

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

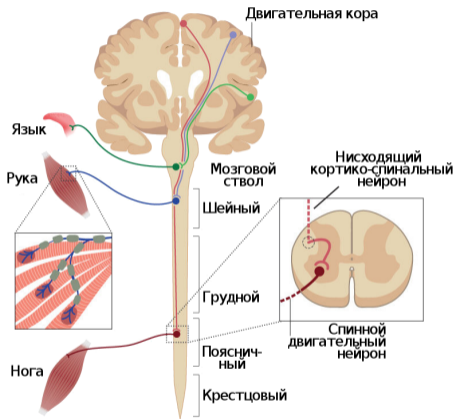
Н. А. Алемасов

*Руководитель:* к.б.н., доц. В. А. Иванисенко

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

Новосибирск, 2018

# Боковой амиотрофический склероз (БАС)



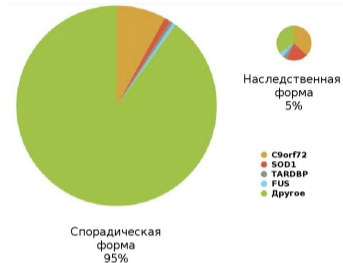
Nature, 2016, doi:10.1038/nature20413

- Неотвратимая дегенерация центральных и периферических моторных нейронов
- Приводит к прогрессирующему параличу и возможной смерти от дыхательной недостаточности
- Частота: 1-3 случая на 100 000 человек/год (США, Европа)
- Возникает примерно в возрасте 47 лет и длится, в среднем 2-4 года
- Лекарственные средства:
  - Рилузол (1995). Продлевает жизнь на 2-3 месяца
  - Радикава (2017). Облегчает симптомы

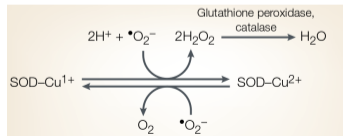
- 1824 г: Чарльз Бэллом БАС впервые описан, как заболевание
- 1874 г: Термин «боковой амиотрофический склероз» впервые упомянут в работе Жана Мартена Шарко

# Мутации в *SOD1* – важный фактор патогенеза

- До 25 % случаев БАС имеют наследственную природу
  - Доминантное наследование, с высокой пенетрантностью
- Из секвенирования полного экзона известно более 30 генов, ассоциированных с БАС
  - SOD1, TARDBP, FUS, OPTN и VCP (до 25 % наследственных случаев)
  - некодирующий регион C9orf72 (до 23 % наследственных случаев)
- Ген *SOD1*
  - Супероксиддисмутаза-1 катализирует реакцию превращения анионов супероксида на молекулярный кислород и пероксид водорода
  - Мутации в гене SOD1 встречаются в < 20 % наследственных случаев
  - Известно свыше 180 мутаций в белке SOD1 у пациентов с БАС

