

Санкт-Петербургский Государственный университет

Ганкин Матвей Михайлович

**Методы предсказания амилоидной структуры и прионных
свойств белков: состояние и перспективы**

Выпускная квалификационная работа
по направлению подготовки «Биология»
основная образовательная программа магистратуры
«Биоинформатика»

Работа выполнена на базе кафедры
Генетики и Биотехнологии
Санкт-Петербургского Государственного Университета

Научный руководитель: к.б.н., доцент Нижников Антон Александрович

Санкт-Петербург, 2021 г.

Содержание

Содержание	2
Введение	3
1. Обзор литературы.....	5
1.1. Понятие «амилоид»	5
1.2. Классификация амилоидов	6
1.3. Экспериментальные методы исследования амилоидов	8
1.4. Биоинформатическое предсказание амилоидогенных свойств белков.....	10
1.4.1 Предикторы, основанные на машинном обучении	11
1.4.2 Предикторы, основанные на использовании индивидуальных свойств аминокислот	13
1.4.3 Предикторы основанные на композиционном подходе	15
1.4.4 Консенсусный предиктор	18
2. Материалы и методы.....	28
2.1. Данные	28
2.2. Тестирование предикторов	28
2.3. Обработка полученных результатов.....	29
3. Результаты и обсуждение	30
4. Выводы	35
Список источников литературы.....	36
Благодарности	43

Введение

Термин «амилоид» был впервые использован Матиасом Шлейденом в 1838 г. применительно к конгломератам крахмала в клетках растений. Спустя 16 лет этот термин использовался Рудольфом Людвигом Карлом Вирховым для описания патологических включений в тканях «восковой» печени, которые по аналогии с крахмалом окрашивались йодом. С тех пор наши представления об этих агрегатах изменились и на данный момент амилоиды можно определить как фибриллярные белки, обладающие кросс-бета структурой: четвертичная структура бета-тяжей образует набор связанных параллельных слоев, перпендикулярных длинной оси фибриллы [Нижников, Антонец, Инге-Вечтомов, 2015]

Множество известных амилоидов лежат в основе заболеваний ЦНС, таких как болезнь Альцгеймера, болезнь Якоба-Крейцфельда, Синдром Герстмана-Штраусслера-Шейнкера у человека, скрепи у овец и губчатая энцефалопатии у крупного рогатого скота, а также целого ряда болезней, называемых амилоидозами, имеющих первичное или вторичное происхождение, локализованный или системный характер [Hazenbergh, 2013; Taverna, 2008]. В последние десятилетия с накоплением научных данных произошла смена парадигмы амилоидов только как патологической формы белков, на точку зрения при которой амилоиды также рассматриваются как важный компонент биологических функций множества организмов [Taverna, 2008]. Существует множество работ, которые свидетельствуют о широком функциональном значении амилоидов в природе: хранение гормонов в неактивном состоянии и участие в синтезе меланина у млекопитающих [Watt B, van Niel G, Raposo G, 2011], компонент формирования биопленок у бактерий и воздушных гифов у грибов [Taglialegna *et al.*, 2016; Yang *et al.*, 2017], вхождение в состав хориона чешуекрылых [Iconomidou, Vriend, Hamodrakas, 2000], запасующая функция растений [Antonets *et al.*, 2020], компонент сигнальных путей *Saccharomyces cerevisiae*. [Ваха *et al.*, 2005; Otzen, Riek, 2019; Shorter, Lindquist, 2005]. Особую группу амилоидов составляют прионы, обладающие инфекционными свойствами *in vivo* [Bolton *et al.*, 1987; Iadanza *et al.*, 2018; Wickner, 1994]

Строение амилоидов обуславливает наличие у них ряда важных физико-химических свойств. Структура амилоидов делает их устойчивыми к большому числу денатурирующих белок веществ, таких как ионные детергенты и хаотропные агенты, а также они имеют высокую устойчивость к действию протеаз [Gallardo, Ranson, Radford, 2020]. Другим важным свойством является автокаталитическая способность амилоидов

воспроизводить свою структуру в белках с той же последовательностью. Соответствующие структурные белки в их нативном состоянии в присутствии амилоидов, которые выступают в роли конформационной матрицы, начинают менять свою структуру. Наименьший известный на данный полипептид имеющие данное свойство содержит 7 аминокислотных остатков. [Jucker, Walker, 2013; Nelson *et al.*, 2005].

Особенности кросс-бета структуры налагают серьезные ограничения на способы идентификации и исследования амилоидных белков. Доступными подходами для определения амилоидной структуры *in vitro* являются методы окрашивания специфическими красителями: самый часто используемый анилиновый краситель конго-красный, дает яблочно-зеленый цвет в поляризованном свете при связывании с амилоидами. Для определения амилоидной структуры *in vivo* возможно использование методов иммуногистохимии, однако их применение имеет свои ограничения. Для исследования структуры амилоидов в практике используются методы ЯМР-спектроскопии и двухмерной рентгеновской дифракции. Такие способы исследования эффективны, но дорогостоящи [Yakupova *et al.*, 2019].

Широкий практический интерес к исследованию амилоидов и ограничения, которые накладывает их структура, делают значимым развитие биоинформатических подходов для изучения амилоидных белков и прионов *in silico*. За последние 20 лет было разработано более 30 алгоритмов, которые идентифицируют амилоидогенные участки в полипептидах, на основе известных последовательностей белков. В нашей работе мы делаем попытку оценить эффективность существующих инструментов предсказания амилоидных свойств белков, разделив известные амилоиды на три группы: патогенные, функциональные и прионы.

Целью данной работы является сравнительный анализ эффективности существующих биоинформатических инструментов предсказания амилоидной структуры и прионных свойств полноразмерных белков

В рамках цели были сформулированы следующие **задачи**:

1. Проанализировать литературные данные о существующих биоинформатических подходах для предсказания амилоидогенных свойств белков.
2. Предложить классификацию таких подходов, основываясь на принципах их работы.
3. Изучить эффективность предсказания амилоидных свойств белков существующими биоинформатическими инструментами.

1. Обзор литературы

1.1. Понятие «амилоид»

Понятием «амилоид» обозначают определенную структурную конформацию полипептидной цепи, которую, как предполагают некоторые авторы, может принять практически любой белок [Padanza *et al.*, 2018]. В основе этой конформации лежит повторяющийся мотив кросс-бета структуры, также называемый ядром амилоида – близкорасположенные взаимосвязанные межмолекулярные бета-слои. Подобная конформация остается устойчивой благодаря действию водородных связей между соседними тяжами и за счет «стерической застезки», когда боковые цепи одного слоя располагаются между такими же цепями соседнего слоя, вызывая действие ван-дер-ваальсовых сил [Gallardo, Ranson, Radford, 2020],[Nelson *et al.*, 2005]. Вкупе такое строение амилоидов делает их чрезвычайно стабильными термодинамически [Baldwin *et al.*, 2011] и механически [Knowles *et al.*, 2007]. В рамках последовательностей амилоид-образующих белков выделяют амилоидогенные участки, также известные как hot-spot. Эти участки могут являться отправной точкой для образования амилоида и направлять этот процесс [Ivanova *et al.*, 2004][Ventura *et al.*, 2004].

Переход из состояния растворимого мономера белка в состояние нерастворимого амилоидного полимера в конечном счете приводит к образованию внутри- и внеклеточных белковых агрегатов. Процесс образования амилоидных агрегатов из мономеров может протекать различными путями и зависит от ряда внутренних и внешних факторов. Так, описаны случаи зависимости образования амилоидов от аминокислотной последовательности белка, конформации мономера, температуры и кислотности среды [Almeida, Brito, 2020]. В общем случае выделяют зависимые и независимые от нуклеации механизмы агрегации. В первом случае, процесс происходит в три фазы: нуклеация, элонгация и фаза насыщения. Нуклеация это медленный процесс, связанный с образованием ядра амилоида, который в дальнейшем служит направляющим элементом при присоединении мономеров и объединении их в олигомеры с кросс-бета структурой. В фазе элонгации процесс ускоряется, ядро, мономеров и олигомеры продолжают взаимодействие создают префибриллярные структуры, которые вскоре образуют упорядоченные протофибриллы. Во время фазы насыщения, концентрация мономеров достигает минимального значения, что приводит к сборке протофибрилл в зрелые амилоидные фибриллы с различными морфологическими структурами и различной

степенью полиморфизма. Примером такого механизма агрегации служит агрегация лизоцима [Hill *et al.*, 2009]. Независимый от нуклеации механизм образования агрегатов представляет собой цепочку реакции линейной полимеризации. Как только процесс агрегации начинается он превращается в каскад реакций, когда присоединение амилоидогенного мономера к растущему агрегату энергетически предпочтителен без необходимости присутствия ядра амилоида. Примером может служить агрегация транстиретина [Hurshman *et al.*, 2004].

1.2. Классификация амилоидов

Амилоидные структуры встречаются в клетках представителей и прокариот и эукариот. При этом их природа и влияние на жизнедеятельность организмов может коренным образом отличаться от случая к случаю. В зависимости от происхождения и процессов, в которых они принимают участие, амилоиды можно разделить на три группы: патологические, функциональные и прионные амилоиды. К патологическим амилоидам относят белки, для которых образование амилоидных структур не является нативным свойством, но они приобретают его в рамках различных патологий. Это может быть результатом процессинга другими белками [O'Brien, Wong, 2011] или различного рода мутаций [Planté-Bordeneuve, Said, 2011] [Leeuwen Van *et al.*, 2000]. Патогенные амилоиды ассоциированы с различного рода заболеваниями ЦНС, амилоидозами и некоторыми видами рака [Chiti, Dobson, 2017]. Болезнь Альцгеймера, наиболее известный пример нейродегенеративного заболевания ассоциированного с амилоидами, является официальной причиной смерти в более чем 120 000 случаях в Америке за 2018 год [2020 Alzheimer's disease facts and figures, 2020]. Не все, но значимая часть патологических амилоидных структур обладают цитотоксическими свойствами. Точный молекулярный механизм токсичности амилоидов не установлен в полной мере. Установлено что амилоидные агрегаты могут связывать целый ряд рецепторов на поверхности клеток, вмешиваясь в работу сигнальных путей, увеличивать проницаемость клеточных мембран, также показано что амилоидные фибриллы могут связывать глюкозаминогликаны, гепарин, прочие белковые молекулы и некоторые молекулы РНК [Iadanza *et al.*, 2018; Kaye, Lasagna-Reeves, 2013].

К функциональным амилоидам можно отнести белки, которые образуют амилоидные структуры, обеспечивающие нормальное протекание процессов в организме.

