subsets of mutant superoxide dismutases associated with familial amyotrophic lateral sclerosis / L. J. Hayward [et al.] // Journal of Biological Chemistry. — 2002. — Vol. 277, no. 18. — P. 15923–15931.
147. SOD1 misplacing and mitochondrial dysfunction in amyotrophic lateral sclerosis pathogenesis. / F. Tafuri [et al.] // Frontiers in cellular neuroscience. — 2015. — Vol. 9, August. — P. 336.
148. Unveiling the unfolding pathway of FALS associated G37R SOD1 mutant: a computational study. / D. Milardi [et al.] // Molecular bioSystems. — 2010. — Vol. 6, no. 6. — P. 1032–9.
149. Tildesley D. J., Allen M. P. Computer simulation of liquids // Clarendon, Oxford. — New York, 1987. — P. 110–140.
150. Coarse-Grained Protein Models and Their Applications / S. Kmiecik [et al.] // Chemical Reviews. — 2016. — Vol. 116, no. 14. — P. 7898–7936.
151. Suhre K., Sanejouand Y.-H. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. // Nucleic acids research. — 2004. — Vol. 32, Web Server issue. — W610–4.
152. Yang L., Song G., Jernigan R. L. How Well Can We Understand Large-Scale Protein Motions Using Normal Modes of Elastic Network Models? // Biophysical Journal. — 2007. — Vol. 93, no. 3. — P. 920–929.

153. Frappier V., Najmanovich R. J. A Coarse-Grained Elastic Network Atom Contact Model and Its Use in the Simulation of Protein Dynamics and the Prediction of the Effect of Mutations // *PLoS Computational Biology*. — 2014. — Vol. 10, no. 4. — e1003569.
154. Dynamics of radiation damage / J. B. Gibson [et al.] // *Physical Review*. — 1960. — Vol. 120, no. 4. — P. 1229–1253.
155. Zhou R. H., Berne B. J. Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? // *Proceedings of the National Academy of Sciences of the United States of America*. — 2002. — Vol. 99, no. 20. — P. 12777–12782.
156. Nymeyer H., García A. E. Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized Born approximation with explicit solvent. // *Proceedings of the National Academy of Sciences of the United States of America*. — 2003. — Vol. 100, no. 24. — P. 13934–9.
157. Molecular dynamics simulations of the complete satellite tobacco mosaic virus / P. L. Freddolino [et al.] // *Structure*. — 2006. — Vol. 14, no. 3. — P. 437–449.
158. Molecular dynamics simulations of large macromolecular complexes / J. R. Perilla [et al.] // *Current Opinion in Structural Biology*. — 2015. — Vol. 31. — P. 64–74.
159. Шайтан К. В. Молекулярная динамика биополимеров // Проблемы регуляции в биологических системах/Под общей ред. АБ Рубина.—М.-Ижевск: НИЦ { \guillemotleft }Регулярная и хаотическая динамика. — 2006.
160. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules / W. D. Cornell [et al.] // *Journal of the American Chemical Society*. — 1995. — Vol. 117, no. 19. — P. 5179–5197.
161. Kollman P. A. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules // *Accounts of Chemical Research*. — 1996. — Vol. 29, no. 10. — P. 461–469.
162. Wang J., Cieplak P., Kollman P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? // *Journal of Computational Chemistry*. — 2000. — Vol. 21, no. 12. — P. 1049–1074.

163. Comparison of multiple amber force fields and development of improved protein backbone parameters / V. Hornak [et al.]. — 2006.
164. Improved side-chain torsion potentials for the Amber ff99SB protein force field / K. Lindorff-Larsen [et al.] // *Proteins: Structure, Function and Bioinformatics*. — 2010. — Vol. 78, no. 8. — P. 1950–1958.
165. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations / Y. Duan [et al.] // *Journal of Computational Chemistry*. — 2003. — Vol. 24, no. 16. — P. 1999–2012.
166. Garcia A. E., Sanbonmatsu K. Y. α -Helical stabilization by side chain shielding of backbone hydrogen bonds // *Proceedings of the National Academy of Sciences*. — 2002. — Vol. 99, no. 5. — P. 2782–2787.
167. Mackerell A. D., Feig M., Brooks C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation // *Journal of Computational Chemistry*. — 2004. — Vol. 25, no. 11. — P. 1400–1415.
168. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins / a. D. MacKerell [et al.] // *The Journal of Physical Chemistry B*. — 1998. — Vol. 102, no. 18. — P. 3586–3616.
169. Feller S. E., MacKerell A. D. J. An improved empirical potential energy function for molecular simulations of phospholipids // *Journal of Physical Chemistry B*. — 2000. — Vol. 104, no. 31. — P. 7510–7515.
170. Foloppe N., MacKerell A. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data // *J. Comput. Chem.* — 2000. — Vol. 21, no. 2. — P. 86–104.
171. Biomolecular Simulation: The GROMOS96 Manual and User Guide / F. W. Van Gunsteren [et al.]. — 1996. — P. 1–1042.
172. Jorgensen W. L., Maxwell D. S., Tiradorives J. Development and testing of the OPLS all atom force field on conformational energetics and properties of organic liquids // *J. Am. Chem. Soc.* — 1996. — Vol. 118. — P. 11225–11236.

