

Шагин Дмитрий Алексеевич

**Термостабильная дезоксирибонуклеаза из
Paralithodes camtschaticus –
новый инструмент исследования сложных геномов**

Специальность 1.5.3 – «молекулярная биология»

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
доктора биологических наук

Работа выполнена в отделе молекулярных технологий для биологии и медицины в лаборатории молекулярных технологий Федерального государственного бюджетного учреждения науки Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук.

Научный консультант:

доктор биологических наук, профессор, академик РАН **Лукьянов Сергей Анатольевич**

Официальные оппоненты:

Янковский Николай Казимирович - доктор биологических наук, профессор, академик РАН, научный руководитель ФГБУН Институт общей генетики им. Н.И. Вавилова Российской академии наук.

Лазарев Василий Николаевич - доктор биологических наук, доцент, заместитель генерального директора по научной работе ФГБУ «Федеральный научно-клинический центр физико-химической медицины имени академика Ю.М. Лопухина Федерального Медико-биологического Агентства»

Михайлович Владимир Михайлович - доктор биологических наук, старший научный сотрудник лаборатории биологических микрочипов ФГБУН Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук.

Ведущая организация:

Федеральное государственное бюджетное учреждение науки «Институт биологии гена Российской академии наук»

Защита диссертации состоится 3 апреля 2024 года в 11 часов на заседании диссертационного совета 24.1.037.01 при Федеральном государственном бюджетном учреждении науки Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук по адресу: 117997, ГСП-7, Москва, В-437, ул. Миклухо-Маклая, д.16/10.

С диссертацией можно ознакомиться в библиотеке ФГБУН Института биоорганической химии им. академиков М.М.Шемякина и Ю.А.Овчинникова РАН по адресу: 117997, ГСП-7, Москва, В-437, ул. Миклухо-Маклая, д.16/10.

Автореферат разослан « _____ » _____ 202__ г.

Ученый секретарь диссертационного совета,
доктор физико-математических наук

В.А. Олейников



ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Развитие современной биологии и медицинской науки тесно связано с исследованиями молекулярных механизмов, лежащих в основе нормальных и патологических процессов в живых организмах, а также с изучением генетических основ, определяющих эти механизмы. Для анализа геномов и транскриптомов в последние десятилетия был создан мощный технологический аппарат, включающий высокопроизводительные методы секвенирования ДНК и РНК, методы исследования генома и транскриптома единичных клеток и методы масштабного анализа генной экспрессии на микрочипах.

Многочисленные технологические прорывы в молекулярной биологии часто связаны с открытием и внедрением в практику ферментов, обладающих новыми активностями или ранее не описанными комбинациями свойств. Так, обнаружение бактериальных эндонуклеаз рестрикции и ДНК-лигазы в 1960-70-е годы сделало возможным создание химерных нуклеиновых кислот и открыло эру технологий рекомбинантных ДНК. Выявление термостабильной ДНК-полимеразы из бактерий *Thermophilus aquaticus* (Taq-полимераза) привело к широкому использованию в биологических науках, медицинской диагностике и криминалистике метода полимеразной цепной реакции (ПЦР), а обнаружение систем CRISPR-Cas, ставшее одним из значимых событий последних десятилетий, сделало реальностью редактирование геномов.

Нуклеазы являются группой ферментов, которые широко используются в биотехнологии и медицине. Однако среди нуклеаз, применяемых на момент инициации данной работы, отсутствовали клонированные термостабильные ферменты, способные избирательно разрушать только двухцепочечную (дц) ДНК, сохраняя одноцепочечную (оц) ДНК и РНК неповрежденными. В то же время фермент с такой активностью был высоко востребован для разработки ряда технологий анализа сложных смесей нуклеиновых кислот (НК). Вот лишь несколько примеров.

Значительная часть геномной ДНК эукариот представлена последовательностями, которые не несут смысловую нагрузку и, как правило, не представляют особого интереса для исследователей. К таким последовательностям относятся повторяющиеся некодирующие элементы генома: ретроэлементы, сателлитные повторы, простые двух-, трех-, четырех нуклеотидные повторы и т.д. Повторяющиеся последовательности составляют более половины генома человека (Balzano E. et al., 2021), а в геноме кукурузы и пшеницы их количество достигает 85% и 87% соответственно (Janicki M. et al., 2011). С учетом того, что для большинства эукариотических геномов все еще стоит задача определения первичной нуклеотидной последовательности, актуальной является разработка методических подходов, позволяющих значительно снизить содержание повторяющихся последовательностей в образцах геномной ДНК перед секвенированием.

Другая проблема обусловлена широким диапазоном вариаций в концентрации различных мРНК в клетках эукариот. Так, в большинстве тканей эукариот около 85% всей мРНК является продуктом небольшого числа так называемых «мажорных» генов (González-Porta M. et al., 2013). Для примера, в тканях сердца человека около

60% общей мРНК представлены всего пятью типами РНК, кодирующими белки митохондрий (Wang Q. et al., 2018). Из-за этого выявление редких и крайне редких транскриптов оказывается нетривиальной задачей, требующей либо многократного увеличения объемов секвенирования, либо применения специальных подходов для выравнивания концентраций разнопредставленных транскриптов в образце.

Еще одна проблема, стоящая перед исследователями, связана с функциональным скринингом кодирующих последовательностей, продукты которых обладают определенной активностью. Кроме разной представленности таких транскриптов, выявление целевых молекул затрудняет функциональная избыточность биологических систем, в которых одна и та же активность может быть присуща нескольким белкам (Ngara T.R. et al., 2018).

Эффективным решением этих проблем может быть нормализация образцов геномной ДНК/кДНК (выравнивание уровня представленности различных последовательностей) или удаление определенных заданных последовательностей из популяции исследуемой кДНК (деплеция нецелевых последовательностей кДНК). Указанные технологии требуют изолирования (выделения) оц ДНК из смеси НК – задача, которая легко решается при использовании фермента, обладающего такой активностью, и трудно разрешима при его отсутствии.

Таким образом, отсутствие в широкой палитре нуклеаз, используемых в биотехнологии и молекулярной биологии, ферментов, избирательно расщепляющих дц ДНК существенно ограничивает развитие методов, связанных с необходимостью выделения оц ДНК-фракции из сложных смесей НК. Выявление и клонирование таких ферментов с последующим созданием на их основе новых технологий является актуальной задачей.