В работе [Otzen, Riek, 2019] авторы разделяют функциональные амилоиды на пять групп в зависимости от их роли в жизнедеятельности, при этом один и тот же белок может относиться более чем к одной группе:

- I. Накопление химически-активных компонентов. Некоторые пептидные молекулы образуют амилоидные структуры, иногда теряя при этом биологическую активность. Это позволяет организму регулировать активность таких молекул, высвобождая их по мере необходимости. К этой группе можно отнести главный основной белок в клетках-эозинофилах млекопитающих [Soragni *et al.*, 2015] или пируват киназу дрожжей [Saad *et al.*, 2017].
- II. Структурные. К этой группе можно отнести те амилоидные белки, которые выполняют структурную роль в ходе различных клеточных процессов. Белки Curli и Fap входящие в состав биопленок некоторых бактерий относятся к этой группе [Taverna, 2008; Zeng *et al.*, 2015].
- III. Информационные. К этой группе относятся амилоиды, которые выступают как элементы эпигенетического наследования и памяти. Сюда можно отнести целый ряд белков *Saccharomyces cerevisiae* [Alberti, 2009; Ваха *et al.*, 2005; Derkatch *et al.*, 2000]
- IV. Потеря функции. В ряде случаев при образовании амилоидной структуры белки теряют свою химическую активность. Сюда же можно отнести уже упомянутые во II и III группах белки.
- V. Приобретение функции. Также существует обратный случай, когда белки в растворимой форме инактивированы и, для того чтобы включиться в биологические процессы, им необходимо принять амилоидную конформацию. Сюда можно отнести белки RIP1 и RIP3 млекопитающих, чей процесс образования амилоидных структур задействован в некрозе клеток [Li *et al.*, 2012].

Прионы – группа инфекционных амилоидных белков, которые могут как нести функциональную роль в организмах, так и быть ассоциированными с заболеваниями [Bolton *et al.*, 1987; Wickner, 1994]. От других групп амилоидов прионы отличает их инфекционность и трансмиссивность. Как и прочие амилоиды, прионы обладают автокаталитичностью и при попадании незараженный организм прионы могут сравнительно быстро начать трансформацию нативных гомологичных белков. При этом у прионов существуют механизмы, которые позволяют прионным агрегатам делиться на полимеры меньшего размера, что обеспечивает их эффективную передачу в дочерние

клетки, в отличие от неинфекционных амилоидов. Например, в клетках дрожжей фрагментацию прионных агрегатов обеспечивают белки-шапероны, центральную роль среди которых играет Hsp104 [Chernova, Wilkinson, Chernoff, 2017]. В результате фрагментации, каждый новый полимер или олигомер сохраняет автокаталитические свойства и становится новой точкой образования агрегатов [Kushnirov *et al.*, 2007; Marrero-Winkens, Sankaran, Schätzl, 2020]. К прионам относится белок PrP в конформации Sc (scrapie, скрепи) [Bolton *et al.*, 1987], который связывают с рядом нейродегенеративных заболеваний, включая коровье бешенство, скрепи (почесуха) у овец и болезнь Крейтцфельда-Якоба [Wille, Requena, 2018].

1.3. Экспериментальные методы исследования амилоидов

Уникальные физико-химические свойства амилоидов усложняют их изучение. Тем не менее амилоиды активно исследуются с применением разных подходов. Далее представлены основные на текущий момент экспериментальные методы исследования амилоидов.

Методы иммунохимического анализа основываются на создании и использовании специфических афинных к объекту исследования антител. В рамках исследования амилоидных белков, антитела получают как к нативным неструктурированным мономерам белков, так и к олигомерам и амилоидным фибриллам [Galvin, 2003; Genst De, Messer, Dobson, 2014]. Эти методы позволяют обнаруживать амилоиды и их прекурсоры в культурах клеток и тканях, а также *in vivo*, но, сами по себе, не дают возможности изучать структуру или кинетику агрегации амилоидов, равно как и не позволяют делать однозначное заключение об амилоидных свойствах исследуемых белковых агрегатов, поэтому данный подход используют совместно с другими. Например, иммунофлуоресцентный анализ в клинических исследованиях амилоидов, когда антитела используют вместе с флуоресцирующими метками и таким образом визуализируют кинетику образования амилоидных агрегатов [Wisniowski, Wechalekar, 2020].

Курли-зависимый метод продукции амилоидов (Curli-dependent amyloid generator) основывается на способности *E. coli* секретировать на поверхность клетки амилоидогенные белки CsgA и CsgB, которые в свою очередь образуют амилоидные фибриллы курли. Механизм транспорта этих белков из цитоплазмы во внеклеточную среду способствует принятию ими амилоидной конформации в случае, если исследуемые

белки способны образовывать амилоиды. Авторы показали, что, используя генно-модифицированные штаммы *E. coli*, можно использовать этот механизм для исследования амилоидных свойств разнообразных белков [Sivanathan, Hochschild, 2013].

Метод двумерной рентгеновской дифракции основывается на явлении дифракции рентгеновских лучей на трехмерной кристаллической решетке и позволяет определять атомную пространственную структуру исследуемого вещества [Makin, Serpell, 2005]. Именно с помощью этого подхода была описана атомная структура амилоидов и определена их уникальная кросс-бета структура. Анализ образования крестообразной «кросс-бета картины», отражающей характерные расстояния между бета-цепями (около 4,5 Å) и бета-слоями (около 10 Å) в амилоидных фибриллах при рентгеновской дифракции является ключевым способом доказательства амилоидной структуры изучаемых белковых агрегатов [Iadanza *et al.*, 2018].

Для исследования амилоидов также применяются различные методы световой микроскопии с использованием афинных к амилоидным структурам красителей: конго красный и тиофлавин -Т/-S. Конго красный начали использовать еще в начале 20 века для обнаружения амилоидов в тканях [Yakupova *et al.*, 2019]. Связавшись с амилоидной структурой, этот краситель дает яблочно-зеленое свечение при поляризационной микроскопии [Frid, Anisimov, Popovic, 2007]. Окрашивание конго красным с последующим анализом образцов в поляризованном свете является «золотым стандартом» в гистологических исследованиях амилоидов. Более новая группа красителей, тиофлавины-Т/-S, также как и конго красный имеют высокую специфичность и афинность к амилоидным структурам. Эти красители являются флуоресцентными и используются в сочетании с конфокальной микроскопией и спектрофотометрией как для идентификации амилоидов, так и для исследования кинетики агрегатов [Naiki *et al.*, 1990; Westermarck, Johnson, Westermarck, 1999].

Трансмиссионная электронная микроскопия (ТЭМ) представляет собой подход, при котором через исследуемый объект, обычно представленный ультратонким срезом, пропускают пучок ускоренных электронов. За счет неоднородного поглощения электронов образцом, формируется двумерная картина распределения плотности прошедшего электронного потока. Таким образом получают черно-белые изображения высокого разрешения [Harris, 2015]. На таких изображениях отложения амилоидов выглядят как скопление прямых, неветвящихся фибрилл от 6 до 13 нм в диаметре и 100-1600 нм в длину. Этот метод используют для обнаружения амилоидов и изучения их

морфологии [Yakupova *et al.*, 2019], а также при определении амилоидозов в клинике [Leung, Nasr, Sethi, 2012].

Другой распространённой в практике разновидностью электронной микроскопии является растровая электронная микроскопия (РЭМ). В основе РЭМ лежит поточечное сканирование объекта исследования пучком электронов. Установка улавливает спектр различных сигналов, полученных в результате столкновения пучка с материалом и на основе этого получается изображение высокого разрешения вплоть до субнанометровых размеров [Fischer *et al.*, 2012]. Также как и с ТЭМ, этот подход используется для идентификации и изучения морфологии, а также в клинических исследованиях для обнаружения амилоидозов [Cunha Filho Da *et al.*, 2016].

Метод протеомного скрининга амилоидов используют в целях идентификации амилоидных белков в организме [Nizhnikov *et al.*, 2014]. Данный метод позволяет обнаружить амилоидные белки в клетках, вне зависимости от их последовательности и уровня экспрессии в клетке. Данный подход базируется на общей для всех амилоидов высокой устойчивости к ионным детергентам. Белковую фракцию, насыщенную амилоидными агрегатами, получают путем ультрацентрифугирования лизата клетки в присутствии ионных детергентов, которые разрушают большинство неамилоидных белковых комплексов, однако амилоидные агрегаты остаются частично или полностью устойчивыми к таким обработкам. Полученную фракцию разделяют на отдельные белки или пептиды путем гель-электрофореза или трипсинолиза с последующей высокоэффективной жидкостной хроматографией [Nizhnikov *et al.*, 2016], которые в дальнейшем идентифицируют путем масс-спектрометрии.

1.4. Биоинформатическое предсказание амилоидогенных свойств белков

Естественным решением проблемы идентификации и исследования амилоидов в лабораторных условиях является разработка биоинформатических подходов к предсказанию амилоидной структуры. В рамках нашей работы мы отобрали 24 существующих предиктора. Для дальнейшей работы мы разделили их на четыре категории, в зависимости от подхода к предсказанию амилоидогенных и прионных свойств: машинное обучение, индивидуальные свойства аминокислот, композиционный подход и консенсусный алгоритм.

1.4.1 Предикторы, основанные на машинном обучении

К этой категории нами были отнесены те предикторы, в основе которых лежит использование нейронных сетей, машинного обучения и/или марковских моделей.

Предиктор Arpp основывается на предварительно отобранной и обученной нейронной сети. Авторы спроектировали набор нейронных сетей исходя из двух предположений: 1) в полипептидной цепи амилоида существуют короткие последовательности, которые направляют рефолдинг белка по пути формирования амилоидных фибрилл; 2) минимальная длина такой подпоследовательности равна шести аминокислотам. Для оценки амилоидогенности последовательностей авторы собрали данные о физико-химических и биохимических свойствах аминокислот с помощью баз данных APDBase и AAindex. Далее путем рекурсивного отбора признаков авторам удалось отобрать свойства аминокислот наиболее значимые для склонности к формированию амилоидов. Эти данные легли в основу набора нейронных сетей построенных на разных видах классификаторов. В результате наибольшую эффективность показала нейронная сеть, построенная на наивном байесовском классификаторе и 14 отобранных свойствах аминокислот которая и является конечным предиктором. Авторы подчеркивают, что их подход основывается на амилоидогенных свойствах коротких участков последовательностей, но не учитывает, что таких участков может быть больше одного и процесс образования фибрилл может быть построен на их взаимодействии. Авторы отмечают, что предиктор обучен на коротких последовательностях и с увеличением длины таких участков точность предсказания будет снижаться [Família *et al.*, 2015].