173. Hagler A. T., Huler E., Lifson S. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals // *Journal of the American Chemical Society*. — 1974. — Vol. 96, no. 17. — P. 5319–5327.
174. Rudd R. E., Broughton J. Q. Coarse-grained molecular dynamics and the atomic limit of finite elements // *Physical Review B*. — 1998. — Vol. 58, no. 10. — R5893–R5896.
175. Fuglebakk E., Reuter N., Hinsen K. Evaluation of Protein Elastic Network Models Based on an Analysis of Collective Motions // *Journal of Chemical Theory and Computation*. — 2013. — Vol. 9, no. 12. — P. 5618–5628.
176. Kurkcuoglu Z., Doruker P. Ligand Docking to Intermediate and Close-To-Bound Conformers Generated by an Elastic Network Model Based Algorithm for Highly Flexible Proteins // *PLOS ONE*. — 2016. — Vol. 11, no. 6. — P. 1–22.
177. Haliloglu T., Bahar I. Adaptability of protein structures to enable functional interactions and evolutionary implications // *Current Opinion in Structural Biology*. — 2015. — Vol. 35. — P. 17–23.
178. Bastolla U. Computing protein dynamics from protein structure with elastic network models. — 2014.
179. Case D. A. Normal mode analysis of protein dynamics // *Current Opinion in Structural Biology*. — 1994. — Vol. 4, no. 2. — P. 285–290.
180. Brüschweiler R. Collective protein dynamics and nuclear spin relaxation // *The Journal of Chemical Physics*. — 1995. — Vol. 102, no. 8. — P. 3396–3403.
181. Ahmed A., Gohlke H. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory // *Proteins: Structure, Function, and Bioinformatics*. — 2006. — Vol. 63, no. 4. — P. 1038–1051.
182. Bakan A., Meireles L. M., Bahar I. ProDy: Protein Dynamics Inferred from Theory and Experiments // *Bioinformatics*. — 2011. — Vol. 27, no. 11. — P. 1575–1577.
183. Eyal E., Bahar I. Toward a Molecular Understanding of the Anisotropic Response of Proteins to External Forces: Insights from Elastic Network Models // *Biophysical Journal*. — 2008. — Vol. 94, no. 9. — P. 3424–3435.

184. Sarkar R. Native flexibility of structurally homologous proteins: insights from anisotropic network model // *BMC Biophysics*. — 2017. — Vol. 10, no. 1. — P. 1.
185. Mechanical characterization of amyloid fibrils using coarse-grained normal mode analysis / G. Yoon [et al.] // *Advanced Functional Materials*. — 2011. — Vol. 21, no. 18. — P. 3454–3463.
186. Anisotropic elastic network modeling of entire microtubules / M. A. Deriu [et al.] // *Biophysical Journal*. — 2010. — Vol. 99, no. 7. — P. 2190–2199.
187. Myosin-V as a mechanical sensor: An elastic network study / M. Düttmann [et al.] // *Biophysical Journal*. — 2012. — Vol. 102, no. 3. — P. 542–551.
188. Sacquin-Mora S., Delalande O., Baaden M. Functional modes and residue flexibility control the anisotropic response of guanylate kinase to mechanical stress // *Biophysical Journal*. — 2010. — Vol. 99, no. 10. — P. 3412–3419.
189. Gibbs Energy of Superoxide Dismutase Heterodimerization Accounts for Variable Survival in Amyotrophic Lateral Sclerosis / Y. Shi [et al.] // *Journal of the American Chemical Society*. — 2016. — Vol. 138, no. 16. — P. 5351–5362.
190. Salomon-Ferrer R., Case D. A., Walker R. C. An overview of the Amber biomolecular simulation package // *Wiley Interdisciplinary Reviews: Computational Molecular Science*. — 2013. — Vol. 3, no. 2. — P. 198–210.
191. Comparison of simple potential functions for simulating liquid water / W. L. Jorgensen [et al.] // *The Journal of Chemical Physics*. — 1983. — Vol. 79, no. 2. — P. 926.
192. Zhao Y., Truhlar D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other function // *Theoretical Chemistry Accounts*. — 2007. — Vol. 120, no. 1–3. — P. 215–241.
193. Gaussian 09 Revision A.01 / M. J. Frisch [et al.]. — 2009.
194. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF) / M. B. Peters [et al.] // *Journal of Chemical Theory and Computation*. — 2010. — Vol. 6, no. 9. — P. 2935–2947.