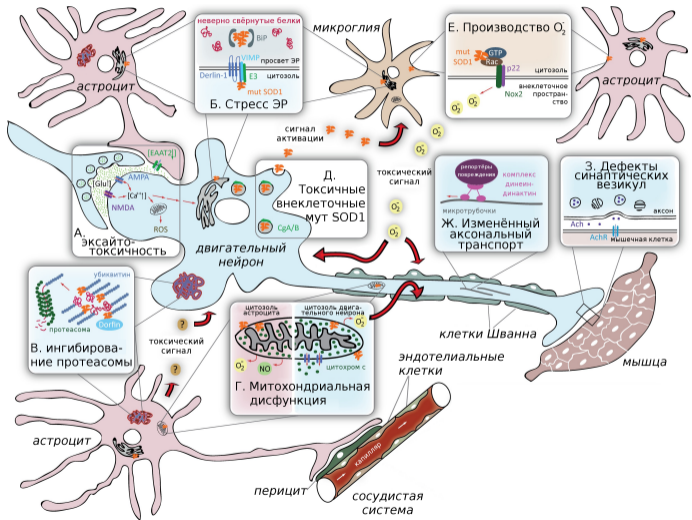


J Neurol Neurosurg Psych, 2017,  
doi:10.1136/jnnp-2017-315995



Nature Reviews Neuroscience, 2001,  
doi:10.1038/35097565

# Механизм возникновения и развития БАС

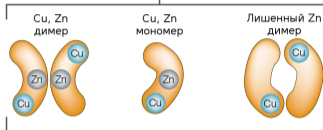


Journal of Cell Biology, 2009, doi:10.1083/jcb.200908164

**SOD1**  
дикого типа



**Мутантные SOD1**

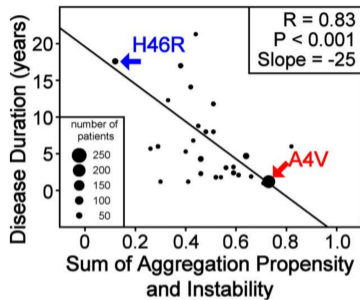


- Нарушенная окислительная активность
- Агрегация
- Потеря функции из-за агрегации
- Агрегация
- Сниженная активность шаперонов
- Агрегация
- Сниженная активность протеасомы

Nature Reviews Neuroscience, 2001,  
doi:10.1038/35097565

Накопление белковых агрегатов с участием SOD1, вызванное неверной укладкой в результате мутаций – одна из причин БАС

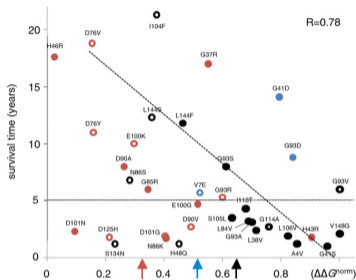
# Существующие регрессионные модели для предсказания тяжести БАС



PLOS Biology, 2008,  
10.1371/journal.pbio.0060170

Модель Wang et al., 2008 ( $R = -0.83$ )

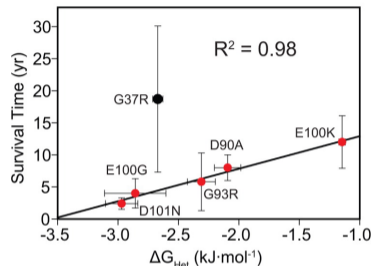
- подверженность мутантов SOD1 агрегации ( $\Delta G_f$ ,  $\Delta(\alpha, \beta)$ ,  $\Delta$ заряд)
- вероятность разупорядочения мутантов SOD1



Journal of Biological Chemistry, 2010, doi:  
10.1074/jbc.M109.086074

Модель Bystrom et al., 2010 ( $R = 0.78$ )

- изменение термостабильности в мутантах SOD1



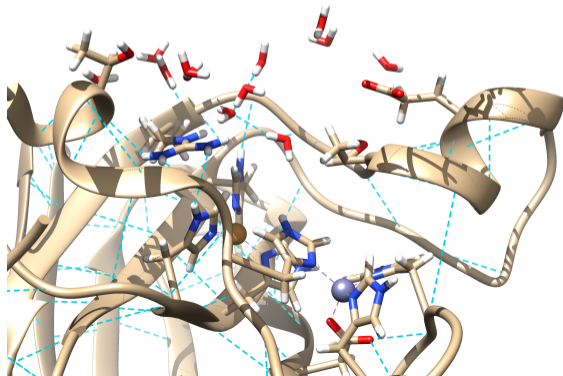
J. Am. Chem. Soc., 2016, doi:  
10.1021/jacs.6b01742

Модель Shi et al., 2016 ( $R = 0.99$ )

- свободная энергия гетеродимеризации

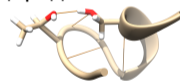
# Выдвинутая гипотеза о механизме БАС

Как дестабилизация, так и стабилизация структуры SOD1 влияет на увеличение вероятности нахождения его в метастабильном «патогенном» состоянии. 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.

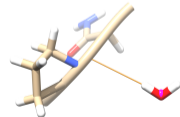


## Водородные связи

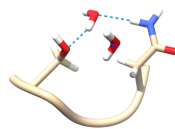
Между атомами  
белка



Между белком и  
водой



Водные мостики



# Цель работы

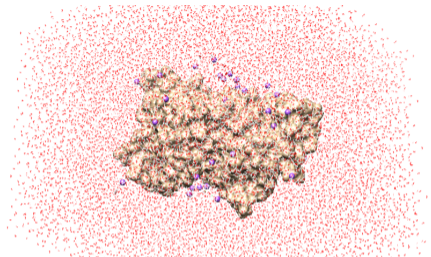
Компьютерный анализ связи между конформационными свойствами мутантных белков SOD1 и наследственной формой бокового амиотрофического склероза на основе оценки стабильности водородных связей и водных мостиков.