Веб-сервер Net-CSSP является интерактивной онлайн версией предиктора CSSP2 [Yoon *et al.*, 2007]. В основе предиктора лежит использование нейронных сетей с прямой связью и обратным распространением ошибок. Данные сети рассматривают любую последовательность как набор 7-членных полипептидов, в которых на основе информации об энергии третичных взаимодействий между соседними аминокислотными остатками делается заключение о склонности участка полипептидной цепи образовывать структуру α -спирали, β -тяжи или кольца. Сети были обучены на наборе последовательностей из 3676 уникальных белковых доменов. Валидация предиктора была проведена на наборе из последовательностей 1629 белковых доменов [Kim *et al.*, 2009].

Предиктор MILAMP разработан для решения сразу нескольких задач: определения является ли исследуемая последовательность амилоидом, поиск амилоидогенных участков и влияние точечных мутаций на амилоидогенность белка. Для этого разработчики создали модель многовариантного машинного обучения (Multiple-instance learning). Эта модель берет в расчет исключительно сочетания аминокислот в рамках гексапептида. Для обучения модели авторы создали несколько разных наборов данных: набор из 304 гексапептидов, амилоидогенных и неамилоидогенных, а также набор из 483 белков, включая амилоиды и контроль. Тестирование модели проводилось на наборе из 33 амилоидных белков и наборе из 53 мутантных белков с известным влиянием мутации на амилоидогенность [Munir *et al.*, 2019].

Предиктор FISH основывается на использовании методов машинного обучения. Авторы исходят из предположения что амилоидогенные участки обладают определенным паттерном распределения аминокислот в последовательности. Авторы подчеркивают, что амилоидогенность участка определяют не просто аминокислоты на конкретных позициях, а именно сочетания нескольких различных аминокислот в определенных позициях подпоследовательности. Модель обучалась на нескольких наборах данных, включая два набора гексапептидов, амилоидогенных и неамилоидогенных, и наборе из последовательностей переменной длины, также включающий амилоидогенную группу и контроль. Валидацию метода проводили на наборе из 436 пептидов, амилоидогенных и неамилоидогенных [Gasior, Kotulska, 2014].

Предиктор AmyloGram основывается на использовании методов машинного обучения. Предлагаемая модель построена на классификаторе случайного леса. Обучающая выборка для модели была создана с использованием N-грамм анализа на основе набора из 1430 пептидных последовательностей, амилоидогенных и неамилоидогенных. Валидацию метода проводили на двух наборах данных: набор из 424 пептидов, амилоидогенных и неамилоидогенных, а также наборе из 33 амилоидных белков [Burdukiewicz *et al.*, 2017].

Предиктор GAP разработан с целью не только идентифицировать амилоидогенные участки последовательности, но и отличать такие участки, от участков образующих аморфные β -агрегаты. Для этого авторы собрали данные о влиянии позиции аминокислот в рамках гексапептида на склонность последовательности к агрегации. Кроме этого, авторы на основе набора из 139 амилоидных гексапептидов и 168 гексапептидов образующих аморфные β -агрегаты рассчитали частоты встречаемости пар

аминокислотных остатков для обеих групп агрегатов. На основе этих факторов авторы вычислили термодинамическую потенциальную энергию для всех пар аминокислотных остатков. Предиктор, представляющий собой модель машинного обучения, использует эти значения при определении склонности участка последовательности к образованию агрегирующих структур и определении являются ли эти структуры потенциально амилоидными или аморфными. Валидацию метода проводили на наборе из 310 амилоидогенных пептидов [Thangakani *et al.*, 2014].

Предиктор RFAmyloid представляет собой модель машинного обучения, базирующуюся на алгоритме случайного леса. В качестве метрики для определения амилоидогенности последовательности предиктор использует данные о сочетаниях аминокислот в амилоидных белках и их физико-химические свойства. Для обучения и тестирования предиктора авторы создали набор из 165 амилоидогенных белков и 382 неамилоидогенных [Niu *et al.*, 2018].

Предиктор Pafig представляет собой модель машинного обучения, построенную на использовании метода опорных векторов. На основе базы данных AAindex авторы отобрали 41 физико-химических свойств аминокислот, наиболее влияющих на амилоидогенность последовательности. Модель обучалась и тестировалась на выборке из 2452 гексапептида, включая известные амилоидогенные последовательности и неамилоидогенные фрагменты белков [Tian *et al.*, 2009].

Предиктор PLAAC разработан для поиска в последовательности белка доменов с прион-подобным сочетанием аминокислот. Авторы рассчитали частоты встречаемости различных сочетаний аминокислот в прионах *Saccharomyces cerevisiae* имплементировали в данном предикторе с использованием скрытых марковских моделей. Данных о тестировании предиктора в посвященной ему научной работе не представлено [Lancaster *et al.*, 2014].

1.4.2 Предикторы, основанные на использовании индивидуальных свойств аминокислот

К этой категории были отнесены те предикторы, которые предсказывают амилоидогенность участков последовательности белка на основе данных об индивидуальных физико-химических свойствах аминокислот.

Предиктор AbAmyloid разработан для предсказания амилоидогенных свойств антител, в том числе, не только в рамках одной зародышевой линии, но на последовательностях из разных линий. Для предсказания амилоидогенности белка авторы отобрали три свойства полипептидной последовательности: сочетания аминокислот, сочетания дипептидов и физико-химические свойства аминокислот. На основе этих данных были построены модели машинного обучения с использованием алгоритма случайного леса (Random Forest). Каждая модель брала в расчет один, несколько или сразу все свойства, а обучалась и тестировалась на последовательностях антител сперва в рамках одной линии, а впоследствии на наборах из 12 зародышевых линий антител. Лучший результат среди всех тестов показала модель, использующая сочетания дипептидов и обученная на наборе из различных зародышевых линий. Именно эта модель и легла в основу предиктора [Liaw, Tung, Ho, 2013].

Предиктор Zyggregator анализирует амилоидогенность последовательности на основе значений набора физико-химических свойств индивидуальных аминокислот: склонность к образованию альфа-спиралей, склонность к образованию бета-листов, гидрофобность, заряд, гидрофобные паттерны и является ли аминокислота привратником (gatekeeper residue). Конечная предсказанная амилоидогенность последовательности сравнивается с аналогичными значениями в наборе случайно сгенерированных полипептидов той же длины, и, в случае если амилоидогенность исследуемого участка выше, он признается амилоидогенным. Для подтверждения работоспособности своего предиктора, в ряде работ авторы успешно использовали этот метод для поиска амилоидогенных участков в известных патологических амилоидах, таких как Tau, A β и α -синуклеин [Tartaglia, Vendruscolo, 2008].

Предиктор FoldAmyloid предсказывает амилоидогенные участки последовательности белка на основе двух физико-химических параметров аминокислот: ожидаемая вероятность образования водородных связей и ожидаемая плотность упаковки аминокислотных остатков. Значения этих параметров для каждой из 20 протеиногенных аминокислот были вычислены авторами на основе базы данных пространственных структур 3769 белков. Валидация метода была произведена на наборе из 407 известных пептидов, как амилоидогенных, так и не-амилоидогенных и четырех глобулярных белках [Garbuzynskiy, Lobanov, Galzitskaya, 2009].

Предиктор BETASCAN производит поиск подпоследовательностей полипептидной цепи с высокой долей вероятности образующих β -тяжи. Для этого вся последовательность рассматривается как набор подпоследовательностей конечной длины от 2 до k . Параметр k

задается вручную, но по умолчанию $k=13$, как максимальная длина среди известных последовательностей образующих β -тяжи. Далее предиктор для каждой такой пары подпоследовательностей высчитывает вероятность того, что они образуют пару параллельных тяжей. Для этого предиктор берет в расчет экспериментально полученные данные о склонности различных пар аминокислотных остатков образовывать водородные связи. Среди всех полученных значений, наиболее значимые отбираются при помощи алгоритма поиска восхождения к вершине (Hill-climbing algorithm). Алгоритм тестировали на 34 фрагментах белков, для которых экспериментально было подтверждено формирование вторичных структур [Bryan *et al.*, 2009].

Предиктор PASTA 2.0 в рамках исследуемой полипептидной последовательности ищет короткие участки, потенциально стабилизирующие кросс-бета структуру амилоидов. Для этого предиктор рассчитывает вероятность для пары аминокислотных остатков в подпоследовательности образовать водородные связи. Вывод об этом делается на основе экспериментально полученных параметров энергетической функции для различных протеиногенных аминокислот. Валидацию метода проводили на двух наборах данных: набор из 424 пептидов, амилоидогенных и неамилоидогенных, и наборе из 33 амилоидных белков [Walsh *et al.*, 2014].

Принцип работы предиктора AGGRESCAN исходит из предположения, что амилоидогенные участки являются решающим фактором образования амилоидных фибрилл и что амилоидогенность всей последовательности зависит от физико-химических свойств аминокислот в рамках таких участков. Для предсказания таких участков авторы получили провели серию опытов с точечными мутациями над гидрофобным ядром белка А β . Полученные данные о влиянии отдельных протеиногенных аминокислот на склонность белка к агрегации и легли в основу работы предиктора. Валидацию предиктора проводили на 18 амилоидогенных белках, для части амилоидогенность была показана исключительно *in vitro* [Conchillo-Solé *et al.*, 2007].

1.4.3 Предикторы основанные на композиционном подходе

В эту категорию были помещены те предикторы, которые рассматривают полипептид как набор подпоследовательностей конечной длины и рассчитывает амилоидогенность каждого участка на основе данных о взаимовлиянии аминокислот.

Принцип работы предиктора Waltz исходит из утверждения что минимальной амилоидогенной последовательностью является гексапептид. Waltz рассматривает исследуемую полипептидную последовательность как набор шестичленных подпоследовательностей. Авторы на основе информации из базы данных AmylHex и с помощью лабораторных методов идентификации амилоидов создали набор из более чем 200 гексапептидов, как амилоидогенных, так и не амилоидогенных. На основе этих данных была создана позиционная матрица оценок, которая рассчитывала влияние отдельной аминокислоты на амилоидогенность участка в зависимости от позиции аминокислоты в этом гексапептиде. В основе работы Waltz, помимо этой матрицы используются данные о 19 физико-химических показателях индивидуальных аминокислот, а также позиционная матрица псевдоэнергии аминокислот, полученная из данных структурного моделирования амилоидной фибриллы «GNNQQNY». Для валидации работоспособности предиктора на основе базы данных UniProt были проанализированы более 9000 белков. Свыше 21000 гексамеров были определены как амилоидогенные. Случайно отобранные 30 гексамеров синтезировали *in vitro*. После периода инкубации в фосфатном буфере 80% полученных полипептидов образовали амилоиды амилоиды. Сходным образом алгоритм был протестирован на N-домене прионного белка sup35 и наборе из 35 функциональных амилоидов [Maurer-Stroh *et al.*, 2010].