Степень разработанности темы. В разработку методов анализа сложных смесей НК, в частности методов нормализации и деплеции ДНК и кДНК, вовлечено большое количество исследовательских групп, результаты деятельности которых подробно рассмотрены в обзоре литературы диссертации. Хотя показано, что некоторые ДНКазы обладают предпочтительностью к дц ДНК по сравнению с оц, эти ферменты не отличаются термостабильностью и/или не имеют достаточной специфичности и избирательности к дц ДНК, что серьезно ограничивает их использование при работе со сложными смесями НК *in vitro*. Отсутствие термостабильных ферментов, способных удалять дц ДНК, оставляя интактными одноцепочечные молекулы, привело к созданию многочисленных обходных путей, которые отличаются высокой трудоемкостью и многостадийностью, требуют больших количеств биологического материала и высокого профессионализма исследователя, что делает их неудобными, а в ряде случаев неприменимыми для работы.

Цель исследования. Целью настоящей работы было выявление, выделение и характеристика термостабильной нуклеазы, обладающей субстратной специфичностью к дц ДНК; и разработка с использованием этой нуклеазы технологий анализа сложных смесей нуклеиновых кислот.

Задачи исследования:

- клонирование кодирующей последовательности нуклеазы из гепатопанкреаса камчатского краба, специфичной к дц ДНК;

- выявление гомологов нуклеазы камчатского краба и структурно-функциональный анализ аминокислотных последовательностей для обнаружения ключевых участков, определяющих их специфическую ферментную активность;
- разработка новых высокоэффективных технологий, направленных на структурно-функциональные и сравнительные исследования геномов и транскриптомов как эукариот, так и прокариотических организмов;
- валидация разработанных технологий в модельных экспериментах;
- применение разработанных технологий для решения исследовательских задач и внедрение их в лабораторную практику.

Научная новизна исследования. В рамках настоящей работы была клонирована полноразмерная кодирующая последовательность нового фермента из гепатопанкреаса камчатского краба – дуплекс-специфическая нуклеаза (Par_DSN), которая демонстрировала уникальную комбинацию свойств, включающих термостабильность, высокую оптимальную температуру катализа, способность гидролизовать только ДНК в составе ДНК-ДНК и ДНК-РНК дуплексов.

Был идентифицирован ряд новых гомологов Par_DSN из других членистоногих и впервые охарактеризовано новое семейство дуплекс-специфических нуклеаз.

Были предложены новые высокоэффективные технологии для выявления однонуклеотидных различий в молекулах ДНК, идентификации целевых молекул ДНК и РНК в биологических образцах, нормализации кДНК и геномной ДНК, деплеции нецелевых последовательностей в образцах кДНК и ДНК.

Теоретическая и практическая значимость работы. В рамках работы был охарактеризован новый фермент, обладающий уникальной селективностью по отношению к дц ДНК и термостабильностью.

Был проведен структурно-функциональный анализ Par_DSN и ее гомологов и выявлены аминокислотные участки, определяющие субстратную специфичность и термостабильность нуклеазы камчатского краба.

Сравнительный и структурно-функциональный анализа Par_DSN и ее гомологов позволил выделить дуплекс-специфические нуклеазы в отдельное семейство внутри суперсемейства His-Me нуклеаз.

Уникальные свойства Par_DSN послужили толчком для разработки целого спектра молекулярно-биологических технологий анализа сложных геномов.

Данные настоящей работы легли в основу ряда коммерчески доступных продуктов, что привело к широкому внедрению полученных в исследовании результатов в мировую лабораторную практику. Нормализация и деплеция с использованием Par_DSN стали методами выбора при подготовке биологических образцов для высокопроизводительного секвенирования – NGS (от англ. next generation sequencing, секвенирование нового поколения). На основе уникальных свойств Par_DSN и разработанных нами технологий рядом западных лабораторий были предложены новые методы анализа микро РНК. В настоящее время число статей, упоминающих дуплекс-специфическую нуклеазу камчатского краба, в базе данных Google Scholar превышает 2000.

Методология и методы исследования. Работа выполнялась в период 2002-2012 гг. в лаборатории молекулярных технологий Института биоорганической химии имени академиков М.М. Шемякина и Ю.А. Овчинникова РАН. В работе

использовались наборы реактивов российских и зарубежных производителей, стандартные и оригинальные (разработанные в лаборатории молекулярных технологий) методы генной инженерии.

Определение последовательностей нуклеиновых кислот осуществлялось секвенированием по методу Сэнгера в лаборатории молекулярных технологий или с использованием автоматического секвенатора серии 454 GS FLX Standard (Roche, Швейцария) в центре «Биоинженерия» РАН. ПЦР проводили с использованием амплификатора MJ Research PTC-200 DNA Thermal Cycler (BioRad, США), ПЦР в режиме реального времени – с использованием амплификатора MX3005P QPCR Systems (Stratagene, США). Для электропорации использовали микропульсатор Biorad Micropulser (Biorad, США). Масс-спектрометрический анализ проводили с использованием MALDI-TOF/TOF с возможностью LIFT (Ultraflex TOF/TOF, Bruker Daltonik GmbH, Бремен, Германия) на базе ЦКП ИБХ РАН при участии Р.Х. Заганшина. Фрагментацию ДНК осуществляли с использованием ультразвукового процессора Cole-Parmer CP750 (Vernon Hills, IL, США).

Для анализа последовательностей нуклеиновых кислот и аминокислот применяли пакеты программ и программные продукты BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>), SMART (<http://smart.embl-heidelberg.de>), SignalP (версии 3.0, 5.0 и 6.0; <https://services.healthtech.dtu.dk/service.php/SignalP>), ТМНММ-2.0 (<https://services.healthtech.dtu.dk/service.php?TMNMM-2.0>) и PredictProtein (<https://predictprotein.org/>), FGESH (<http://sun1.softberry.com>). Repeat-Masker (www.repeatmasker.org/cgi-bin/WEBRepeatMasker), TREE-PUZZLE, версия 5.0, ClustalX, GeneDoc. Анализ макрочипов осуществляли с помощью пакета программ AtlasImage™ Software. Статистическую обработку данных осуществляли с использованием моделей дисперсионного анализа ANOVA, Systat Software SigmaPlot 11. Для анализа данных масс-спектрометрии применяли пакеты программ FlexAnalysis версии 2.4, BioTools версии 3.0 (Bruker Daltonics GmbH), Mascot (<http://www.matrixscience.com/>).