# 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.



# Используемые методы молекулярного моделирования

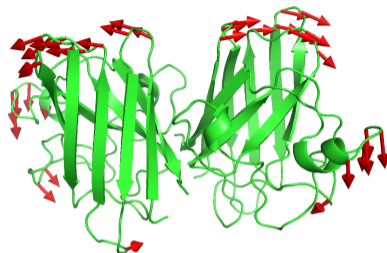


$$U = U_{\text{связи}} + U_{\text{углы}} + U_{\text{дв.углы}} + U_{\text{LJ}} + U_{\text{эл}}$$

## Метод молекулярной динамики (МД)

Явный растворитель + ионы  
 Детализированная форма уравнения потенциальной энергии  
 Высокое разрешение ( $\sim 1$  фс)  
 Высокие вычислительные затраты  
 Небольшие флуктуации атомов  
 Траектория движения частиц

*AMBER 12, ff99SB, 300 K, 5\*50 нс*



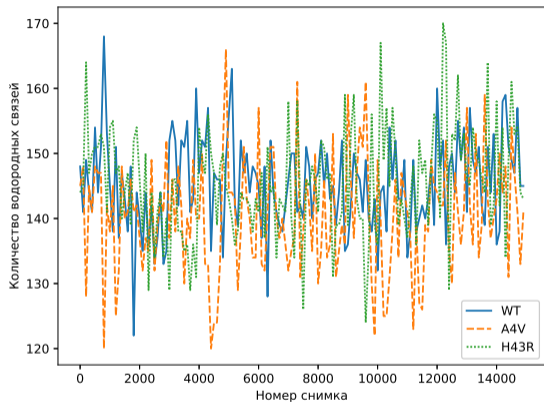
$$U = k \times \Delta x$$

## Эластичные сетевые модели (ЭС)

Нет растворителя  
 Упрощённая форма уравнения потенциальной энергии  
 Крупнозернистая модель белка  
 Низкие вычислительные затраты  
 Крупномасштабная динамика  
 Ансамбль конформаций

*ElNeto, 5 мод, 11 конформаций*

# Стабильность водородных связей



Водородные связи в ходе флуктуаций атомов белка образуются и разрушаются

Стабильность связи  $i$  в мутанте  $m$  по снимкам (МД) или конформациям (ЭС) из  $[1; n]$ :

$$O_i^m = \frac{1}{n} \sum_{j=1}^n o_{ij}^m$$

$o_{ij}^m = 1$ , если связь присутствует

$o_{ij}^m = 0$ , если связь отсутствует

	WT	A4V	H43R	D101N
HIS_120@HD1 HIS_120@ND1 ALA_140@O	0.14162	0.08076	0.1106	0.1199
HIS_120@HD1 HIS_120@ND1 GLY_141@O	0.25114	0.31684	0.30268	0.14948
HIS_120@HD1 HIS_120@ND1 SER_142@O	0.01698	0.02466	0.02006	0.0325
HIS_120@NE2 ARG_143@HE ARG_143@NE	0.00012	0.00012	0.10442	0.00374
HIS_120@NE2 ARG_143@HH12 ARG_143@NH1	0.00154	0	0	0
HIS_43@HD1 HIS_43@ND1 HIS_120@O	0.58168	0.57438	0	0.54576

Рассматривались водородные связи:

МД	между атомами белка	(PPhb)
МД	между атомами белка и атомами молекул воды	(PWhb)
МД	водные мостики	(Wbr)
ЭС	между атомами белка	(PPhb <sub>ЭС</sub> )



# Обучающие выборки для построения регрессионных моделей

Данные о дожитии пациентов брались из работы Wang et al., 2008.

Entries	Mutation	Reference	Duration ± SD in Years (number of patients)	Average Duration	Number of Patients
1	A4S	[156]	>3 (1)	>3	0
3	A4V	[67]	1.4 ± 0.9 (84)	1.2	205
		[158]	1.0 ± 0.4 (75)		
		[160]	1.0 ± 0.5 (8)		
		[161]	1.2 ± 0.8 (24)		
		[162]	0.9 (4)		

Для одной мутации имелось несколько различных значений  $\mu$ ,  $\sigma$  и  $n$ . Строилось нормальное распределение с параметрами:

$$\bar{\mu} = \frac{1}{K} \times \sum_{i=1}^K \mu_i$$

$$\bar{\sigma} = \frac{1}{K} \times \sqrt{\sum_{i=1}^K \sigma_i^2}$$

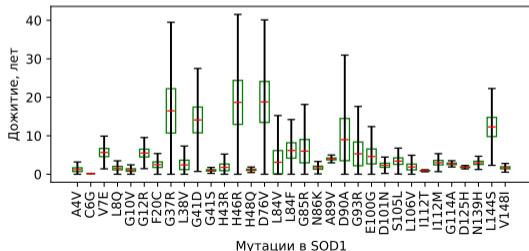
где  $K$  – количество значений  $\mu$ ,  $\sigma$  и  $n$  для заданной мутации.