В основе TANGO лежит статистическая модель, которая рассчитывает склонность участков к образованию различных структур: α -спирали, β -слои и β -агрегаты. Принцип работы TANGO также исходит из нескольких предположений: в структурированных β -слоистых агрегатах основной вторичной структурой являются β -тяги; участки образующие β -агрегаты стремятся к образованию максимально возможного числа водородных связей и комплементарные заряды в каждой подпоследовательности способствуют наиболее предпочтительным электростатическим взаимодействиям, а общий заряд пептида вне и внутри подпоследовательности препятствуют образованию агрегата. Таким образом, предиктор на основе данных о физико-химических основах образования β -слоев и беря в расчет внешние физико-химические условия (рН, концентрация вещества, ионов и трифлурэтанола в растворе), делает предсказания касательно склонности участков последовательности образовывать различные вторичные структуры и агрегировать. Предиктор тестировался на двух наборах данных: первый набор из 179 пептидов, полученных из 21 различных белков, и второй набор из 71 пептида

из ассоциированных с заболеваниями человека белков, включая лизоцим и β 2-микроглобулин [Fernandez-Escamilla *et al.*, 2004].

Предиктор ZipperDB использует метод 3D-профилирования для предсказания участков последовательности, которые могут образовывать «стерическую застежку». Предиктор разработан на основе данных полученных из атомной структуры кристаллов коротких амилоидогенных полипептидов «GNNQQNY» и «NNQQNY». Исследуемая последовательность сравнивается с трехмерной моделью «NNQQNY» и, если подпоследовательность может согласно расчетам образовывать сходную структуру застежки с минимальными затратами энергии, она признается амилоидогенной. Предсказательную способность предиктора тестировали на РНКазе *A Bos taurus*: те фрагменты последовательности, которые были определены предиктором как амилоидогенные, были синтезированы в лаборатории, и их способность образовывать амилоидные фибриллы была подтверждена *in vitro* [Goldschmidt *et al.*, 2010].

Предиктор PAPA разработан для предсказания прионных свойств у аспарагин-глутамин-обогащенных белков. Авторы провели серию экспериментов с точечным замещением различных аминокислот в прион-образующих доменов известных прионных белков, получили таким образом данные о влиянии отдельных аминокислот и их сочетаний на образовании прионов. Совместив полученные данные с предиктором FoldIndex [Prilusky *et al.*, 2005], который предсказывает внутреннюю склонность белковой последовательности к фолдингу, авторы получили предиктор способный предсказывать прионные свойства белков. Предиктор тестировался на наборе из 100 аспарагин-глутамин-обогащенных белков [Toombs *et al.*, 2012].

Алгоритм DIANA создан для поиска аспарагин-/глутамин-насыщенных участков последовательности белка. Данный метод исследует последовательность при помощи «скользящего окна» длиной в 80 аминокислотных остатков. Алгоритм отмечает участок как аспарагин-/глутамин-насыщенный если он содержит хотя бы 30 аспарагин и/или глутамин аминокислот. Таким образом были исследованы геномы трех эукариотических организмов, шести термофильных архей, двух термофильных и 20 мезофильных бактерий [Michelitsch, Weissman, 2000].

Предиктор LPS нацелен на поиск последовательностей с низкой вероятностью возникновения (Low-probability-subsequences/compositionally-biased-regions). Предиктор вычисляет частоты возникновения отдельных аминокислот во всей последовательности белка и в подпоследовательностях переменной длины. Далее алгоритм на основе

вычислений идентифицирует подпоследовательности с низкой вероятностью возникновения. В качестве примера использования такого подхода, авторы просчитали частоты возникновения аминокислот для 4 известных прионов и на основе этих результатов проанализировали шесть эукариотических протеомов [Harrison, Gerstein, 2003].

Предиктор SARP является усовершенствованной версией предиктора LPS. Сохранив тот же подход к определению последовательностей с низкой вероятностью возникновения, авторы оптимизировали процесс поиска, многократно сократив необходимое для расчетов время. Предиктор тестировали на наборе из 1000 белков *Saccharomyces cerevisiae*. Кроме этого предиктор использовали при исследовании протеомов растений и таким образом была предсказана амилоидогенность запасных белков с доменами Cupin [Antonets, Nizhnikov, 2017], что впоследствии было подтверждено в экспериментальном анализе амилоидных свойств запасного белка семян гороха *Pisum sativum L.* вицилина [Antonets *et al.*, 2020].

Предиктор ArchCandy разработан с целью поиска участков последовательности способных образовывать β -арки – белковые структуры, образованные объединенными β -тяжами. Для решения этой задачи авторы на основе исследования структур β -амилоида, белка CA150 человека и β 2-микроглобулина, отобрали набор факторов взаимодействия аминокислот в подпоследовательности, которые влияют на конечную пространственную структуру белка: стерическое напряжение, взаимодействие заряженный аминокислотных остатков, электростатическое отталкивание, плотность и насыщенность водородными связями боковых цепей молекул, влияние пролина и длину последовательности. Метод тестировали на наборе из 23 амилоидов и амилоидогенных пептидов, и наборе из 50 неамилоидогенных белков [Ahmed *et al.*, 2015].

1.4.4 Консенсусный предиктор

В эту категорию был вынесен предиктор, который в своей работе использует сразу несколько независимых подходов и дает консенсусный вывод на основе их работы.

В состав консенсусного предиктора Amyl2Pred входит 11 различных вычислительных методов: Aggrescan, AmyloidMutants, Amyloidogenic Pattern, Average Packing Density, Beta-strand contiguity, Hexapeptide Conformational Energy, NetCSSP, Pafig, SecStr (Possible Conformational Switches), Tango и Waltz. На основе результатов работы

всех подходов Amyl2Pred делает консенсусное предсказание касемо наличия амилоидогенных участков в последовательности белка. При этом предиктор позволяет определять состав используемых методов и их настройки. Авторы протестировали метод на наборе из 33 амилоидных белков и консенсусное предсказание оказался точнее индивидуальных результатов для каждого метода [Tsolis *et al.*, 2013].

Таким образом, проведенный нами анализ литературных данных, позволил классифицировать существующие биоинформатические подходы к предсказанию амилоидогенных свойств белков на четыре группы согласно принципам их работы. Общая информация по этим подходам представлена в Таблице 1.

Таблица 1. Сводная характеристика методов предсказания амилоидогенных свойств белков

Название	Используемый подход	Принцип работы	Валидация метода	Доступность (на 24.05.2021)
Appnp	Машинное обучение	С помощью машинного обучения были выделены семь физико-химических и биохимических свойств отдельных аминокислот, которые связывают с амилоидогенностью белка. На этих параметрах и основана работа непосредственно предиктора	Обучение нейросетей происходило на датасете из 6-членных пептидов, включающие как амилоидные, так и неамилоидные последовательности. Проверка результатов осуществлялась на аналогичном наборе данных, но включающем последовательности большей длины.	http://cran.r-project.org/web/packages/appnp/index.html

Название	Используемый подход	Принцип работы	Валидация метода	Доступность (на 24.05.2021)
Amyl2Pred	Консенсусный предиктор	Amyl2Pred представляет собой алгоритм, включающий в себя набор прочих методов предсказания белковой структуры: Aggrescan, AmyloidMutants, Amyloidogenic Pattern, Average Packing Density, Beta-strand contiguity, Hexapeptide Conformational Energy, NetCSSP, Pafig, SecStr (Possible Conformational Switches), Tango, Waltz.	Набор из 33 амилоидогенных белков	http://aias.bio1.uoa.gr/AMYLPRED2/
AbAmyloid	Машинное обучение	На основе ряда свойств (сочетание АК, сочетание дипептидов и физико-химические св-ва) данный алгоритм предсказывает амилоидогенность антител.	Набор последовательностей легких цепей антител из 12 линий.	http://iclab.life.nctu.edu.tw/abamyloid/ (Веб-сервер не доступен)
Zyggregator	Свойства индивидуальных аминокислот	Данный предиктор предсказывает наличие амилоидогенных участков на основе набора физико-химических свойств отдельных аминокислот в подпоследовательности	В ряде работ авторы использовали этот метод относительно для поиска амилоидогенных участков в известных амилоидах, таких как Tau, A β , α -Synuclein и т.д.	https://www-cohsoftware.ch.cam.ac.uk/index.php/zyggregator (Веб-сервер не доступен)

Название	Используемый подход	Принцип работы	Валидация метода	Доступность (на 24.05.2021)
WALTZ	Композиционный подход	Алгоритм WALTZ рассматривает все возможные шестичленные пептиды в рамках последовательности белка. Пользуясь собственной матрицей оценок (алгоритм учитывает вид АК и ее положение в рамках гексапептида) и учитывая 19 различных физических свойств АК, WALTZ предсказывает амилоидогенность белка.	На основе базы данных UniProt были проанализированы более 9000 белков. Свыше 21000 гексапептидов были определены как амилоид-формирующие. Случайно отобранные 30 гексапептидов синтезировали in vitro, 80% из которых образовали амилоиды. Схожим образом алгоритм был протестирован на N-домене прионного белка sup35 и наборе из 35 функциональных амилоидов.	https://waltz.switchlab.org/
Net-CSSP	Машинное обучение	Предиктор является предварительно обученной нейросетью, в основе обучения которой лежит алгоритм CSSP, который определяет склонность агрегации 7-членных подпоследовательностей белка на основе расчета наиболее энергетически выгодной конформации и третичных взаимодействий атомов.	Метод тестировался на наборе из 22707 полипептидов, полученных из 1629 белковых доменов.	http://cssp2.sookmyung.ac.kr/