Сотрудники ЗАО Евроген оказали значительную помощь в синтезе олигонуклеотидов для экспериментальной работы.

Основные положения, выносимые на защиту:

1. Открыто новое семейство нуклеаз, специфически разрушающих двухцепочечную ДНК, и получившее название дуплекс специфические нуклеазы (ДСН).
2. Проведен филогенетический анализ белков данного семейства, а также исследованы функциональные и структурные особенности ферментов данного семейства на примере дуплекс специфической нуклеазы Par_DSN из камчатского краба (*Paralithodes camtschaticus*).
3. Определены границы нуклеазного домена Par_DSN, а также проведен анализ сайтов, ответственных за каталитическую активность данного фермента.
4. Разработана технология мутационного анализа в генах эукариот с использованием Par_DSN.
5. Разработана технология выявления ДНК-мишеней в комплексном продукте ПЦР с использованием Par_DSN.

6. Разработана технология селективного удаления транскриптов из популяций кДНК с использованием Par_DSN.

7. Разработана технология создания нормализованных библиотек кДНК из клеток и тканей эукариотических организмов с использованием Par_DSN.

8. Разработана технология создания нормализованных библиотек геномной ДНК из клеток и тканей эукариотических организмов с использованием Par_DSN.

9. Разработана технология удаления кДНК рРНК последовательностей при создании библиотек кДНК прокариотических организмов с использованием Par_DSN.

Личный вклад автора. Работа по клонированию и очистке нуклеазы камчатского краба Par_DSN, создание технологий с её использованием осуществлялась в лаборатории молекулярных технологий Института биоорганической химии имени академиков М.М. Шемякина и Ю.А. Овчинникова РАН рабочей группой в составе: Д.А. Шагина, Д.В. Ребрикова, А.С. Щеглова, В.Е. Анисимовой, Е.А. Богдановой и П.А. Жулидова, под общим руководством д.б.н. С.А. Лукьянова. Автором диссертации был разработан общий дизайн исследования и предложена стратегия клонирования дуплекс-специфической нуклеазы и ее гомологов, ему принадлежит ведущая роль в систематизации и обработке результатов экспериментов, координации и продвижении исследований, выполнении основных частей экспериментальной работы. Автор диссертации является автором всех предложенных технологий с использованием Par_DSN и их модификаций. Общая стратегия нормализации кДНК является совместной разработкой с С.А. Лукьяновым. Разработка протокола нормализации кДНК выполнялась совместно с П.А. Жулидовым. Первая модификация метода мутационного анализа была разработана совместно с Д.В. Ребриковым на основании идеи автора диссертации. Основной объем экспериментальной работы по клонированию нуклеаз из членистоногих был выполнен А.С. Щегловым, В.Е. Анисимовой и Д.В. Ребриковым. Характеристика физико-химических свойств нового фермента осуществлялась диссертантом совместно с А.С. Щегловым и П.А. Жулидовым. Протокол препаративного выделения нуклеазы был разработан А.С. Щегловым при участии диссертанта. Процедура очистки и активации рекомбинантных белков была разработана В.Е. Анисимовой и А.С. Щегловым. Филогенетические и структурные исследования нового семейства нуклеаз проводились диссертантом совместно с В.Е. Анисимовой и Е.А. Богдановой. В выполнении отдельных этапов экспериментальной работы участвовали сотрудники лаборатории молекулярных технологий ИБХ РАН.

Сотрудники Дальневосточного института биоорганической химии В.А. Рассказов, В.Б. Кожемяко, Н.И. Мензорова принимали участие в обсуждении результатов исследования свойств Par_DSN.

Работа по изучению аллельных вариантов гена COMT, ассоциированных с высокой болевой чувствительностью, выполнялась диссертантом в рамках международного научного консорциума по исследованию хронической боли.

Работа по исследованию аберрантно-слитых транскриптов в клеточной линии клеток рака молочной железы MCF-7 проводилась диссертантом в сотрудничестве с группой лабораторий, входящих в международную программу по исследованию рака молочной железы (Breast Cancer Research Program).

Работа по выявлению маркеров старения у рыбы *Nothobranchius furzeri* выполнялась диссертантом в рамках сотрудничества с группой лабораторий Германии в рамках проекта по изучению механизмов старения.

Соответствие диссертации паспорту научной деятельности. Научные положения диссертации соответствуют паспорту специальности 1.5.3 – Молекулярная биология (биологические науки).

Степень достоверности и апробация результатов. Результаты исследования были доложены и обсуждены на всероссийских и международных конференциях: «Третий съезд биотехнологов России им. Ю.А. Овчинникова» (Москва, 2005), «Генетика в России и мире» (Москва, 2006), «Human Genome Meeting» (Bethesda, 2006), «Ломоносов 2007» (Москва, 2007), «Биология – наука XXI века» (Пушино, 2007), «The 12th biological sciences graduate congress» (Kuala Lumpur, 2007), «Молекулярная диагностика» (Москва, 2017), «Высокопроизводительное секвенирование в геномике» (Новосибирск, 2017), научном семинаре отдела молекулярной диагностики и эпидемиологии ФБУН ЦНИИ Эпидемиологии (Москва, 2017).

Результаты исследования были использованы при разработке учебного курса «Основы биологии» для студентов кафедры молекулярной физики факультета молекулярной и биологической физики ГОУ ВПО Московского физико-технического института; учебных курсов: «Современные методы в биологии» и «Биоинженерия» для студентов кафедр биохимии, молекулярной биологии, вирусологии Московского государственного университета им. М.В. Ломоносова; учебного пособия для студентов Московского инженерно-физического института «Применение современных молекулярно-биологических методов для поиска и клонирования полноразмерных нуклеотидных последовательностей кДНК»

Объективность и достоверность полученных результатов проверена в ряде модельных экспериментов и подтверждена многочисленными исследованиями, проведенными во множестве научных лабораторий мира. Результаты исследования внедрены в международную лабораторную практику.