Получены две обучающие выборки:

- $\Theta$  средние значения дожития  $\bar{\mu}$  для каждой мутации (33 точки);
- $\Theta_{exp}$  для каждой мутации  $n$  значений дожития, полученные из нормального распределения с параметрами  $\bar{\mu}$ ,  $\bar{\sigma}$  (566 точек).

Мутация	Дожитие пациентов, лет	Количество пациентов
Ала4Вал	1.2 ± 0.7	205
Гли12Арг	5.5 ± 1.5	2
Гли41Сер	1.0 ± 0.3	16
Лей84Вал	3.1 ± 4.5	10
Асп90Ала	9.0 ± 8.1	14
Лей106Вал	1.9 ± 1.1	6
Асн139Гис	3.0 ± 0.6	6

Обучающая выборка  $\Theta$



Обучающая выборка  $\Theta_{exp}$

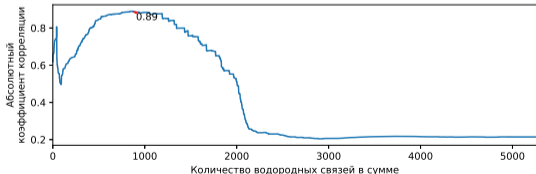
# Построение линейных регрессионных моделей

## Метод OSCAR

Оптимальная сумма абсолютных разностей стабильностей водородных связей в белке SOD1 дикого типа и его мутантах:

$$\Omega_{\mathcal{R}}^m = \sum_{k \in \mathcal{R}, R_k > 0} \bar{O}_k^m - \sum_{k \in \mathcal{R}, R_k < 0} \bar{O}_k^m, \quad \bar{O}_i^m = |O_i^m - O_i^{\Delta T}|$$

где  $\mathcal{R} \subseteq [1, N]$  – подмножество номеров связей, удовлетворяющих условию:  $|R_k| > R_{th}$ ,  $R_k$  – коэффициент корреляции  $\{\bar{O}_k^m\}$  с  $\{\Theta^m\}$  (или  $\{\Theta_{exp}^m\}$ ).  $R_{th}$  получен путём максимизации  $|R|$  между  $\Omega_{\mathcal{R}}^m$  и  $\Theta^m$



## Гребневая регрессия RR

Множественная линейная регрессия, более устойчивая к переобучению по сравнению с обычной линейной регрессией за счёт штрафов, накладываемых на величину коэффициентов регрессионного уравнения.

Предикторы – водородные связи.

Параметры:

Глубина регуляризации (alpha)	1
Центрирование данных (fit_intercept)	используется
Нормализация данных (normalize)	не используется
Максимальное количество итераций для метода сопряжённых градиентов (max_iter)	выбирается автоматически
Точность решения (tol)	0.0001
Решатель (solver)	«auto»
Состояние датчика псевдослучайных чисел (random_state)	0

# Построение нелинейных регрессионных моделей

## Модель random forest RF

Множественная регрессия. Учитываются сложные нелинейные взаимные связи между предикторами. Есть неявная перекрёстная проверка.

Предикторы – водородные связи.

Параметры:

Количество деревьев (n_estimators)	100
Мера качества разбиения (criterion)	«mse»
Максимальное количество предикторов (max_features)	«auto»
Максимальная глубина дерева (max_depth)	неограниченна
Минимальное количество элементов выборки для разбиения внутренней вершины (min_samples_split)	2
Минимальное количество элементов выборки в вершине-листе (min_samples_leaf)	1
Максимальное количество вершин-листов (max_leaf_nodes)	неограниченно
Минимальное снижение меры «impurity» (min_impurity_decrease)	0
Использование bootstrap (bootstrap)	используется
Использование элементов выборки для оценки $R^2$ на тестовой выборке (oob_score)	не используется
Состояние датчика псевдослучайных чисел (random_state)	0

## Метод SSRF

С помощью программы DSSP была получена вторичная структура белка SOD1. В структуре для каждой из двух субъединиц выделяется 23 элемента:

$\alpha$ -спирали		петли		$\beta$ -тяжи	
H1	57–60	C1	1–2	E1	3–9
H2	108–111	C2	10–14	E2	15–22
H3	132–135	C3	23–28	E3	29–36
		C4	37–40	E4	41–48
		C5	49–56	E5	83–89
		C6	61–82	E6	95–101
		C7	90–94	E7	116–120
		C8	102–107	E8	143–151
		C9	112–115		
		C10	121–131		
		C11	136–142		
		C12	152–153		

Все водородные связи группировались в зависимости от принадлежности формирующих их аминокислотных остатков одному из элементов вторичной структуры.

Стабильности  $O_i^m$  водородных связей суммировались для водородных связей из одной группы.

Всего 46 предикторов.

Строилась регрессия на основе метода random forest.

# Построение комбинированных регрессионных моделей и bootstrap

## Модели MR

Множественная гребневая регрессия.

Предикторы – предсказанные по параметрам PPhb, PWhb, Wbr и PPhbЭС значения  $\hat{\Theta}^m$  дожития пациентов одной моделью из: OCAPC, SSRF, RF и RR.

$$\hat{\Theta}^m = a_1 * \hat{\Theta}^m_{PPhb} + a_2 * \hat{\Theta}^m_{PWhb} + a_3 * \hat{\Theta}^m_{Wbr} + a_4 * \hat{\Theta}^m_{PPhbЭС} + a_0$$

$a_0$ – $a_4$  – коэффициенты уравнения гребневой регрессии.

## Проверочные модели Rhb

Линейная регрессия. Предиктор –  $O_i^m$  одной водородной связи, выбранной случайным образом.

При обучении модели  $\Theta$  (или  $\Theta_{exp}$ ) случайным образом перемешивались.

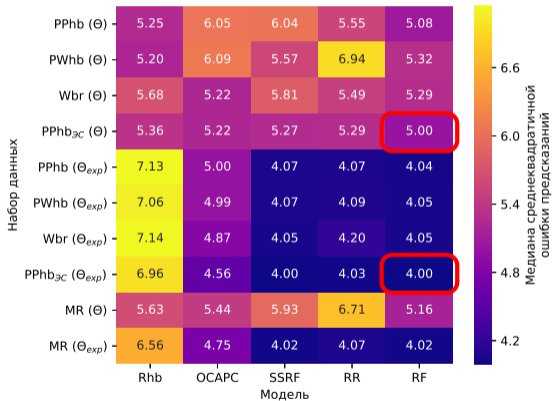
## Перекрёстная проверка

$$\mathfrak{M} = \Theta \text{ (или } \mathfrak{M} = \Theta_{exp}\text{)}.$$

- обучающая выборка ( $\frac{3}{5}$  элементов  $\mathfrak{M}$ );
- тестовая выборка ( $\frac{2}{5}$  элементов  $\mathfrak{M}$ );
- 100 шагов bootstrap;
- оценивалась величина стандартной ошибки предсказаний (RMSPE);
- распределения величины RMSPE сравнивались с распределением RMSPE для Rhb с помощью непараметрического критерия Манна-Уитни-Вилкоксона.

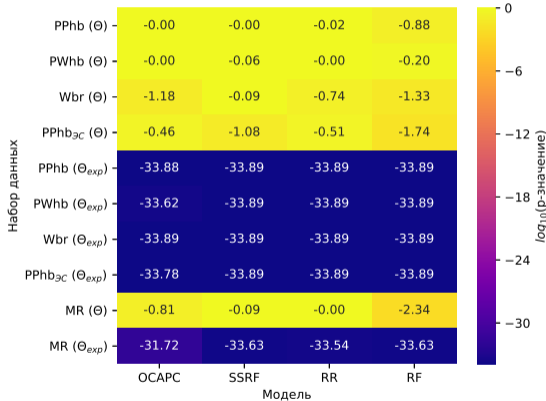
# Анализ точности регрессионных моделей

## Медиана RMSPE



Наиболее точные – **RF** (PPhbЭC)