195. Distinctive features of the D101N and D101G variants of superoxide dismutase 1; two mutations that produce rapidly progressing motor neuron disease. / J. Ayers [et al.] // *Journal of neurochemistry*. — 2014. — Vol. 128, no. 2. — P. 305–14.
196. Guerois R., Nielsen J. E., Serrano L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations // *Journal of Molecular Biology*. — 2002. — Vol. 320, no. 2. — P. 369–387.
197. Scikit-learn: Machine Learning in Python / F. Pedregosa [et al.] // *Journal of Machine Learning Research*. — 2011. — Vol. 12, Oct. — P. 2825–2830.
198. McKinney W. Data Structures for Statistical Computing in Python // *Proceedings of the 9th Python in Science Conference*. Vol. 1697900. — 2010. — P. 51–56.
199. Hunter J. D. Matplotlib: A 2D graphics environment // *Computing in Science and Engineering*. — 2007. — Vol. 9, no. 3. — P. 99–104.
200. Comaniciu D., Meer P. Mean shift: A robust approach toward feature space analysis // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2002. — Vol. 24, no. 5. — P. 603–619.
201. Benjamini Y., Yekutieli D. The control of the false discovery rate in multiple testing under dependency // *Annals of Statistics*. — 2001. — Vol. 29, no. 4. — P. 1165–1188.
202. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules / H. Ashkenazy [et al.] // *Nucleic Acids Research*. — 2016. — Vol. 44, W1. — W344–W350.
203. UCSF Chimera—a visualization system for exploratory research and analysis. / E. F. Pettersen [et al.] // *Journal of computational chemistry*. — 2004. — Vol. 25, no. 13. — P. 1605–12.
204. Shrake A., Rupley J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin // *Journal of Molecular Biology*. — 1973. — T. 79, № 2. — C. 351–371.
205. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories / R. T. McGibbon [et al.] // *Biophysical Journal*. — 2015. — Vol. 109, no. 8. — P. 1528–1532.

206. Maximum allowed solvent accessibilities of residues in proteins. / M. Z. Tien [и др.] // PLoS one. — 2013. — Т. 8, № 11. — e80635.
207. Mann H. B., Whitney D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other // The Annals of Mathematical Statistics. — 1947. — Vol. 18, no. 1. — P. 50–60.
208. Wilcoxon F. Individual comparisons by ranking methods // Biometrics bulletin. — 1945. — Vol. 1, no. 6. — P. 80–83.
209. G41S SOD1 mutation: A common ancestor for six ALS Italian families with an aggressive phenotype / S. Battistini [et al.] // Amyotrophic Lateral Sclerosis. — 2010. — Vol. 11, no. 1–2. — P. 210–215.
210. Valentine J. S., Doucette P. A., Zittin Potter S. Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis // Annual Review of Biochemistry. — 2005. — Vol. 74, no. 1. — P. 563–593.
211. Keerthana S. P., Kolandaivel P. Study of mutation and misfolding of Cu-Zn SOD1 protein. // Journal of biomolecular structure & dynamics. — 2015. — Vol. 33, no. 1. — P. 167–83.
212. Kabsch W., Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features // Biopolymers. — 1983. — Vol. 22, no. 12. — P. 2577–2637.
213. Breiman L. Random Forests // Machine Learning. — 2001. — Vol. 45, no. 1. — P. 5–32.
214. Hoerl A. E., Kennard R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems // Technometrics. — 1970. — Vol. 12, no. 1. — P. 55–67.
215. Predicting the spread of aquatic invaders: insight from 200 years of invasion by zebra mussels / A. Y. Karatayev [et al.] // Ecological Applications. — 2015. — Vol. 25, no. 2. — P. 430–440.
216. The unusually stable quaternary structure of human Cu,Zn-superoxide dismutase 1 is controlled by both metal occupancy and disulfide status. / F. Arnesano [et al.] // The Journal of biological chemistry. — 2004. — Vol. 279, no. 46. — P. 47998–8003.