На основании результатов работы созданы наборы реактивов:

- *Duplex-specific nuclease, cat.# EA001; EA002; EA003; EA008 (Evrogen),*
- *Trimmer-2 cDNA normalization kit, cat.# NK003 (Evrogen)*

Работа прошла апробацию на открытом заседании Отдела геномики и постгеномных технологий ФГБУН Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН 17.05.2023 г.

Публикации. По результатам исследования опубликована 21 работа, включая 15 исследовательских и одну обзорную статью в рецензируемых журналах, представленных в международных базах данных, а также три главы в книгах международных издательств. Получено два патента на изобретения США.

Структура и объем диссертации. Диссертация изложена на 342 страницах машинописного текста, включает 32 таблицы, 78 рисунков. Диссертация состоит из введения, обзора литературы, материалов и методов, результатов исследования и их обсуждения, заключения, выводов, списка сокращений, а также списка использованной литературы. Список литературы включает 669 работ.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Результаты и обсуждение.

Идентификация термостабильной нуклеазы, селективно расщепляющей двухцепочную ДНК, и её характеристика. Отсутствие термостабильных ферментов специфических к дц ДНК в палитре нуклеаз, доступных для разработки методов выделения оц фракции ДНК из смесей НК, ограничивало возможности создания технологий анализа сложных образцов ДНК и кДНК. В 1990 г. была очищена ДНКазы креветки *Marsipaneus japonicus*, обладающая функциональным сходством с ДНКазой I быка (Chou M.Y. et al., 1990). Кодированная ее последовательность была идентифицирована в 2000 г. (Wang W.Y. et al., 2000) Было показано, что фермент обладает низким уровнем гидролитической активности по отношению к РНК в присутствии Mg^{2+} и Ca^{2+} , однако не было получено данных о различиях в его субстратной предпочтительности к оц и дц ДНК. Также было продемонстрировано наличие Ca^{2+} , Mg^{2+} -зависимой ДНКазы К у камчатского краба *Paralithodes camtschaticus* (Menzorova N.I. et al., 1994). Фермент демонстрировал устойчивость к прогреванию при $60^{\circ}C$ и предпочтительность расщепления дц ДНК по сравнению с оц ДНК. К сожалению, фермент не был клонирован и детально охарактеризован.

Нами было принято решение клонировать нуклеазу из камчатского краба. Было выдвинуто предположение, что термостабильный фермент камчатского краба может быть гомологичен ДНКазе креветки *M. japonicus*. Поэтому для осуществления задачи клонирования кДНК фермента было проведено сравнение аминокислотных последовательностей известных нуклеаз с нуклеазой *M. japonicus* и выявлены регионы для конструирования вырожденных праймеров (Рисунок 1).

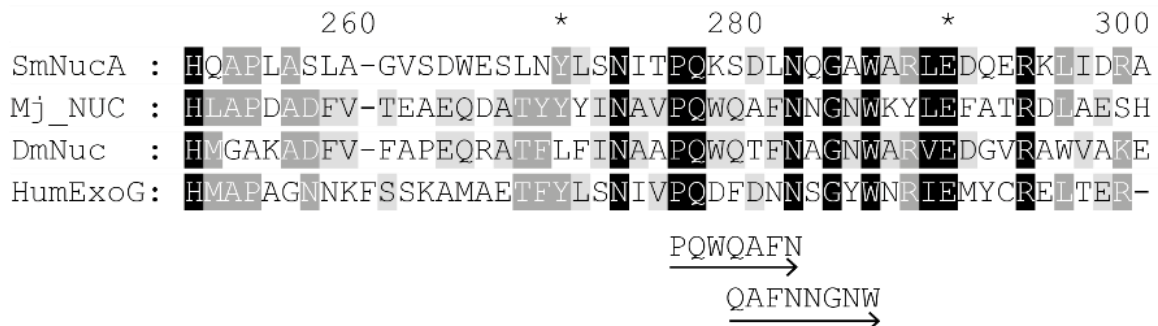


Рисунок 1. Множественное выравнивание аминокислотных последовательностей наиболее консервативного региона нуклеазы из *M. japonicus* (Mj_NUC), нуклеазы из *S. marcescens* (SmNucA), эндонуклеазы G человека (HumEndoG) и предсказанной на основании сравнения нуклеазы из *D. melanogaster* (DmNuc). Под выравниванием указаны регионы, выбранные для конструирования вырожденных праймеров для двух раундов амплификации 3'-концевого фрагмента кодирующей последовательности.

Последовательности нуклеаз из разных организмов отличаются низким процентом идентичности. В связи с этим не удалось подобрать пары праймеров для амплификации внутреннего фрагмента кодирующей последовательности целевого белка, поэтому для изолирования нуклеотидной последовательности нуклеазы камчатского краба была применена ранее разработанная нами стратегия, названная Step-Out RACE (Matz M. et al., 1999). Данные о нуклеотидной последовательности

полученного 3'-концевого фрагмента были использованы для изолирования и клонирования полной кодирующей последовательности нуклеазы. Последовательность нуклеотидов, кодирующая полноразмерную нуклеазу, была депонирована в базу данных GenBank под номером AF520591 (номер последовательности аминокислот — AAN86143). Анализ нуклеотидной последовательности нуклеазы выявил открытую рамку считывания длиной в 1221 п.о. (начиная с иницирующего кодона), кодирующую белок, состоящий из 407 аминокислот. Предсказанная аминокислотная последовательность белка демонстрировала 64% идентичности с аминокислотной последовательностью нуклеазы из *M. japonicus*, включала в себя сигнальный пептид и нуклеазный домен. Доменная организация предсказанного белка представлена на рисунке 2.

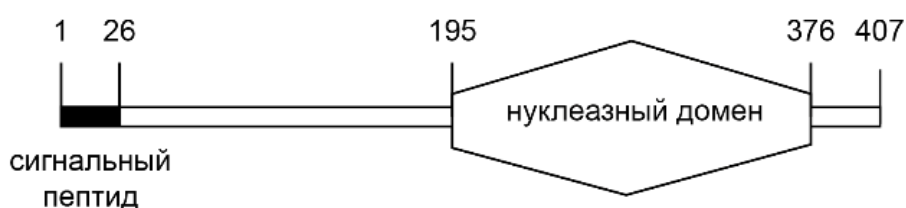


Рисунок 2. Доменная организация нуклеазы камчатского краба, предсказанная на основе анализа ее аминокислотной последовательности с помощью программного пакета SMART.