Название	Используемый подход	Принцип работы	Валидация метода	Доступность (на 24.05.2021)
FoldAmyloid	Свойства индивидуальных аминокислот	Метод учитывает две характеристики АК: вероятность формирования водородных связей и плотность упаковки (вольный перевод). Высокие значения характеристик в подпоследовательности белка говорит о наличии амилоидогенного участка.	Метод тестировался на 407 известных пептидах (амилоидогенных и не-амилоидогенных) и четырех глобулярных белках.	http://bioinfo.protres.ru/fold-amyloid/
BETASCAN	Свойства индивидуальных аминокислот	Метод оценивает вероятность образования подпоследовательности белка длиной от 2 до k (k=13 по умолчанию) формировать вторичную структуру (бета-тяжи). Оценка делается на основе ожидаемой вероятности формирования водородных связей между АК.	Алгоритм тестировался на 34 фрагментах белков, для которых экспериментально было подтверждено формирование вторичных структур.	http://cb.csail.mit.edu/cb/betascan/ (Веб-сервер доступен, но не работает)
PASTA 2.0	Свойства индивидуальных аминокислот	Метод определяет какая часть последовательности стабилизирует кросс-бета структуру амилоида, на основе оценки потенциальной энергии водородных связей пар АК остатков.	Метод тестировался на двух наборах данных: 1) Набор из 424 пептидов (агрегирующих и не-агрегирующих) 2) Набор из 33 амилоидных белков	http://old.protein.bio.unipd.it/pasta2/index.html

Название	Используемый подход	Принцип работы	Валидация метода	Доступность (на 24.05.2021)
MILAMP	Машинное обучение	Предиктор представляет собой нейросеть обученную на данных о различных гексапептидах: амилоидогенных и неамилоидогенных. Данный метод предсказывает является ли белок амилоидном, идентифицирует амилоидогенные участки и предсказывает влияние точечных мутаций на амилоидогенность белка.	Модель обучалась на наборах амилоидогенных и неамилоидогенных гексапептидов и белков. Валидация проходила на наборах из 33 амилоидных и 53 мутантных белков.	http://milamp.pythonanywhere.com/
FISH	Машинное обучение	Предиктор построен на алгоритме машинного обучения и берет в расчет сочетание АК в рамках подпоследовательности определенной длины (5-6 АК остатков).	Алгоритм обучен на трех наборах последовательностей из разных источников (два набора гексапептидов и набор из последовательностей переменной длины) Предиктор тестировался на наборе из 436 пептидов длиной до 10 аминокислотных остатков.	http://compres-lin.iar.pwr.edu.pl/fishInput/ (Веб-сервер доступен, но не работает)
AGGRESKAN	Свойства индивидуальных аминокислот	Предиктор делает упор на роль коротких последовательностей в амилоидогенности белка. Авторы на основе исследований белка $A\beta$ <i>in vivo</i> создали собственную систему оценок влияния отдельных АК в подпоследовательностях на амилоидогенность.	18 амилоидогенных белков (для части амилоидогенность была показана исключительно <i>in vitro</i>)	http://bioinf.uab.es/aggrescan/

Название	Используемый подход	Принцип работы	Валидация метода	Доступность (на 24.05.2021)
Amylogram	Машинное обучение	Предиктор работает на основе предварительно обученной нейронной сети. При разработке сети авторы учли и 17 физико-химических свойств АК и особенности композиции различных АК в гексамерах.	Использовался тот же набор данных что и для PASTA 2.0	http://biongram.biotech.uni.wroc.pl/AmyloGram/
GAP	Машинное обучение	На основе имеющихся данных о сочетаниях АК в гексапептидах, которые увеличивают склонность белка к созданию агрегирующих фибрилл (в т.ч. амилоидов), авторы рассчитали потенциальную термодинамическую энергию для пар АК в белковых фибриллах. К полученным результатам они применили методы машинного обучения и на основе этого создали данный предиктор	55 гексапептидов и 310 пептидов переменной длины.	https://www.itm.ac.in/bioinfo/GAP/
RFAmyloid	Машинное обучение	В основе метода лежит машинное обучения, для которого использовались данные о сочетании АК и их физико-химические свойства	Для обучения и тестирования предиктора авторы создали набор 165 амилоидных и 382 неамилоидных белков.	http://server.malab.cn/RFAmyloid/

Название	Используемый подход	Принцип работы	Валидация метода	Доступность (на 24.05.2021)
TANGO	Композиционный подход	Данный алгоритм на основе стабильности белка и физико-химических свойств последовательности определяет участки, участвующие в формировании агрегирующих вторичных бета-структур. Авторы отмечают, что алгоритм не определяет непосредственно амилоидогенные участки, но существует корреляция между определением участков бета-структур и амилоидами.	250 пептидов, включая несколько прионных белков.	http://tango.crg.es/
ZipperDB	Композиционный подход	Предиктор делает упор на определение возможности белка формировать “стерическую молнию” – основу для формирования амилоидного фибрилла	Предиктор тестировался на фрагментах коровьей РНКазы А	http://services.mbi.ucla.edu/zipperdb/
Pafig	Машинное обучение	Предиктор представляет собой нейросеть обученную на данных о гексапептидах и физико-химических свойствах отдельных АК.	Набор из 2452 гексапептида	http://www.mobiinform.cn/pafig (Веб-сервер не доступен)
PLAAC	Машинное обучение	Данный предиктор использует метод скрытых марковских моделей для поиска прион-подобных подпоследовательностей в белке.	-	http://plaac.wi.mit.edu/

Название	Используемый подход	Принцип работы	Валидация метода	Доступность (на 24.05.2021)
PAPA	Композиционный подход	Предиктор берет в расчет сочетание АК в последовательности и влияние каждой индивидуальной АК на склонность белка образовывать прион.	Предиктор тестировался на наборе из ста белков дрожжей композиционно сходных с прионными.	https://combi.cs.colostate.edu/supplements/papa/
LPS	Композиционный подход	Данный предиктор определяет участки последовательностей с низкой вероятностью возникновения (Low-probability-subsequences/compositionally-biased-regions). Авторы отдельно отмечают возможность поиска Q/N-насыщенных регионов, которые связывают с амилоидогенезом.	6 эукариотических протеомов, включая два вида дрожжей.	https://doi.org/10.1186/gb-2003-4-6-r40 , (Отсутствует программная реализация)
SARP	Композиционный подход	Данный предиктор является улучшенной и оптимизированной версией предиктора LPS и использует тот же подход.	Набор из 1000 случайных белков дрожжей. Кроме этого, предиктор использовался при исследовании протеомов растений и таким образом была предсказана амилоидогенность запасных белков гороха, что впоследствии было подтверждено экспериментально на вицилине <i>Pisum sativum</i> .	https://doi.org/10.4137/EVO.S12299 (Доступ к предиктору по запросу),
DIANA	Композиционный подход	Данный предиктор определяет участки последовательности с повышенным содержанием Q/N аминокислот.	Геномы трех эукариот, шесть термофильных архей, двух термофильных бактерий и 20 мезофильных бактерий.	http://itsa.ucsf.edu/~mmichel/DIANA (Веб-сервер не доступен)

Название	Используемый подход	Принцип работы	Валидация метода	Доступность (на 24.05.2021)
ArchCandy-1.0	Композиционный подход	Предиктор основывается на предсказании структуры белка известной как β -арка. На основе имеющихся данных об этой структуре, авторы создали систему оценки взаимодействия между соседними АК и их вероятность образовать эту структуру.	Наборы из 23 амилоидов и амилоидогенных пептидов, а также 50 неамилоидных белков.	https://bioinformatics.crbm.cnrs.fr/index.php?route=tools&tool=7 (Доступ к предиктору по запросу)

2. Материалы и методы

2.1. Данные

Для тестирования предикторов на основе базы данных AmyPro [Varadi *et al.*, 2018] и научной литературы были отобраны белки, для которых наличие амилоидной конформации подтверждено в исследованиях *in vivo*. Таким образом было сформировано три подгруппы амилоидов: 38 патологических, 37 функциональных и 11 прионов.

Для создания контрольной группы неамилоидных белков были выбраны несколько организмов, для которых в других работах был проведен протеомный скрининг методом PSIA [Nizhnikov *et al.*, 2014]. Сюда вошли: *Pisum sativum* [Antonets *et al.*, 2020], *Rattus norvegicus* [Sopova *et al.*, 2019], *Saccharomyces cerevisiae* [Nizhnikov *et al.*, 2016], *Escherichia coli* [Antonets *et al.*, 2016] и *Rhizobium leguminosarum* [Kosolapova *et al.*, 2019]. Контрольные образцы отбирали в две подгруппы согласно медианной и максимальной длинам белковых последовательностей в экспериментальной группе: от 0 до 261 и от 262 до 1565 аминокислотных остатков. Исключив из протеомов отобранных организмов все потенциальные амилоиды, случайным образом были отобраны контрольные неамилоидные белки. Случайность выборки была обеспечена специальной программой на языке программирования Python, с использованием пакетов «Random» и «Bioservices» [Cokelaer *et al.*, 2013]. Полные последовательности белков были получены из базы данных UniProt [Bateman *et al.*, 2017].

2.2. Тестирование предикторов

Для тестирования предсказательной способности были выбраны биоинформатические методы предсказания амилоидогенных свойств белков, которые находятся в свободном доступе. Таким образом мы протестировали 14 предикторов. Все предикторы использовались с настройками, рекомендуемыми в оригинальных статьях. При отсутствии прямых указаний использовались настройки по умолчанию.

Amyl2Pred: Консенсусный предиктор использовался с использованием всех доступных методов.

PAPA: Настройки отсутствуют.

TANGO: Использовалась оффлайн-версия предиктора для ОС Windows со следующими настройками: Nterm protected – no protection; Cterm protected – no protection; Temperature[K] – 298; Ionic strength [M] – 0.02; Concentration [M] – 1;

Waltz: Threshold – Best overall performance; pH – 7.0;

AGGRESCAN: Настройки отсутствуют.

FoldAmyloid: Averaging frame – 5; Threshold 21.4;

PASTA 2.0: Top pairing energies – 20; Large-scale – no; Energy threshold – -5;

AmyloGram: Настройки отсутствуют.

Appnn: Предиктор имплементирован в виде одноименного пакета для языка программирования R. Тестирование производилось с помощью IDE RStudio [RStudio Team, 2021]. Настройки отсутствуют.

GAP: Настройки отсутствуют.

MILAMP: Настройки отсутствуют.

NET-CSSP: Method for CSSP calculation – Single network

PLAAC: Core length – 60; Relative weighting of background probabilities – 100;

Rfamilyoid: Настройки отсутствуют.

Белок считался амилоидогенным согласно предсказанию предиктора, если в результатах работы на это было прямое указание, или если в последовательности белка был предсказан хотя бы один амилоидогенный участок.

2.3. Обработка полученных результатов

На основе результатов работы предикторов были построены ROC-кривые и вычислены специфичность и чувствительность предсказания, а также вычислены значения площади под кривой (AUC) для всех кривых. Все численные результаты обработки данных и графики были получены при помощи пакета «ROCR» [Sing *et al.*, 2005] на языке программирования R.

3. Результаты и обсуждение

Результаты обработки выборок амилоидных и неамилоидных белков при помощи разных предикторов представлены в таблице 2 и на рисунках 1-3.