217. Two novel SOD1 mutations in patients with familial amyotrophic lateral sclerosis. / H. X. Deng [et al.] // Human molecular genetics. — 1995. — Vol. 4, no. 6. — P. 1113–6.
218. Variable metallation of human superoxide dismutase: atomic resolution crystal structures of Cu-Zn, Zn-Zn and as-isolated wild-type enzymes. / R. W. Strange [et al.] // Journal of molecular biology. — 2006. — Vol. 356, no. 5. — P. 1152–62.
219. Ligand binding and aggregation of pathogenic SOD1. / G. S. A. Wright [et al.] // Nature communications. — 2013. — Vol. 4. — P. 1758.
220. Autonomic failure in ALS with a novel SOD1 gene mutation / T. Shimizu [et al.] // Neurology. — 2000. — Vol. 54, no. 7. — P. 1534–1537.
221. G127R: A novel SOD1 mutation associated with rapidly evolving ALS and severe pain syndrome. / T. Holmøy [et al.] // Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases. — 2010. — Vol. 11, no. 5. — P. 478–80.
222. Lepock J. R., Frey H. E., Hallewell R. A. Contribution of conformational stability and reversibility of unfolding to the increased thermostability of human and bovine superoxide dismutase mutated at free cysteines. // J. Biol. Chem. — 1990. — Vol. 265, no. 35. — P. 21612–21618.
223. Cunningham B., Wells J. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis // Science. — 1989. — Vol. 244, no. 4908. — P. 1081–1085.
224. Is SOD1 loss of function involved in amyotrophic lateral sclerosis? / R. A. Saccon [et al.]. — 2013.
225. A novel exon 3 mutation (D76V) in the SOD1 gene associated with slowly progressive ALS. / T. Segovia-Silvestre [et al.] // Amyotrophic lateral sclerosis and other motor neuron disorders : official publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases. — 2002. — Vol. 3, no. 2. — P. 69–74.
226. Stabilization of the spectrin-like domains of nesprin-1?? by the evolutionarily conserved ”adaptive” domain / Z. Zhong [et al.] // Cellular and Molecular Bioengineering. — 2010. — Vol. 3, no. 2. — P. 139–150.

227. Туроверов К. К., Уверский В. Н., Кузнецова И. М. Нативные глобулярные и нативные частично или полностью неупорядоченные белки. Фолдинг, образование надмолекулярных комплексов, агрегация // Цитология. — 2009. — Т. 51, № 3. — С. 190.
228. Timofeeff-Ressovsky N. Studies on the phenotypic manifestation of hereditary factors. I. On the phenotypic manifestation of the genovariation radius incompletus in *Drosophila funebris* // Genetics. — 1927. — Vol. 12. — P. 128–198.

Список рисунков

1.1	Пространственная структура белка SOD1.	26
1.2	Пример амилоидных волокон (а) линейного вида (б) в виде «зигзага». По рис. 2 и 3 из работы [131].	28
1.3	Пример агрегата в виде спирали. По рис. 4 из работы [131].	28
1.4	Области в первичной структуре белка, соответствующие пептидам, устойчивым к действию протеазы. По рис. 1 из работы [134].	30
2.1	Зависимость между позицией аминокислотного остатка в последовательности SOD1 и количеством упоминаний остатка в этой позиции в качестве участвующего в образовании нестандартных контактов между SOD1 в литературе.	44
2.2	Графики величин $SASA_n$, H_n и SAHR для аминокислотных остатков белка SOD1.	45
3.1	Стабильность водородных связей между атомами белка.	49
3.2	Стабильность водородных связей между атомами белка и атомами молекул воды.	49
3.3	Стабильность водных мостиков.	50
3.4	Зависимость RMSD пространственной структуры мутантов SOD1 и белка дикого типа в ходе моделирования МД.	50
3.5	Диаграмма Венна для пересечения множеств контактов через водородные связи PPhb, Wbr и PPhb _{ЭС} между аминокислотными остатками донора и акцептора.	53
3.6	Средняя дистанция в белковой последовательности между позициями аминокислотных остатков-доноров и акцепторов из контактов через водородные связи PPhb, Wbr и PPhb _{ЭС}	54
3.7	Стабильность водородных связей PPhb, Wbr и PPhb _{ЭС} , уникальных для каждой группы связей и общих для всех групп связей.	55
3.8	Распределения вкладов ξ_j^m для водородных связей PPhb (а), PWhb (б), Wbr (в) и PPhb _{ЭС} (г).	57
3.9	Семь водородных связей PPhb (а), PWhb (б), Wbr (в) и PPhb _{ЭС} (г) с наивысшими вкладами ξ_j^m	58

3.10	Кластеризация мутантов SOD1 по стабильности всех обнаруженных водородных связей PPhb (а), PWhb (б), Wbr (в) и PPhb _{ЭС} (г).	59
3.11	Зависимость между нормализованной термостабильностью мутантов SOD1 и их клиническими проявлениями, выраженными в форме дожития пациентов с БАС с соответствующими мутациями.	62
4.1	Коэффициент корреляции между $\Omega_{\mathcal{R}}^m$ и Θ^m для водородных связей из группы PPhb.	67
4.2	Коэффициент корреляции между $\Omega_{\mathcal{R}}^m$ и Θ^m для водородных связей из группы PWhb.	67
4.3	Коэффициент корреляции между $\Omega_{\mathcal{R}}^m$ и Θ^m для водородных связей из группы Wbr.	68
4.4	Коэффициент корреляции между $\Omega_{\mathcal{R}}^m$ и Θ^m для водородных связей из группы PPhb _{ЭС}	68
4.5	Значения коэффициентов корреляции между конформационными свойствами мутантов белка SOD1 и дожитием пациентов с БАС с соответствующими мутациями.	79
4.6	Доля предсказанных значений дожития в диапазоне наблюдавшегося среднего значения дожития (из Θ) $\pm 3\sigma$	80
4.7	Значения медианы среднеквадратичной ошибки предсказаний дожития пациентов с изучаемыми мутациями в SOD1, полученной в 100 шагах bootstrap.	81
4.8	Десятичный логарифм р-значения отличий среднего среднеквадратичной ошибки S , которую совершают регрессионные модели OSCAR, SSRF, RF и RR при предсказании дожития пациентов, от среднего ошибки проверочной модели Rhb в соответствии с односторонним критерием Манна-Уитни-Вилкоксона.	82
4.9	Зависимость между наблюдаемым дожитием пациентов и предсказанным с помощью модели RF _{PPhb_{ЭС}} , построенной на основе выборок Θ (а) и Θ_{exp} (б).	83
4.10	Тепловая карта, отражающая расположение в первичной структуре SOD1 аминокислотных остатков \mathcal{T} , $\hat{\mathcal{T}}$, \mathcal{S}_{mut} , \mathcal{S}_{dest} , \mathcal{S}_{agr}^9 , а также меру SAHR.	84
4.11	Диаграмма, отражающая элементы структуры SOD1, позиции остатков которых значимо пересекаются с остатками из \mathcal{T}	85