Получить активный рекомбинантный белок в гетерологической бактериальной системе экспрессии не удалось – весь продукт обнаруживался в нерастворимой фракции в тельцах включения. Поэтому было принято решение осуществить очистку природного фермента. Для контроля выделения и для подтверждения связи выделяемого природного фермента с клонированным нами белком было решено использовать поликлональные антитела (ПА) против неактивного рекомбинантного продукта, выделенного в денатурирующих условиях. Для выделения природного фермента на первом этапе работы использовали свежий гепатопанкреас, в дальнейшем перешли на коммерчески доступный ацетоновый порошок из камчатского краба. Стадии хроматографического разделения фракций контролировали по анализу продуктов в денатурирующем полиакриламидном геле, наличию нуклеазной активности к ДНК и связыванию фракции в денатурирующих условиях с ПА. Используя комбинацию из нескольких последовательных стадий хроматографии: анионообменная хроматография, два цикла очистки на фенол-сефарозе, хроматография на гидроксиапатитной колонке, хроматография на гепарин-сефарозе и гель-фильтрация, нами был получен гомогенный белок с высокой удельной нуклеазной активностью, и степенью очистки примерно в 618 раз от исходного. Молекулярная масса выделенного белка, определенная с помощью эксклюзионной хроматографии (~40 кДа), хорошо согласовывалась с молекулярной массой, определенной с помощью гель-электрофореза (~44 кДа), что указывало на то, что выделенный белок присутствовал в растворе в мономерной форме. Для подтверждения идентичности первичной структуры выделенного белка и аминокислотной последовательности, предсказанной по клонированной кодирующей нуклеотидной последовательности,

дополнительно к окрашиванию антителами был использован масс-спектрометрический анализ. Параллельная масс-спектрометрия продуктов обработки очищенного препарата природной нуклеазы трипсином выявила с высокой степенью достоверности четыре пептида: DIETSRPSFK, YLEYATR, DLAESHGSDLR и EVVPVPSLTWK. Каждая из этих аминокислотных последовательностей была представлена в аминокислотной последовательности, предсказанной на основе анализа клонированной НК. Таким образом, идентичность денатурированного рекомбинантного белка (кодируемого клонированной нуклеотидной последовательностью гена нуклеазы) и активного фермента, очищенного из гепатопанкреаса краба, была подтверждена масс-спектрометрическим анализом пептидов и иммуноблоттингом.

Анализ физико-химических свойств выделенного природного фермента осуществляли с использованием в качестве субстрата дц ДНК. Было показано, что нуклеаза проявляет ферментативную активность только в присутствии двухвалентных катионов металлов и ингибируется ЭДТА, то есть является металл-зависимым ферментом. Максимальный уровень активности наблюдался в присутствии ионов марганца в концентрации 20 мМ. В присутствии оптимальной концентрации ионов кобальта (10 мМ) фермент демонстрировал 70% от максимальной активности, а в присутствии оптимальных концентраций ионов магния (7–30 мМ) только 10%. В присутствии только ионов Ca^{2+} и Cd^{2+} фермент не проявлял нуклеазной активности (рисунок 3).

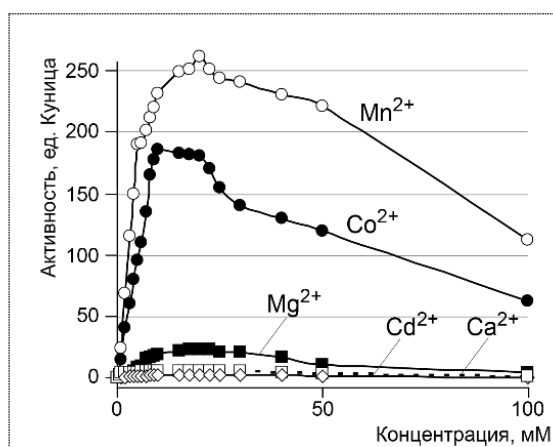


Рисунок 3. Зависимость активности нуклеазы от концентрации двухвалентных катионов

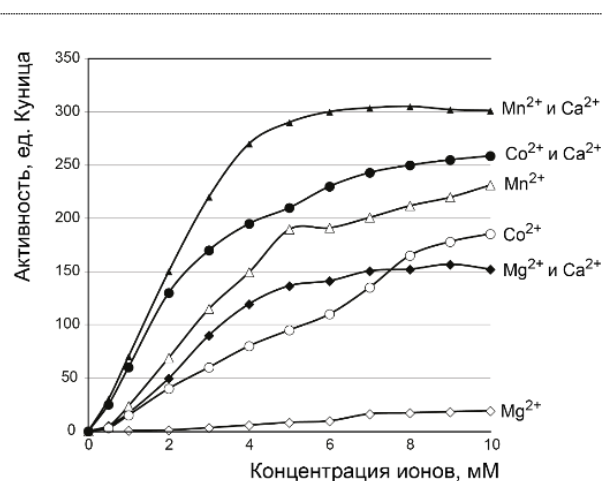


Рисунок 4. Зависимость активности нуклеазы от концентрации комбинаций двухвалентных катионов

Добавление к реакционной смеси, содержащей ионы магния, ионов Ca^{2+} производило синергетический эффект, увеличивая активность фермента примерно в пять раз при оптимальной концентрации обоих ионов. Этот эффект при комбинациях ионов Ca^{2+} и Co^{2+} или Ca^{2+} и Mn^{2+} был выражен в значительно меньшей степени (рисунок 4). Для остальных комбинаций исследуемых ионов синергетического эффекта обнаружено не было.

Фермент проявлял активность в широком диапазоне температур. При использовании в качестве субстрата дц ДНК температурный оптимум наблюдался

при $\sim 60^{\circ}\text{C}$. Дальнейшее повышение температуры сопровождалось резким снижением каталитической активности (рисунок 5). Это могло быть связано как с денатурацией фермента, так и (с большей вероятностью) с плавлением дц ДНК-субстрата. Кинетика тепловой денатурации нуклеазы показана на рисунке 6. Инкубация в течение 30 мин при температуре 80°C сопровождалась 60% падением активности фермента. Тем не менее, 5% активности сохранялось даже при тридцатиминутном прогревании фермента при 100°C .