Таблица 2. Оценка предсказательной способности предикторов

	Amyl2Pred	PAPA	TANGO	Waltz	AGGRESCAN	FoldAmyloid	PASTA 2.0
Чувствительность	100%	12.8%	97%	90.7%	100%	98.8%	93%
Специфичность	2.2%	100%	1.1%	14.4%	1.1%	3.3%	12.2%
AUC	0.511	0.564	0.494	0.526	0.5	0.511	0.526
	AmyloGram	Appnn	GAP	MILAMP	NET-CSSP	PLAAC	Rfamilyoid
Чувствительность	100%	100%	98.8%	100%	97.7%	24.4%	100%
Специфичность	1.1%	2.2%	0%	0%	3.3%	94.4%	0%
AUC	0.506	0.511	0.5	0.5	0.511	0.589	0.5

На рисунке 1 представлены ROC-кривые для предикторов, основанных на использовании индивидуальных свойств аминокислот и для консенсусного предиктора Amyl2Pred. Графики всех методов фактически представляют собой прямые, что говорит о равном соотношении верно предсказанных амилоидов и ошибок первого рода, когда белки из контрольной группы предсказываются как амилоиды.

Предикторы основанные на индивидуальных св-ах АК

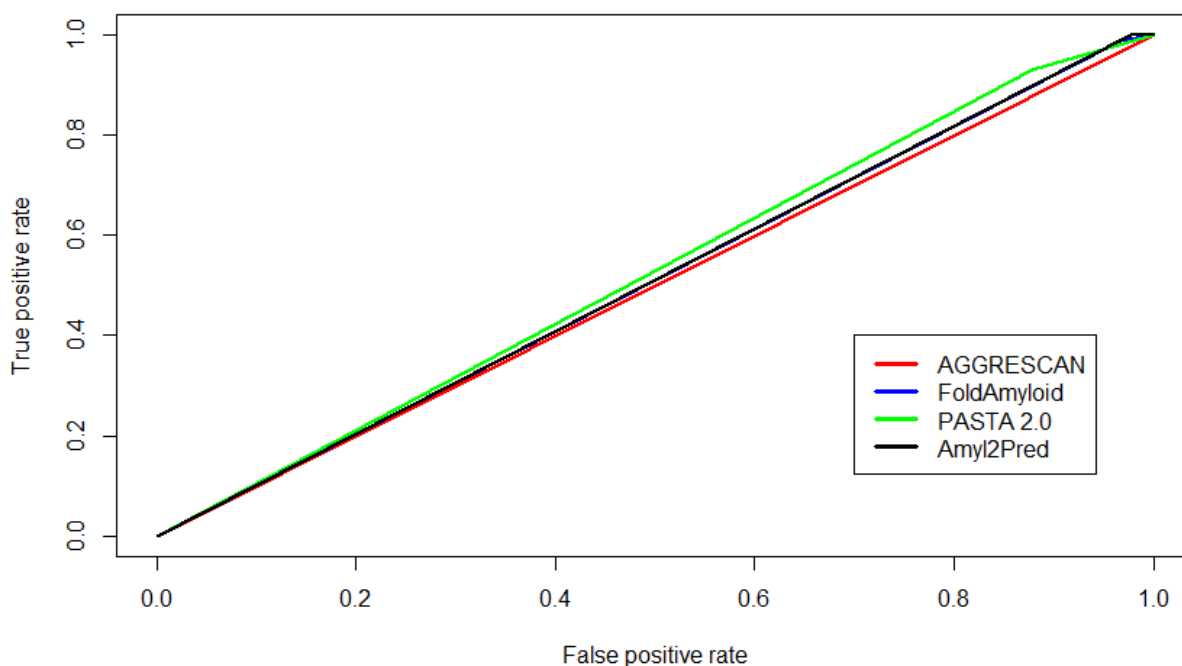


Рис.1. ROC-кривая для предикторов основанных на индивидуальных свойствах аминокислот. На оси абсцисс обозначена доля ошибок первого рода, а на оси ординат обозначена доля верно предсказанных амилоидов от общего числа предсказаний.

На рисунке 2 представлены ROC-кривые для предикторов основанных на машинном обучении и коненсусного предиктора Amyl2Pred. Аналогично предыдущему подходу, ROC-кривые для предикторов представляют собой даже еще более ярко выраженные прямые, что также говорит нам равном соотношении между долей верно предсказанных амилоидных белков и ложно-положительных прогнозов. Из общей картины выделяется результат работы предиктора PLAAC, который показывает несколько более хороший результат и долю верных предсказаний более высокую, чем ошибок первого рода.

Предикторы основанные на машинном обучении

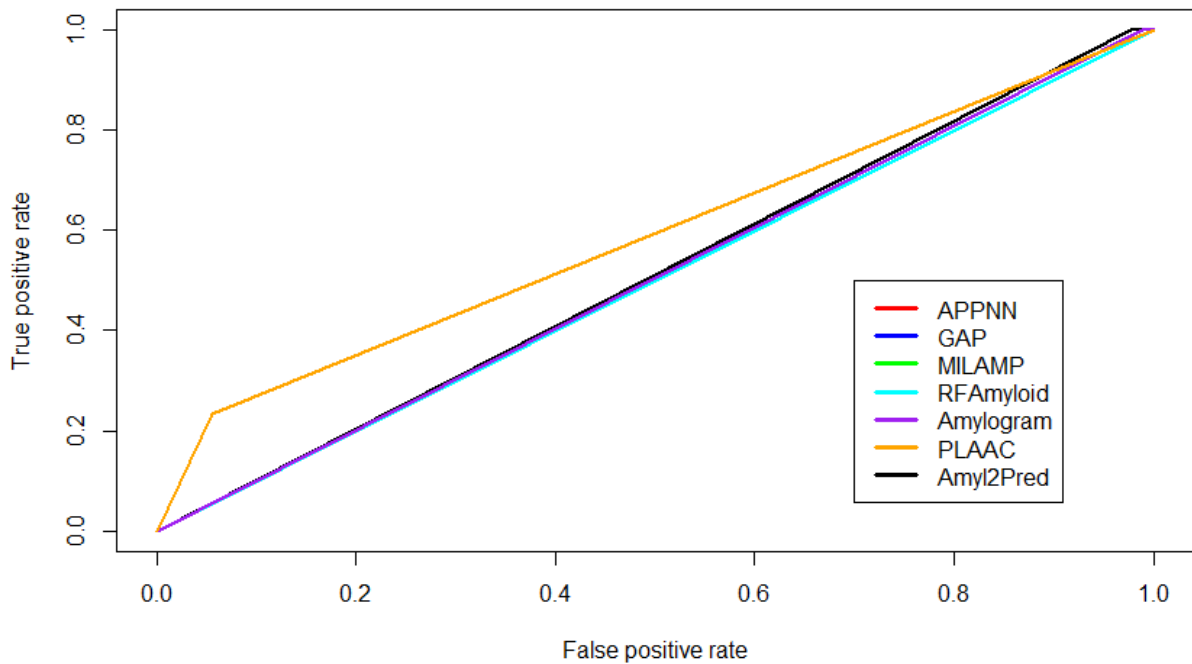


Рис.2. ROC-кривая для предикторов основанных на машинном обучении. На оси абсцисс обозначена доля ошибок первого рода, а на оси ординат обозначена доля верно предсказанных амилоидов от общего числа предсказаний.

На рисунке 3 представлены ROC-кривые для предикторов основанных на композиционном подходе и для консенсусного предиктора Amyl2Pred. Общий результат предсказания амилоидных свойств для этого подхода не отличается от прочих подходов и большинство предикторов в равной мере верно определяют амилоиды в экспериментальной группе и делают ошибочные предсказания касательно контрольной группы. В этой подгруппе предикторов сравнительно лучший результат показал алгоритм PAPA, с меньшей долей ошибок первого рода от общего числа предсказаний.

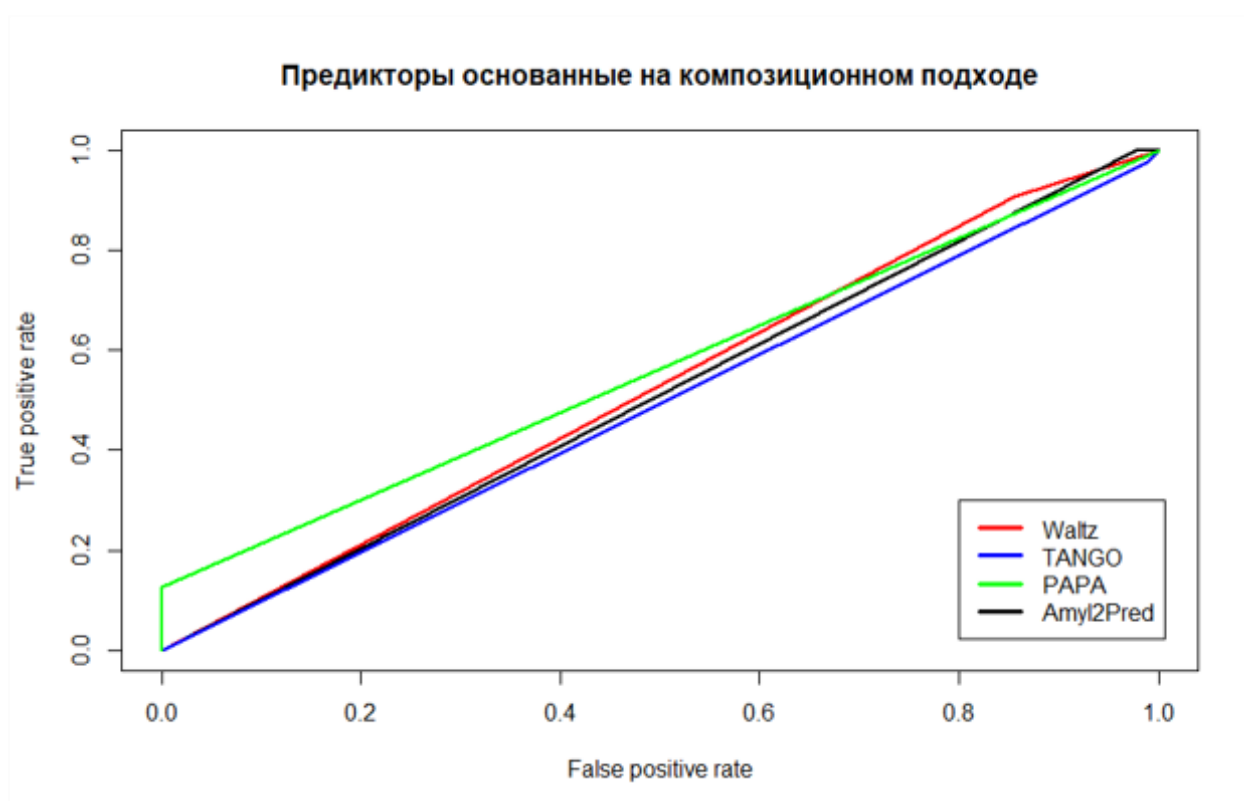


Рис.3. ROC-кривая для предикторов основанных на композиционном подходе. На оси абсцисс обозначена доля ошибок первого рода, а на оси ординат обозначена доля верно предсказанных амилоидов от общего числа предсказаний.