4.12	Стабильность водородных связей Wbr (а) и PPhb (б), сформированных остатками в позициях 11–14, в SOD1 дикого типа и его мутантах.	87
4.13	Тепловая карта, отражающая расположение в первичной структуре SOD1 важных аминокислотных остатков $\mathfrak{T}_{\text{OCAPC}}$, $\mathfrak{T}_{\text{SSRF}}$, \mathfrak{T}_{RR} , \mathfrak{T}_{RF} , $\mathfrak{S}_{\text{мут}}$, $\mathfrak{S}_{\text{дест}}$, $\mathfrak{S}_{\text{агр}}^9$ и меру SAHR в позиции каждого остатка белка.	90
4.14	Пространственное расположение аминокислотных остатков, образующих важные для предсказания дожития пациентов водородные связи в белке SOD1, при обучении регрессионной модели RF.	91
4.15	Тепловая карта, отражающая предсказания дожития пациентов с мутациями Цис6Ала, Вал94Ала, Цис111Сер, Вал118Лей, Асп124Вал и Гли127Арг в SOD1.	94
4.16	Предсказания дожития пациентов с мутациями Цис6Ала, Вал94Ала, Цис111Сер, Вал118Лей, Асп124Вал и Гли127Арг в SOD1, усреднённые по регрессионным моделям SSRF, RR и RF.	95
5.1	Структура белка SOD1 с отмеченной цветом механической жёсткостью $\kappa_{\text{Cu},i}$ между ионом меди и аминокислотными остатками белка	100
5.2	Статистическая значимость отличия средних величин наборов $\bar{\kappa}_{m_k}^s$ и κ_{ij} для соответствующих 32 мутантных позиций в SOD1.	102
5.3	График зависимости между механической жёсткостью для пар $(m_k, 46)$ и дожитием пациентов с мутациями в SOD1 в соответствующих позициях m_k	103
5.4	Зависимость между средней механической жёсткостью и эволюционной консервативностью аминокислотных остатков SOD1	105
5.5	Тепловая карта, отражающая расположение в первичной структуре SOD1 важных аминокислотных остатков \mathfrak{T} и \mathfrak{S} , а также их расположение относительно позиций в последовательности с наибольшей консервативностью.	106
5.6	Средняя эволюционная консервативность аминокислотных остатков SOD1 из наборов \mathfrak{T}_{RF} , $\mathfrak{S}_{\text{мут}}$, $\mathfrak{S}_{\text{дест}}$, $\mathfrak{S}_{\text{мжест}}$, $\mathfrak{S}_{\text{Cu-Cu}}$, $\mathfrak{S}_{\text{агр}}^9$ и $\mathfrak{S}_{\text{Гфоб}}^9$	107
A.1	Таблица, содержащая исходную информацию о дожитии пациентов с мутациями в SOD1, взятая из работы [7].	145

A.2	Продолжение таблицы на рис. A.1.	146
A.3	Продолжение таблицы на рис. A.2.	147
A.4	Продолжение таблицы на рис. A.3.	148
A.5	Продолжение таблицы на рис. A.4.	149

Список таблиц

1	Значимость отличия стабильности водородных связей в различных группах по критерию Манна-Уитни-Вилкоксона.	56
2	Параметры гипергеометрического распределения и уровень значимости пересечения наборов \mathcal{T} и $\mathcal{S}_{\text{agr}}^9$	86
3	Корреляция между компонентами векторов $\tilde{\mathcal{J}}$ для набора $\mathcal{S}_{\text{гфоб}}^9$ и для наборов \mathcal{T}	88
4	Параметры гипергеометрического распределения и уровень значимости пересечения наборов \mathcal{T} и $\mathcal{S}_{\text{дест}}$ с $\mathcal{S}_{\text{мут}}$	89
6	Мутации и дожитие пациентов, их носителей, использованные для построения регрессионных моделей.	150

Приложение А

Таблица наблюдаемых величин дожития пациентов с мутациями в SOD1

В таблице, представленной на рис. **A.1–A.5** для каждой мутации (столбец «Mutation») приведено одно или несколько средних значений дожития пациентов с данной мутацией и, в большинстве случаев, стандартное отклонение дожития от среднего, а также в скобках указано количество пациентов (столбец «Duration \pm SD in Years (number of patients)»). В столбце «Average Duration» для каждой мутации приведено среднее значение дожития по всем пациентам с данной мутацией, столбец «Number of Patients» содержит суммарное количество пациентов с данной мутацией.