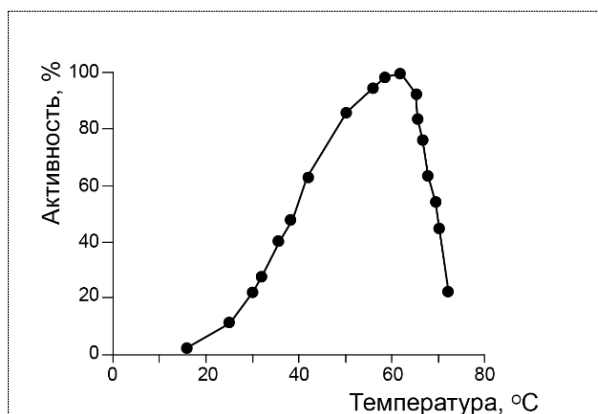


Рисунок 5. Зависимость активности нуклеазы от температуры

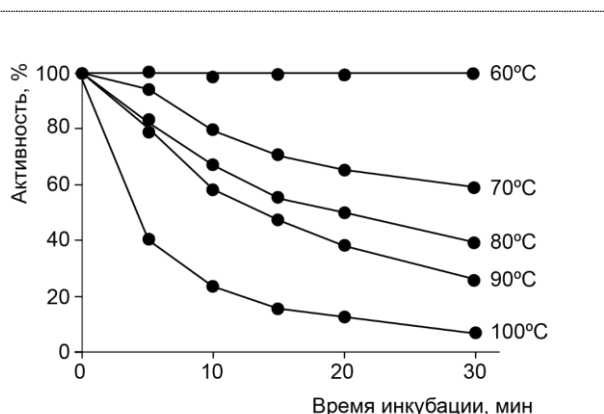


Рисунок 6. Кинетика тепловой денатурации нуклеазы камчатского краба при 30 мин инкубации при разных температурах

Анализ рН стабильности фермента показал, что он сохраняет более 90% активности после 30 мин инкубации в диапазоне рН от 4 до 12. Фермент имеет широкий оптимум рН с максимумом при рН 6,8. При значениях рН менее 3,0 и более 9,0 детектируемой каталитической активности фермента не обнаруживалось.

Повышение ионной силы раствора сопровождалось снижением каталитической активности нуклеазы. Так, добавление 0,2 М NaCl снижало ферментативную активность нуклеазы в 10 раз, а при концентрации 0,5 М NaCl сохранялось всего около 1,5% ее активности. Мочевина была менее эффективна в ингибировании ферментативной активности, чем NaCl, и для 10-кратного ингибирования активности нуклеазы камчатского краба требовалось 3 М мочевины. При 4 М мочевины фермент сохранял 3,5% активности.

Инкубация нуклеазы в течение 1 ч при 37°C в присутствии 1% SDS, 10 мМ меркаптоэтанола или 0,3% перекиси водорода снижала активность фермента только на 8-10%. Однако инкубация в присутствии 1% SDS при 60°C полностью инактивировала фермент; а в присутствии меркаптоэтанола и перекиси водорода в тех же условиях приводила к потере 70% и 80% активности, соответственно. Инкубация со спермидином в концентрациях менее 5 мМ не оказывала ингибирующего действия на нуклеазу камчатского краба, однако при 10 мМ спермидина наблюдалось падение ее активности до 50%, а при концентрации 8 мМ – до 7-8%. Кроме того, было выявлено, что фермент обладает устойчивостью к обработке протеиназой К и папаину. Инкубация в течение 30 мин в присутствии этих протеиназ при 37°C не приводила к изменению активности фермента.

На образцах природной оц и дц ДНК было выявлено, что фермент обладает селективной избирательностью по отношению к дц ДНК, оставляя оц ДНК интактной после 10-15 мин инкубации при 65°C (рисунки 7 – 8). Нуклеаза расщепляла любые дц ДНК субстраты (геномную ДНК, фаговую ДНК, суперскрученную плазмидную ДНК), но не проявляла заметной активности против оц ДНК.

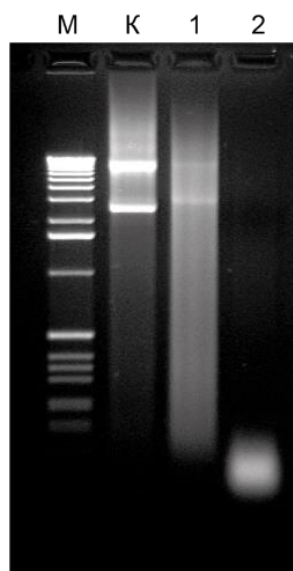


Рисунок 7. Анализ расщепления плазмидной ДНК нуклеазой камчатского краба. Контроль (К) — инкубация 200 нг плазмидной ДНК без нуклеазы; дорожки 1, 2 — инкубация 200 нг плазмидной ДНК при 65°C в течение 1,5 мин (дорожка 1) и 7 мин (дорожка 2) в присутствии 0,2 ед. Куница/мкл нуклеазы. М — маркер длин ДНК, 1 kb ladder (Invitrogen, США).

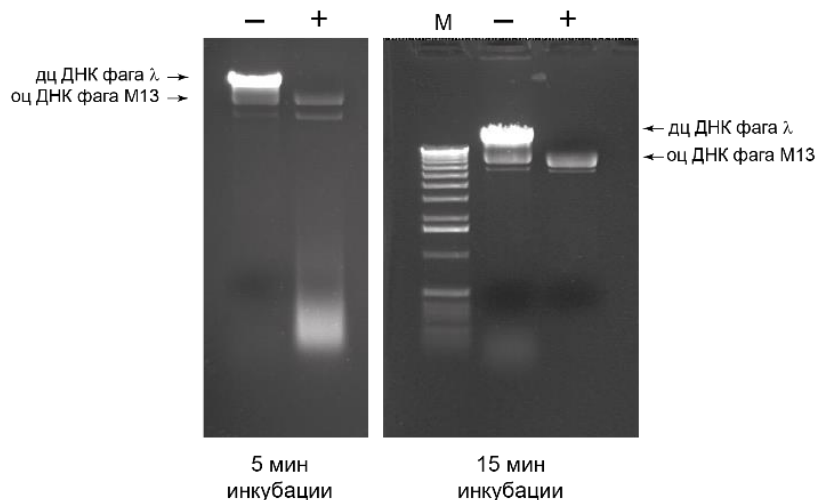


Рисунок 8. Анализ расщепления фаговой дц и оц ДНК нуклеазой камчатского краба. Электрофорез в агарозном геле продуктов реакции расщепления фаговой дц и оц ДНК в присутствии 0,2 ед. Куница/мкл (+) или в отсутствии (-) нуклеазы камчатского краба в ходе инкубация при 65°C. Время инкубации указано под электрофореграммами. М — маркер длин ДНК, 1 kb ladder (Invitrogen, США).