## Значимость отличия ошибок от Rhb

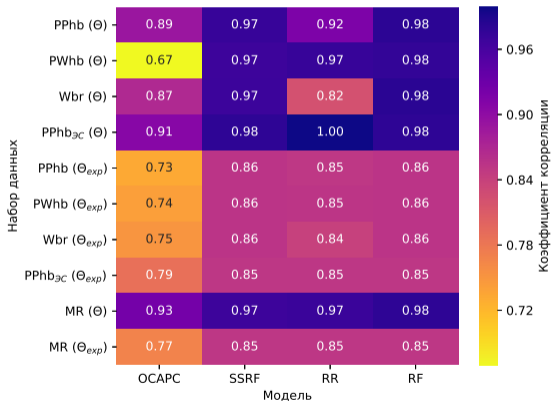


Логарифм, больший 1.3 соответствует  $p > 0.05$



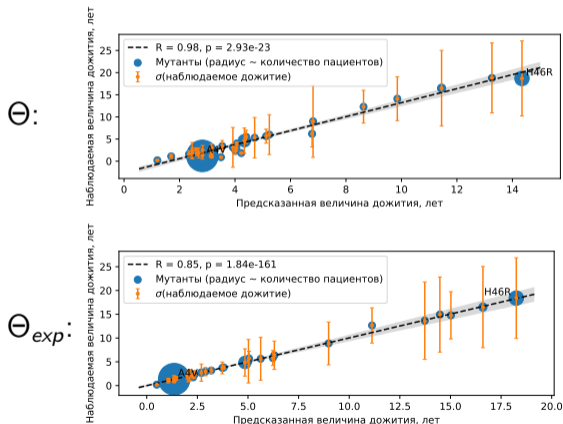
# Анализ точности регрессионных моделей

Коэффициенты корреляции между  $\hat{\Theta}^m$  и  $\Theta$  (или  $\Theta_{exp}$ )



$p < 0.001$

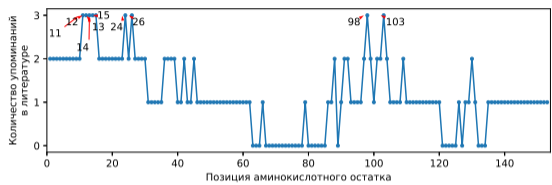
## Модели RF (PPhb<sub>ЭС</sub>)



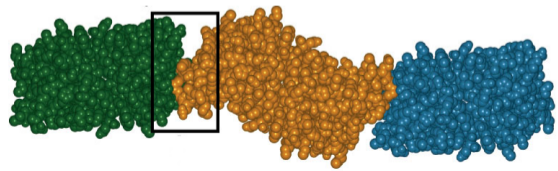
Учёт  $\sigma$  и  $n$  помимо  $\mu$  позволяет снизить RMSEP

# Остатки мутантных SOD1, участвующие в агрегации и гидрофобность

## Частота упоминания агрегирующих остатков

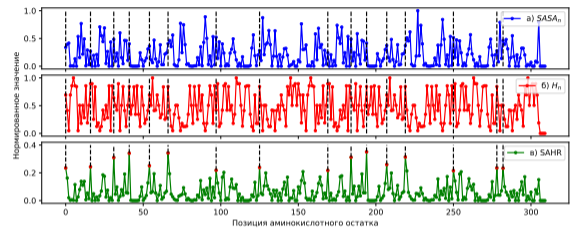


Остатки в позициях 11–15, 24, 26, 98 и 103 упоминаются трижды



Nature Structural Biology, 2003, doi:10.1038/nsb935

## Гидрофобные остатки, доступные растворителю

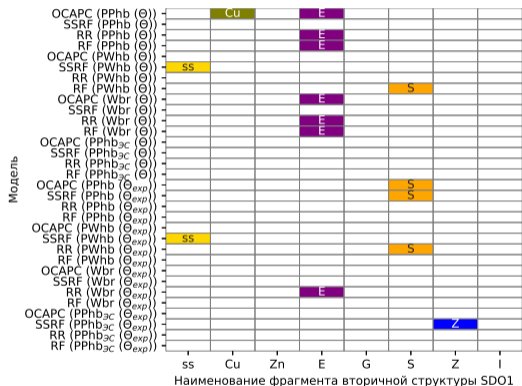


1. методом Шрейка-Рапли находится SASA;
2. рассчитывается гидрофобность каждого остатка ( $H$ );
3. рассчитывается доступность гидрофобных остатков молекулам растворителя (SAHR).