Entries	Mutation	Reference	Duration \pm SD in Years (number of patients)	Average Duration	Number of Patients
1	A4S	[156]	>3 (1)	>3	0
2	A4T	[157]	1.2 (9)	1.5	21
		[157]	4.3 (1) ^a		
		[67]	0.75 (1)		
		[158]	1.5 \pm 0.4 (7)		
		[156]	1.7 (1) ^a		
		[159]	0.83 (1)		
		[159]	1.7 (1)		
3	A4V	[67]	1.4 \pm 0.9 (84)	1.2	205
		[158]	1.0 \pm 0.4 (75)		
		[160]	1.0 \pm 0.5 (8)		
		[161]	1.2 \pm 0.8 (24)		
		[162]	0.9 (4)		
		[156]	1.5 \pm 1.1 (3)		
		[163]	0.7 (2)		
		[164]	0.8 \pm 0.2 (2) ^a		
		[165]	1.5 \pm 0.7 (3)		
		—	—	—	—
4	C6G	[166]	0.2 (2)	0.2	2
5	C6F	[167]	1.0 (1)	1.0	1
		[167]	>0.5 (1)		
6	V7E	[67]	6.0 \pm 2 (2)	5.7	3
		[168]	5.0 (1) ^a		
		[168]	>4 (1)		
7	L8Q	[169]	1.6 \pm 0.7 (5)	1.6	5
8	G10V	[170]	1.1 \pm 0.5 (5) ^a	1.1	5
9	G12A	[171]	>4 (1)	>4	0
10	G12R	[172]	5.5 \pm 1.5 (2)	5.5	2
		[173]	>4(1)		
11	V14G	[156]	1.7 (1)	1.7	1
12	F20C	[174]	1.9 \pm 1.2 (3)	2.0	7
		[174]	>2 \pm 0 (4)		
		[174]	>1 (1)		
13	E21G	[175]	19.3 (3) ^b	17.2	5
		[176]	3.0 (1)		
		[176]	>25 (1)		
14	G37R	[67]	18.7 \pm 11.4 (8)	17.0	27
		[158]	18.0 \pm 0 (3)		
		[177]	9.2 \pm 3.4 (3) ^a		
		[175]	6.7 (3) ^b		
		[178]	24.5 (5)		
		[156]	6.0 (1)		
		[179]	17.0 (3) ^b		
		[178]	>29 (1)		
15	L38R	[175]	>2 (3)	>2	0
16	L38V	[67]	2.8 \pm 1.9 (12)	2.4	22

Рисунок А.1 — Таблица, содержащая исходную информацию о дожитии пациентов с мутациями в SOD1, взятая из работы [7].

Entries	Mutation	Reference	Duration \pm SD in Years (number of patients)	Average Duration	Number of Patients
		[180]	2.0 \pm 1.6 (10)		
		—	—	—	—
17	G41D	[67]	17.0 \pm 6.3 (7)	14.1	15
		[67]	11.6 \pm 1.7 (8)		
		[156]	> 11 (1)		
		[165]	> 14 (1)		
18	G41S	[67]	0.9 \pm 0.5 (4)	1.0	16
		[181]	1.0 \pm 0.1 (8)		
		[172]	1.1 \pm 0.1 (4)		
19	H43R	[67]	2.8 \pm 1.5 (4)	1.8	12
		[158]	1.4 \pm 0.8 (7)		
		[182]	0.6 (1)		
		—	—	—	—
20	F45C	[173]	> 6 (1)	> 6	0
21	H46R	[183]	> 17.2 \pm 8.1 (6)	17.6	49
		[183]	22.0 (1)		
		[184]	> 13.4 (7)		
		[184]	22.0 (1)		
		[184]	47.0 (1)		
		[67]	17.0 \pm 11.0 (5)		
		[158]	17.4 \pm 6.4 (5)		
		[182]	17.0 \pm 7.3 (7)		
		[185]	17.3 \pm 10.7 (4)		
		[185]	16.8 \pm 6.8 (9)		
		[156]	> 11(1)		
		[156]	5.0 (1)		
		[176]	11.5 \pm 10.6 (4) ^a		
		[186]	19.0 (11)		
22	V47F	—	—	—	—
23	H48Q	[187]	1.2 \pm 0.3 (3)	1.2	4
		[188]	1.0 (1)		
24	N65S	[164]	> 14 (1)	> 14	0
25	G72C	[189]	4.4 (1)	4.4	1
26	G72S	[190]	3.3 \pm 0.7 (2)	3.3	2
27	D76Y	[156]	17.0 (1)	17.0	1
		—	—	—	—
28	D76V	[191]	18.8 \pm 7.9 (4)	18.8	4
		[191]	> 2 (1)		
29	L84F	[192]	4.9 \pm 2.7 (12)	5.8	18
		[193]	5.0 (1)		
		[192]	>8.0 \pm 2 (2)		
		[194]	>8.0 \pm 2.6 (3)		
		[192]	> 3 (1)		
		[193]	> 2.1 (1)		
30	L84V	[195]	4.7 \pm 5.8 (5)	3.2	10
		[185]	1.6 \pm 0.5 (5)		