На основе результатов оценки предсказательной способности тестируемых предикторов, которые представлены в таблице 2, можно увидеть, что для всех предикторов значение AUC, которую можно интерпретировать как качественную характеристику предсказательной способности предиктора, близко к значению 0,5, что говорит об общей неспособности сделать различие между экспериментальной и контрольной группами. Превалирующее большинство предикторов демонстрирует крайне высокую степень чувствительности одновременно с низкой степенью специфичности. Это выражается в том, что эти предикторы определяют все или почти все анализируемые белки как амилоиды. Причина этого может крыться в следующем: во-первых, при разработке большинства предикторов были использованы данные о амилоидогенных пептидах, но не полных последовательностях белков. Такие предикторы не учитывают возможного взаимодействия между амилоидогенными и неамилоидогенными участками последовательности белка. Во-вторых, даже в тех случаях, когда при разработке использовались данные о полноразмерных амилоидах, количество таких данных мало, что не позволяет вывести достаточно общие для всех амилоидов закономерности, по которым

предиктор мог бы однозначно отличить амилоидогенную последовательность от неамилоидогенной.

Отдельно стоит отметить предикторы PLAAC и PAPA: от остальных предикторов их отличает низкая чувствительность предсказания одновременно с высокой специфичностью. На практике это выражается в предсказании амилоидогенных свойств исключительно у подгруппы прионов, когда почти все остальные белки, как из экспериментальной, так и из контрольной групп были определены как неамилоидные. Оба этих предиктора разрабатывались на основе данных о прионных белках и для поиска прион-подобных последовательностей. В этом смысле предикторы справляются со своей задачей, но в рамках поставленной нами задачи они также демонстрируют неспособность отличить амилоидный белок от неамилоидного.

Таким образом, проведенное тестирование показывает, что на текущем этапе своего развития биоинформатические подходы к предсказанию амилоидогенных последовательностей малоприменимы к исследованиям полноразмерных белков, например, в рамках исследования протеомов организмов. Большинство предикторов разрабатывали и тестировали для исследований коротких амилоидогенных участков белков, и они больше применимы именно для решения подобных задач. Кроме этого, работа всех рассмотренных нами предикторов сосредоточена вокруг последовательности белка и практически не учитывает возможное влияние внешних факторов на структуру белка, например, пост-трансляционных модификаций и уровня продукции. Дальнейшее развитие предсказательной способности биоинформатических методов можно связать с накоплением данных о существующих в природе амилоидах, образуемых полноразмерными белками *in vivo*; исследовании взаимного влияния амилоидогенных и неамилоидогенных участков последовательностей соответствующих белков; роли внешних факторов в жизненном цикле амилоидов и поиске общих закономерностей формирования кросс-бета структуры.

4. Выводы

На основе результатов нашего исследования можно сделать следующие выводы:

1. Существующие предикторы амилоидов используют в своей работе один из четырех подходов: машинное обучение, композиционный подход, использование индивидуальных свойств аминокислот или консенсусное применение сразу нескольких подходов.
2. Почти все протестированные предикторы вне зависимости от используемого подхода обладают высокой чувствительностью и низкой специфичностью предсказания.
3. Предикторы, которые были разработаны для предсказания прионных свойств, отличаются от остальных предикторов низкой чувствительностью и высокой специфичностью предсказания.
4. Все 14 протестированных предикторов неспособны отличить на основе полной последовательности амилоидные белки от неамилоидных. Вероятно, это связано с тем, что эти предикторы были разработаны на основе данных об амилоидогенных участках, а не полноразмерных амилоидных белках.
5. Для дальнейшего повышения эффективности методов предсказания амилоидогенных белков необходимо накопление экспериментальных данных о полноразмерных белках, проявляющих амилоидные свойства *in vivo*.

Список источников литературы

1. Ahmed A. B. *et al.* A structure-based approach to predict predisposition to amyloidosis // *Alzheimer's Dement.* 2015. Т. 11. № 6. С. 681–690.
2. Alberti 2009. Suppl 1 // *Cell.* 2009. Т. 137. № 1. С. 146–158.
3. Almeida Z. L., Brito R. M. M. Structure and aggregation mechanisms in amyloids // *Molecules.* 2020. Т. 25. № 5.
4. Antonets K. S. *et al.* Proteomic analysis of *Escherichia coli* protein fractions resistant to solubilization by ionic detergents // *Biochem.* 2016. Т. 81. № 1. С. 34–46.
5. Antonets K. S. *et al.* Accumulation of storage proteins in plant seeds is mediated by amyloid formation. , 2020.
6. Antonets K. S., Nizhnikov A. A. Predicting amyloidogenic proteins in the proteomes of plants // *Int. J. Mol. Sci.* 2017. Т. 18. № 10.
7. Baldwin A. J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation // *J. Am. Chem. Soc.* 2011. Т. 133. № 36. С. 14160–14163.
8. Bateman A. *et al.* UniProt: The universal protein knowledgebase // *Nucleic Acids Res.* 2017. Т. 45. № D1. С. D158–D169.
9. Baxa U. *et al.* Filaments of the Ure2p prion protein have a cross- β core structure // *J. Struct. Biol.* 2005. Т. 150. № 2. С. 170–179.
10. Bolton D. C. *et al.* Isolation and structural studies of the intact scrapie agent protein // *Arch. Biochem. Biophys.* 1987. Т. 258. № 2. С. 579–590.
11. Bryan A. W. *et al.* Betascan: Probable β -amyloids identified by pairwise probabilistic analysis // *PLoS Comput. Biol.* 2009. Т. 5. № 3.
12. Burdukiewicz M. *et al.* Amyloidogenic motifs revealed by n-gram analysis // *Sci. Rep.* 2017. Т. 7. № 1. С. 1–10.
13. Chernova T. A., Wilkinson K. D., Chernoff Y. O. Prions, chaperones, and proteostasis in yeast // *Cold Spring Harb. Perspect. Biol.* 2017. Т. 9. № 2. С. 1–18.
14. Chiti F., Dobson C. Amyloid Formation, Protein Homeostasis, and Human Disease: A Summary of Progress Over the Last Decade // *Annu. Rev. Biochem.* 2017. Т. 86. № 1. С. 1–42.

15. Cokelaer T. *et al.* BioServices: A common Python package to access biological Web Services programmatically // *Bioinformatics*. 2013. T. 29. № 24. C. 3241–3242.
16. Conchillo-Solé O. *et al.* AGGREGSCAN: A server for the prediction and evaluation of «hot spots» of aggregation in polypeptides // *BMC Bioinformatics*. 2007. T. 8.
17. Cunha Filho R. R. Da *et al.* Macular amyloidosis: A case report with scanning electron microscopy // *J. Eur. Acad. Dermatology Venereol.* 2016. T. 30. № 5. C. 889–890.
18. Derkatch I. L. *et al.* Dependence and independence of [PSI+] and [PIN+]: A two-prion system in yeast? // *EMBO J.* 2000. T. 19. № 9. C. 1942–1952.
19. Família C. *et al.* Prediction of peptide and protein propensity for amyloid formation // *PLoS One*. 2015. T. 10. № 8. C. 1–16.
20. Fernandez-Escamilla A. M. *et al.* Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins // *Nat. Biotechnol.* 2004. T. 22. № 10. C. 1302–1306.
21. Fischer E. R. *et al.* Scanning electron microscopy. , 2012. 1–47 c.
22. Frid P., Anisimov S. V., Popovic N. Congo red and protein aggregation in neurodegenerative diseases // *Brain Res. Rev.* 2007. T. 53. № 1. C. 135–160.
23. Gallardo R., Ranson N. A., Radford S. E. Amyloid structures: much more than just a cross- β fold // *Curr. Opin. Struct. Biol.* 2020. T. 60. № Lc. C. 7–16.
24. Galvin J. E. Detection of aggregates and protein inclusions by staining of tissues. // *Methods Mol. Biol.* 2003. T. 232. C. 149–164.
25. Garbuzynskiy S. O., Lobanov M. Y., Galzitskaya O. V. FoldAmyloid: A method of prediction of amyloidogenic regions from protein sequence // *Bioinformatics*. 2009. T. 26. № 3. C. 326–332.
26. Gasior P., Kotulska M. FISH Amyloid - a new method for finding amyloidogenic segments in proteins based on site specific co-occurrence of aminoacids // *BMC Bioinformatics*. 2014. T. 15. № 1. C. 1–8.
27. Genst E. De, Messer A., Dobson C. M. Antibodies and protein misfolding: From structural research tools to therapeutic strategies // *Biochim. Biophys. Acta - Proteins Proteomics*. 2014. T. 1844. № 11. C. 1907–1919.
28. Goldschmidt L. *et al.* Identifying the amyloids, proteins capable of forming amyloid-like fibrils // *Proc. Natl. Acad. Sci. U. S. A.* 2010. T. 107. № 8. C. 3487–3492.