Сравнение эффективности расщепления оц и дц ДНК далее проводили с использованием синтетических оц и дц олигонуклеотидов. Для формирования дц ДНК субстратов меченный донором и гасителем флуоресценции 20-мерный олигонуклеотид гибридизовали в эквимольных количествах с комплементарными ему немечеными олигонуклеотидами разной длины. Структуры олигонуклеотидных субстратов и время их полураспада ($T_{1/2}$) представлены в таблице 1. Результаты анализа показаны на рисунке 9.

Подсчитанное соотношение гидролитической активности фермента к дц и оц ДНК субстрату составило не менее 1000:1, что значительно выше, чем для всех известных ранее нуклеаз. При изменении условия гидролиза (добавление в реакционную среду различных комбинаций активирующих ионов двухвалентных металлов, детергентов, хаотропных агентов, полиаминов) это соотношение не менялось. Благодаря столь высокой избирательности нуклеазы по отношению к дц

ДНК, она была нами названа дуплекс-специфической нуклеазой (DSN или Par_DSN, от. англ. Paralithodes camchaticus duplex-specific nuclease).

Таблица 1 – Каталитическая активность нуклеазы по отношению к синтетическим ДНК субстратам

| Субстрат | Длина дц ДНК дуплекса | T _{1/2} , мин |
|---|-----------------------|------------------------|
| 5'-TAMRA-tgggcagtgctcgcttagtg-DABCYL-3' | 0 | 406,8 ± 37,0 |
| 5'-TAMRA-tgggcagtgctcgcttagtg-DABCYL-3' 3'-cacgagc-5' | 7 | 241,9 ± 39,2 |
| 5'-TAMRA-tgggcagtgctcgcttagtg-DABCYL-3' 3'-cacgagcg-5' | 8 | 83,2 ± 15,5 |
| 5'-TAMRA-tgggcagtgctcgcttagtg-DABCYL-3' 3'-tcacgagcg-5' | 9 | 20,8 ± 6,15 |
| 5'-TAMRA-tgggcagtgctcgcttagtg-DABCYL-3' 3'-tcacgagcga-5' | 10 | 2,4 ± 0,4 |
| 5'-TAMRA-tgggcagtgctcgcttagtg-DABCYL-3' 3'-accggtcacgagcgaatcac-5' | 20 | 0,6 ± 0,2 |

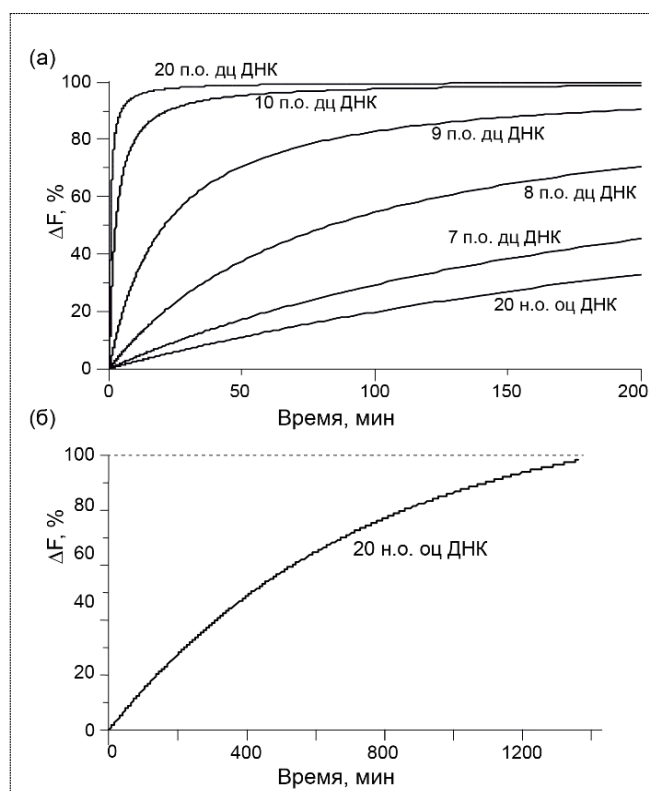


Рисунок 9. Анализ расщепления синтетических оц и дц ДНК субстратов нуклеазой камчатского краба. Относительное изменение флуоресценции реакционной смеси (ΔF) определяли по формуле: $\Delta F = (FR - F0 / FM - F0) \times 100\%$, где FR — уровень флуоресценции реакционной смеси после инкубации в присутствии нуклеазы, FM — уровень флуоресценции реакционной смеси, содержащей полностью гидролизованный субстрат, F0 — фоновая флуоресценция реакционной смеси в отсутствие фермента. Использовали 0,2 ед. Куница / мкл нуклеазы, инкубация проводилась при 30°C.

Для анализа особенностей расщепления субстрата Par_DSN использовали радиоактивно меченые олигонуклеотиды длиной 7–9 н.о. Дц ДНК субстраты формировали гибридизацией этих олигонуклеотидов с комплементарной областью немеченого олигонуклеотида длиной 43 н.о. В ходе инкубации радиоактивно меченых субстратов в течение 1 ч при 30°C детектируемого расщепления 7-звеньевых дц ДНК-дуплексов не наблюдалось, дуплексы длиной 8 п.о. разрушались

частично, а дуплексы длиной 9 п.о. — расщеплялись полностью (данные не приведены).