Рисунок А.2 — Продолжение таблицы на рис. А.1.

Entries	Mutation	Reference	Duration \pm SD in Years (number of patients)	Average Duration	Number of Patients
31	G85R	[158]	6.0 \pm 4.5 (11)	6.0	11
32	N86K	[196]	1.7 \pm 0.6 (7)	1.7	7
33	N86S	[156]	7.5 \pm 1.5 (2)	6.8	4
		[197]	6.0 \pm 3 (2) ^a		
34	A89V	[198]	>3.3 \pm 1.5 (3)	>3.8	0
		[156]	>4.5 \pm 1.5 (2)		
35	D90A	[162]	2.5 \pm 0.5 (2)	8.0	15
		[164]	3.6 \pm 1.2 (2)		
		[175]	2.7 \pm 0.6 (2)		
		[176]	25.0 (1)		
		[199]	14.0 (1)		
		[103]	1.0 \pm 0 (2)		
		[176]	>10.5 \pm 3.5 (2)		
		[199]	>15.5 \pm 9.5 (2)		
		[103]	>9 (1)		
		[172]	>0.8 (1)		
		[103]	>3 (1)		
36	D90V	[200]	2.7 \pm 0.6 (3)	2.7	3
37	G93A	[67]	2.2 \pm 1.5 (9)	3.1	16
		[158]	2.4 \pm 1.4 (6)		
		[201]	15.0 (1)		
38	G93D	[67]	10.5 \pm 5.5 (3)	8.8	7
		[67]	5.7 \pm 4.5 (3)		
		[175]	>2.5 (3)^b		
		[173]	>13 (1)		
39	G93C	[67]	10.1 \pm 6.2 (7)	12.1	27
		[175]	>6.7 (3)^b		
		[180]	12.8 \pm 17 (20)		
40	G93R	[188]	5.3 \pm 4.57 (4)	5.3	4
41	G93S	[202]	8.5 \pm 1.5 (2) ^a	8.0	11
		[203]	6.9 \pm 2.0 (7)		
		[203]	>11.5 \pm 3.5 (2)		
		[203]	>5.3 \pm 2.1 (4)		
42	G93V	[204]	6.0 \pm 1 (2)	6.0	2
43	E100G	[67]	4.0 \pm 2.3 (22)	4.7	50
		[67]	5.2 \pm 0.4 (2)		
		[158]	5.1 \pm 3.3 (23)		
		[205]	3.3 (1)		
		[188]	8.0 (1)		
		[188]	>6 (1)		
44	D101N	[206]	2.4 \pm 0.9 (12)	2.3	17
		[207]	2.2 \pm 0.4 (5)		
45	D101G	[208]	1.4 (2)	1.9	3
		[188]	3.0 (1)		
46	D101H	[209]	1.2 \pm 0.8 (2)	1.2	2
47	D101Y	[210]	0.9 (1)	0.9	1

Рисунок А.3 — Продолжение таблицы на рис. А.2.