29. Harris J. R. Transmission electron microscopy in molecular structural biology: A historical survey Dedication: This article is dedicated to the memory and scientific contribution of the late Robert (Bob) W. Horne, Milan V. Nermut and Marc Adrian, three exceptional han // Arch. Biochem. Biophys. 2015. T. 581. C. 3–18.
30. Harrison P. M., Gerstein M. A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. // Genome Biol. 2003. T. 4. № 6.
31. Hazenberg B. P. C. Amyloidosis. A clinical overview. // Rheum. Dis. Clin. North Am. 2013. T. 39. № 2. C. 323–345.
32. Hill S. E. *et al.* Amyloid protofibrils of lysozyme nucleate and grow via oligomer fusion // Biophys. J. 2009. T. 96. № 9. C. 3781–3790.
33. Hurshman A. R. *et al.* Transthyretin aggregation under partially denaturing conditions is a downhill polymerization // Biochemistry. 2004. T. 43. № 23. C. 7365–7381.
34. Iadanza M. G. *et al.* A new era for understanding amyloid structures and disease // Nat. Rev. Mol. Cell Biol. 2018. T. 19. № 12. C. 755–773.
35. Iconomidou V. A., Vriend G., Hamodrakas S. J. Amyloids protect the silkworm oocyte and embryo // FEBS Lett. 2000. T. 479. № 3. C. 141–145.
36. Ivanova M. I. *et al.* An amyloid-forming segment of β 2-microglobulin suggests a molecular model for the fibril // Proc. Natl. Acad. Sci. U. S. A. 2004. T. 101. № 29. C. 10584–10589.
37. Jucker M., Walker L. C. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases // Nature. 2013. T. 501. № 7465. C. 45–51.
38. Kaye R., Lasagna-Reeves C. A. Molecular mechanisms of amyloid oligomers toxicity // J. Alzheimer's Dis. 2013. T. 33. № SUPPL. 1. C. 1–12.
39. Kim C. *et al.* NetCSSP: Web application for predicting chameleon sequences and amyloid fibril formation // Nucleic Acids Res. 2009. T. 37. № SUPPL. 2. C. 469–473.
40. Knowles T. P. *et al.* Role of intermolecular forces in defining material properties of protein nanofibrils // Science (80-.). 2007. T. 318. № 5858. C. 1900–1903.
41. Kosolapova A. O. *et al.* Two novel amyloid proteins, ropA and ropB, from the root nodule bacterium rhizobium leguminosarum // Biomolecules. 2019. T. 9. № 11. C. 1–25.
42. Kushnirov V. V. *et al.* Prion and nonprion amyloids: a comparison inspired by the yeast

Sup35 protein. // *Prion*. 2007. T. 1. № 3. C. 179–184.

43. Lancaster A. K. *et al.* PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition // *Bioinformatics*. 2014. T. 30. № 17. C. 2501–2502.

44. Leeuwen F. W. Van *et al.* Molecular misreading: A new type of transcript mutation expressed during aging // *Neurobiol. Aging*. 2000. T. 21. № 6. C. 879–891.

45. Leung N., Nasr S. H., Sethi S. How I Treat amyloidosis: The importance of accurate diagnosis and amyloid typing // *Blood*. 2012. T. 120. № 16. C. 3206–3213.

46. Li J. *et al.* The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis // *Cell*. 2012. T. 150. № 2. C. 339–350.

47. Liaw C., Tung C. W., Ho S. Y. Prediction and Analysis of Antibody Amyloidogenesis from Sequences // *PLoS One*. 2013. T. 8. № 1.

48. Makin O. S., Serpell L. C. X-ray diffraction studies of amyloid structure // *Methods Mol. Biol.* 2005. T. 299. C. 67–80.

49. Marrero-Winkens C., Sankaran C., Schätzl H. M. From seeds to fibrils and back: Fragmentation as an overlooked step in the propagation of prions and prion-like proteins // *Biomolecules*. 2020. T. 10. № 9. C. 1–20.

50. Maurer-Stroh S. *et al.* Exploring the sequence determinants of amyloid structure using position-specific scoring matrices // *Nat. Methods*. 2010. T. 7. № 3. C. 237–242.

51. Michelitsch M. D., Weissman J. S. A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions // *Proc. Natl. Acad. Sci. U. S. A.* 2000. T. 97. № 22. C. 11910–11915.

52. Munir F. *et al.* MILAMP: Multiple Instance Prediction of Amyloid Proteins // *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 2019. T. PP. C. 1–1.

53. Naiki H. *et al.* Methods in laboratory investigation. Fluorometric examination of tissue amyloid fibrils in murine senile amyloidosis: Use of the fluorescent indicator, Thioflavine T // *Lab. Investig.* 1990. T. 62. № 6. C. 768–773.

54. Nelson R. *et al.* Structure of the cross- β spine of amyloid-like fibrils // *Nature*. 2005. T. 435. № 7043. C. 773–778.

55. Niu M. *et al.* RFAmyloid: A web server for predicting amyloid proteins // *Int. J. Mol. Sci.* 2018. T. 19. № 7.

56. Nizhnikov A. A. *et al.* Proteomic screening for amyloid proteins // PLoS One. 2014. T. 9. № 12. C. 1–18.
57. Nizhnikov A. A. *et al.* Interaction of Prions Causes Heritable Traits in *Saccharomyces cerevisiae* // PLoS Genet. 2016. T. 12. № 12. C. 1–19.
58. O’Brien R. J., Wong P. C. Amyloid precursor protein processing and alzheimer’s disease // Annu. Rev. Neurosci. 2011. T. 34. C. 185–204.
59. Otzen D., Riek R. Functional amyloids // Cold Spring Harb. Perspect. Biol. 2019. T. 11. № 12.
60. Planté-Bordeneuve V., Said G. Familial amyloid polyneuropathy // Lancet Neurol. 2011. T. 10. № 12. C. 1086–1097.
61. Prilusky J. *et al.* FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded // Bioinformatics. 2005. T. 21. № 16. C. 3435–3438.
62. RStudio Team. RStudio: Integrated Development Environment for R // 2021.
63. Saad S. *et al.* Reversible protein aggregation is a protective mechanism to ensure cell cycle restart after stress // Nat. Cell Biol. 2017. T. 19. № 10. C. 1202–1213.
64. Shorter J., Lindquist S. Prions as adaptive conduits of memory and inheritance // Nat. Rev. Genet. 2005. T. 6. № 6. C. 435–450.
65. Sing T. *et al.* ROCR: Visualizing classifier performance in R // Bioinformatics. 2005. T. 21. № 20. C. 3940–3941.
66. Sivanathan V., Hochschild A. A bacterial export system for generating extracellular amyloid aggregates // Nat. Protoc. 2013. T. 8. № 7. C. 1381–1390.
67. Sopova J. V. *et al.* RNA-binding protein FXR1 is presented in rat brain in amyloid form // Sci. Rep. 2019. T. 9. № 1. C. 1–14.
68. Soragni A. *et al.* Toxicity of Eosinophil MBP Is Repressed by Intracellular Crystallization and Promoted by Extracellular Aggregation // Mol. Cell. 2015. T. 57. № 6. C. 1011–1021.
69. Taglialegna A. *et al.* Staphylococcal Bap Proteins Build Amyloid Scaffold Biofilm Matrices in Response to Environmental Signals // PLoS Pathog. 2016. T. 12. № 6. C. 1–34.
70. Tartaglia G. G., Vendruscolo M. The Zyggregator method for predicting protein aggregation propensities // Chem. Soc. Rev. 2008. T. 37. № 7. C. 1395–1401.

71. Taverna M. A. Friend or foe? // *Aviat. Week Sp. Technol. (New York)*. 2008. T. 168. № 13. C. 53.
72. Thangakani A. M. *et al.* GAP: Towards almost 100 percent prediction for β -strand-mediated aggregating peptides with distinct morphologies // *Bioinformatics*. 2014. T. 30. № 14. C. 1983–1990.
73. Tian J. *et al.* Prediction of amyloid fibril-forming segments based on a support vector machine // *BMC Bioinformatics*. 2009. T. 10. № SUPPL. 1. C. 8–12.
74. Toombs J. A. *et al.* De novo design of synthetic prion domains // *Proc. Natl. Acad. Sci. U. S. A.* 2012. T. 109. № 17. C. 6519–6524.
75. Tsolis A. C. *et al.* A Consensus Method for the Prediction of «Aggregation-Prone» Peptides in Globular Proteins // *PLoS One*. 2013. T. 8. № 1. C. 1–6.
76. Varadi M. *et al.* AmyPro: A database of proteins with validated amyloidogenic regions // *Nucleic Acids Res.* 2018. T. 46. № D1. C. D387–D392.
77. Ventura S. *et al.* Short amino acid stretches can mediate amyloid formation in globular proteins: The Src homology 3 (SH3) case // *Proc. Natl. Acad. Sci. U. S. A.* 2004. T. 101. № 19. C. 7258–7263.
78. Walsh I. *et al.* PASTA 2.0: An improved server for protein aggregation prediction // *Nucleic Acids Res.* 2014. T. 42. № W1. C. 301–307.
79. Watt B, van Niel G, Raposo G M. M. A pigment cell-specific model for functional amyloid formation. *Pigment Cell Melanoma Res.* 2013;26(3):300-315. doi:10.1111/pcmr.12067 // *Bone*. 2011. T. 23. № 1. C. 1–7.
80. Westermark G. T., Johnson K. H., Westermark P. Staining methods for identification of amyloid in tissue // *Methods Enzymol.* 1999. T. 309. № 1922. C. 3–25.
81. Wickner R. B. [URE3] as an altered URE2 protein: Evidence for a prion analog in *Saccharomyces cerevisiae* // *Science (80-.)*. 1994. T. 264. № 5158. C. 566–569.
82. Wille H., Requena J. R. The structure of PrP^{Sc} prions // *Pathogens*. 2018. T. 7. № 1. C. 1–11.
83. Wisniowski B., Wechalekar A. Confirming the diagnosis of amyloidosis // *Acta Haematol.* 2020. T. 143. № 4. C. 312–321.
84. Yakupova E. I. *et al.* Congo Red and amyloids: History and relationship // *Biosci. Rep.* 2019. T. 39. № 1.

85. Yang W. *et al.* The propensity of the bacterial rodlin protein RdlB to form amyloid fibrils determines its function in *Streptomyces coelicolor* // *Sci. Rep.* 2017. Т. 7. № October 2016. С. 1–13.
86. Yoon S. *et al.* CSSP2: An improved method for predicting contact-dependent secondary structure propensity // *Comput. Biol. Chem.* 2007. Т. 31. № 5–6. С. 373–377.
87. Zeng G. *et al.* Functional bacterial amyloid increases *Pseudomonas* biofilm hydrophobicity and stiffness // *Front. Microbiol.* 2015. Т. 6. № OCT. С. 1–14.
88. Нижников А. А., Антонец К. С., Инге-Вечтомов С. Г. Амилоиды: От Патогенеза К Функции // *Биохимия.* 2015. Т. 80. № 9. С. 1356–1375.
89. 2020 Alzheimer’s disease facts and figures // *Alzheimer’s Dement.* 2020. Т. 16. № 3. С. 391–460.

Благодарности

Благодарность выражается научному руководителю ВКР, к.б.н., доценту Нижникову Антону Александровичу за помощь, наставления и бесценный вклад в работу.

Благодарность выражается рецензенту ВКР, к.б.н., Штарк Оксана Юрьевне за оценку итоговой работы.

Благодарность выражается к.б.н. Райко Михаилу Петровичу и к.б.н., доценту Антоненц Кириллу Сергеевичу за бесценные советы повлиявшие на ход работы над ВКР.