$$SAHR = SASA_n * H_n$$

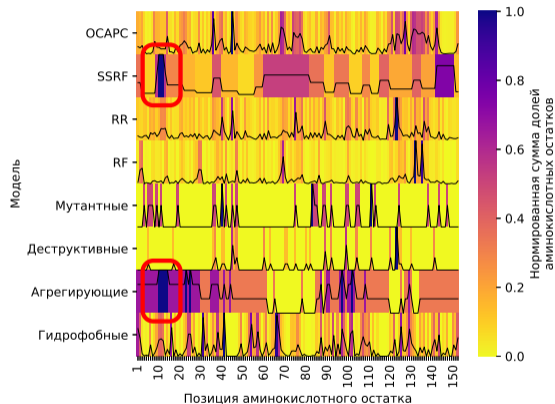
## Расположение важных для предсказаний аминокислотных остатков

$\mathcal{S}_{\text{крит}}$ , перепредставленные среди  $\mathcal{I}$



Критические позиции  $\mathcal{S}_{\text{крит}}$  – формирующие дисульфидную связь («ss»), сайт связывания иона меди («Cu»), сайт связывания иона цинка («Zn»), электростатическую петлю («E»), «greek-key»-петлю («G»), дисульфидную петлю («S»), цинк-связывающую петлю («Z»), интерфейс между субъединицами («I»)

Расположение  $\mathcal{I}$  в первичной структуре SOD1



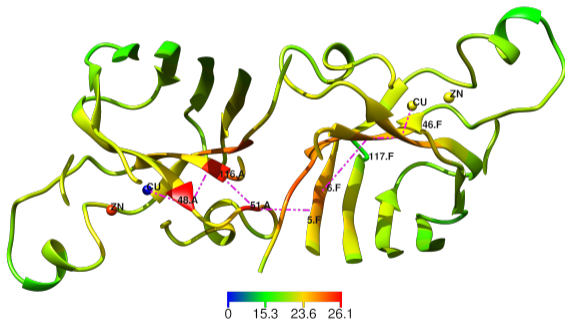
$\mathcal{I}$  – аминокислотные остатки SOD1, формирующие водородные связи, важные для предсказания дожития пациентов.

$\mathcal{S}_{\text{arg}}$  перепредставлены среди  $\mathcal{I}_{\text{SSRF}}$  ( $p < 0.05$ )

# Исследование крупномасштабной динамики белка SOD1 с помощью ЭС

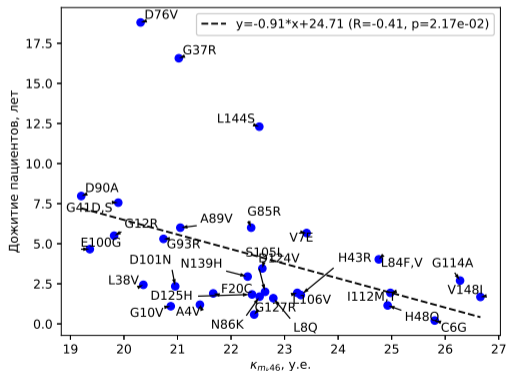
Механическая жёсткость  $\kappa_{ij}$  между парой остатков  $i$  и  $j$  имеет смысл константы жёсткости эквивалентной «пружины» между ними.

## Механическая жёсткость между активными центрами



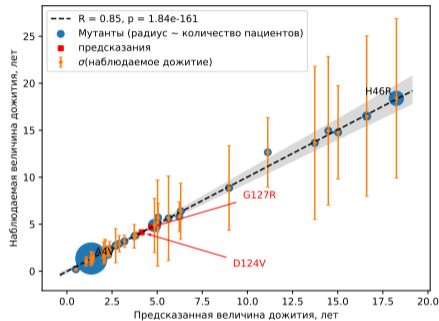
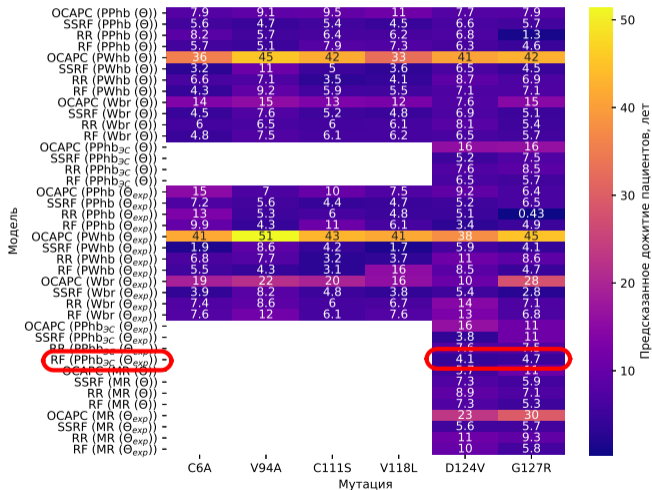
Движения обоих ионов меди скоррелированы: повышенной механической жёсткостью до активных центров обладают остатки из интерфейса между субъединицами.