Entries	Mutation	Reference	Duration \pm SD in Years (number of patients)	Average Duration	Number of Patients
48	I104F	[211]	21.3 \pm 11.8 (3)	21.3	3
49	S105L	[212]	3.2 \pm 1.6 (3)	3.5	7
		[212]	> 2 (1)		
		[178]	3.2 (3) ^b		
		[156]	5.0 (1)		
50	L106V	[67]	2.3 \pm 1.3 (4)	1.9	6
		[158]	1.2 \pm 0.1 (2)		
		—	—	—	—
51	G108V	[213]	4.0 \pm 1 (2)	4.0	2
52	I112T	[67]	0.9 \pm 0.1 (2)	0.9	2
53	I112M	[179]	2.8 (1–4.5) (6)	3.0	9
		[178]	3.4 (3) ^b		
54	I113F	[156]	> 6(1)	> 6	0
55	I113T	[67]	3.5 \pm 2.8 (7)	4.3	38
		[67]	2–20 (3)		
		[158]	4.2 \pm 2.5 (9)		
		[214]	3.0 (1)		
		[215]	4.7 \pm 3.3 (2) ^a		
		[156]	1.6 (1)		
		[216]	1.6 \pm 10 (4)		
		[188]	5.0 \pm 7.5 (6)		
		[217]	1.1 (1)		
		[165]	3.0 \pm 1.7 (3)		
		[165]	>9.3 \pm 0.8 (2)		
		—	—	—	—
56	G114A	[156]	2.4 (1)	2.7	2
		[165]	3.0 (1)		
57	R115G	[156]	> 1.8 (1)	2.5	2
		[219]	2.5 \pm 0.7 (2)		
58	D124V	[204]	> 2 (1)	> 2	0
59	D125H	[220]	1.67 (1)	1.8	2
		[188]	2.0 (1)		
60	L126S	[221]	7.5 \pm 5 (4)	8.0	5
		[222]	>10 (1)		
		[222]	> 5 (1)		
		[156]	> 3 (1)		
61	S134N	[223]	>2 (1)	1.2	3
		[156]	0.8 (2)		
62	N139H	[224]	3.4 \pm 0.7 (3)	3.0	6
		[178]	2.5 (3) ^b		
63	A140G	[225]	10.0 (1)	10.0	1
		[225]	> 2 (1)		
64	G141E	[209]	3.5 \pm 0.5 (2)	3.5	2
		[209]	> 3 (1)		
65	L144F	[226]	4.9 \pm 2.0 (6)	11.8	15
		[158]	11.0 \pm 2.6 (3)		

Рисунок А.4 — Продолжение таблицы на рис. А.3.

Entries	Mutation	Reference	Duration \pm SD in Years (number of patients)	Average Duration	Number of Patients
		[227]	20.4 \pm 16.3 (5)		
		[226]	>13 (1)		
		[156]	>2 \pm 1 (2)		
		[172]	>2 (1)		
		[226]	>2.5 \pm 1.5 (2)		
		[227]	>2 \pm 0 (2)		
66	L144S	[67]	12.3 \pm 3.7 (2)	12.3	2
67	A145T	[67]	1.6 \pm 0.5 (2)	1.6	2
68	C146R	[228]	2.0 (1)	2.8	2
		[228]	>3.5 (1)		
69	V148G	[67]	2.3 \pm 2.2 (4)	2.1	11
		[158]	2.0 \pm 0.9 (7)		
70	V148I	[67]	1.5 (2)	1.7	5
		[211]	1.8 \pm 0.5 (3)		
71	I149T	[229]	2.5 \pm 3.9 (14) ^a	2.7	15
		[188]	>6 (1)		
72	I151T	[230]	20.0 (1)	20.0	1
		[230]	>3 (1)		

Рисунок А.5 — Продолжение таблицы на рис. А.4.

Приложение Б

Наблюдаемые величины дожития пациентов с мутациями в SOD1, выбранные для построения регрессионных моделей

Таблица 6 — Мутации и дожитие пациентов, их носителей, использованные для построения регрессионных моделей.

Мутация	Дожитие пациентов, лет	Количество пациентов
Ала4Вал	1.2 ± 0.7	205
Цис6Гли	0.2	2
Вал7Глу	5.7 ± 1.6	3
Лей8Глн	1.6 ± 0.7	5
Гли10Вал	1.1 ± 0.5	5
Гли12Арг	5.5 ± 1.5	2
Фен20Цис	2.5 ± 1.1	7
Гли37Арг	16.5 ± 8.5	26
Лей38Вал	2.4 ± 1.8	22
Гли41Асп	14.1 ± 5.0	15
Гли41Сер	1.0 ± 0.3	16
Гис43Арг	1.8 ± 1.3	12
Гис46Арг	18.7 ± 8.5	62
Гис48Глн	1.2 ± 0.3	4
Асп76Вал	18.8 ± 7.9	4
Лей84Вал	3.1 ± 4.5	10
Лей84Фен	6.2 ± 3.0	18
Гли85Арг	6.0 ± 4.5	11
Асн86Лиз	1.7 ± 0.6	7
Ала89Вал	4.0 ± 0.4	5
Асп90Ала	9.0 ± 8.1	14
Гли93Арг	5.3 ± 4.6	4
Глу100Гли	4.6 ± 2.9	49
Асп101Асн	2.3 ± 0.8	17
Сер105Лей	3.4 ± 1.2	7
Лей106Вал	1.9 ± 1.1	6
Иле112Тре	0.9 ± 0.1	2
Иле112Мет	3.0 ± 0.9	9
Гли114Ала	2.7 ± 0.3	2
Асп125Гис	1.8 ± 0.2	2
Асн139Гис	3.0 ± 0.6	6
Лей144Сер	12.3 ± 3.7	2

Продолжение таблицы 6

Мутация	Дожитие пациентов, лет	Количество пациентов
Вал148Иле	1.7 ± 0.4	5