## Связь механической жёсткости и дожития



$m_k$  соответствуют изучаемым мутантным позициям. Гис46 участвует в формировании сайта связывания иона меди.

# Предсказания дожития пациентов с мутациями в SOD1



- Известно, что пациент с мутацией D124V жил более 2 лет.
- Известно, что пациент с мутацией G127R жил около 7 месяцев.

# Выводы

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

# 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.

# Кандидатские экзамены, публикации, конференции I

## Экзамены:

- Философия: хорошо
- Английский: отлично
- Специальность: хорошо

## Конференции:

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 мультikonференции 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).

## Публикации:

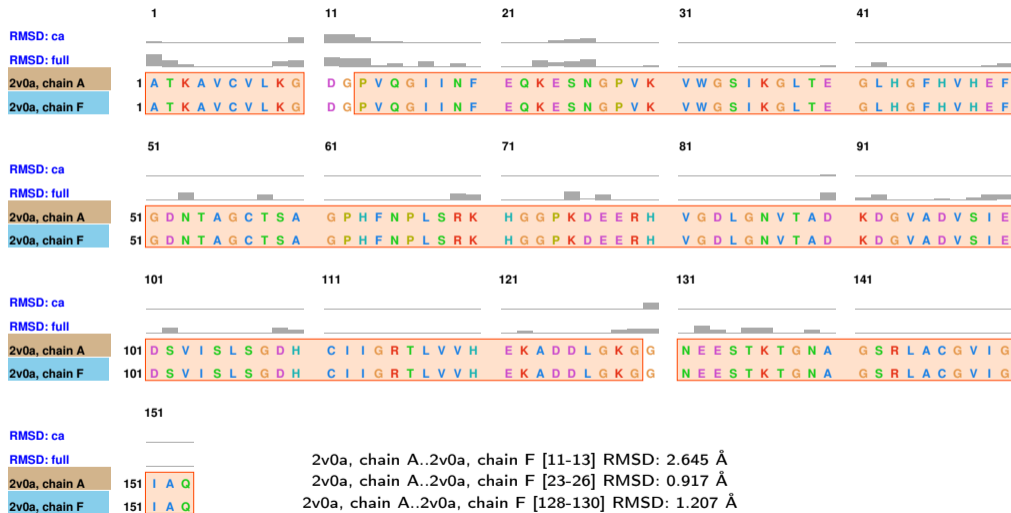
1. Алемасов Н.А., Иванисенко Н.В., Иванисенко В.А. Структурные и динамические особенности мутантов белка SOD1, ассоциированных с боковым амиотрофическим склерозом // Вавиловский журнал генетики и селекции (ИФ 0.440). 2014. Т. 18. No 4\2. С. 920–927.
2. Alemasov, N. A., Ivanisenko, N. V., Medvedev, S. P., Zakian, S. M., Kolchanov, N. A., Ivanisenko, V. A. Dynamic properties of SOD1 mutants can predict survival time of patients carrying familial amyotrophic lateral sclerosis // Journal of Biomolecular Structure and Dynamics (IF 3.123). 2017. Vol. 35. No. 3. P. 645–656.
3. Nikolay A. Alemasov, Nikita V. Ivanisenko, Vladimir A. Ivanisenko. 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 (IF 1.754). 2017. Vol. 77. P. 378–385.
4. Nikolay A. Alemasov, Nikita V. Ivanisenko, Srinivasan Ramachandran, Vladimir A. Ivanisenko. Molecular mechanisms underlying the impact of mutations in SOD1 on its conformational properties associated with amyotrophic lateral sclerosis as revealed with molecular modelling // BMC Structural Biology (IF 1.771). 2018. Vol. 18. Suppl 1. No. 1. P. 1–14.



# Благодарности

- Иванисенко Н.В.
- Ефимов В.М.
- Подколотный Н.Л.
- Закиян С.М.
- Медведев С.П.

# Приложение А: Ассиметрия движения субъединиц SOD1



## Приложение Б: Количество водородных связей в МД и ЭС