Нуклеаза камчатского краба не проявляла детектируемой активности по отношению к оц и дц РНК-субстратам (рисунок 10), в то же время она эффективно расщепляла ДНК цепь в ДНК-РНК гибридах (данные не приведены).

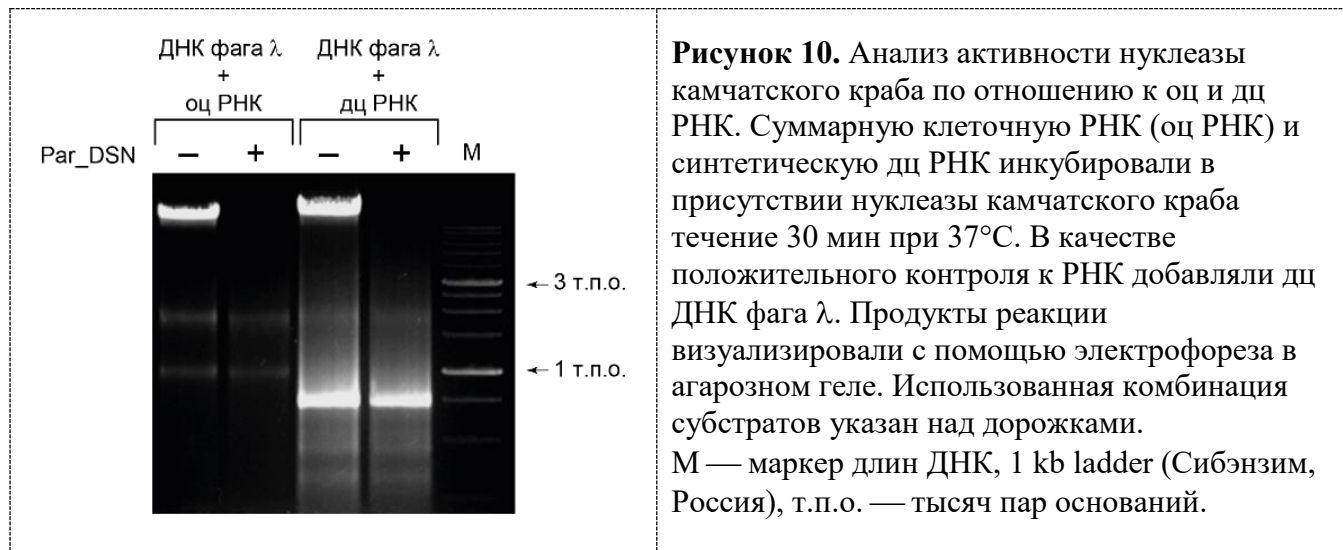


Рисунок 10. Анализ активности нуклеазы камчатского краба по отношению к оц и дц РНК. Суммарную клеточную РНК (оц РНК) и синтетическую дц РНК инкубировали в присутствии нуклеазы камчатского краба течение 30 мин при 37°C. В качестве положительного контроля к РНК добавляли дц ДНК фага λ. Продукты реакции визуализировали с помощью электрофореза в агарозном геле. Использованная комбинация субстратов указан над дорожками. М — маркер длин ДНК, 1 kb ladder (Сибэнзим, Россия), т.п.о. — тысяч пар оснований.

Таким образом, нами был клонирован и охарактеризован термостабильный фермент, обладающий непревзойденной селективностью по отношению к дц ДНК – дуплекс-специфическая нуклеаза Par_DSN. Было показано, что Par_DSN практически неактивна по отношению к РНК и оц-ДНК субстрату и проявляет свойства металл-зависимой ДНКазы по отношению к ДНК в ДНК-ДНК и ДНК-РНК гибридах. Минимальная длина дуплекса, разрушаемого Par_DSN, составляет 8 п.о. Важно отметить, что Par_DSN сочетает в себе высокую специфичность и высокую стабильность, что и обусловило большую привлекательность этого фермента как инструмента для биотехнологии. К сожалению, Par_DSN, экспрессированная в *E. coli*, была локализована в тельцах включения, и наши попытки экспрессировать функциональный рекомбинантный белок сначала не увенчались успехом. Однако мы смогли использовать очищенный, денатурированный рекомбинантный белок для получения препарата антител и разработать успешную стратегию очистки нуклеазы из гепатопанкреаса камчатского краба. Вскоре Анисимовой с соавторами удалось разработать схему реактивации полученного рекомбинантного белка Par_DSN, основанную на устойчивости этой нуклеазы к обработке протеиназой К (Anisimova, V.E. et al., 2006). Свойства реактивированного рекомбинантного белка оказались идентичны свойствам выделенной из камчатского краба нуклеазы. Это стало окончательным подтверждением удачно проведенного исследования.

Идентификация семейства нуклеаз, специфически расщепляющих дц ДНК, и его структурно-функциональная характеристика. Наше исследование структуры и свойств ДСН гепатопанкреаса камчатского краба было в значительной степени индуцировано идентификацией нуклеазы из гепатопанкреаса креветки *M. japonicus*. Анализ ее аминокислотной последовательности выявил сходство ее

нуклеазного домена с нуклеазными доменами *Serratia*-подобных ДНК-РНК неспецифических эндонуклеаз. *Serratia*-подобные ДНК-РНК неспецифические эндонуклеазы (SNF от англ. *Serratia family of nucleases*), архетипом которых является нуклеаза SmNucA, относятся к суперсемейству His-Me (Wu C.-C. et al., 2020). Большинство SNF расщепляют ДНК и РНК в дц и оц форме с одинаковой эффективностью или, как эндонуклеаза из *Syncephalastrum racemosum*, с небольшим предпочтением к ДНК. Для всех представителей семейства был постулирован сходный механизм работы и наличие единственного каталитического центра, связывающего как ДНК, так и РНК. Характерной особенностью структурной организации SNF является наличие так называемого нуклеазного домена (NUC) с консервативным мотивом RGH, в состав которого входит гистидиновый остаток каталитического центра (Meiss G. et al., 2000).

Нуклеаза креветки эффективно расщепляла ДНК и проявляла слабую активность в отношении РНК субстратов. Par_DSN обнаруживала 64% идентичности аминокислотной последовательности с нуклеазой креветки и сходную комбинацию аминокислотных остатков каталитического центра, предсказанного на основе сравнения ее аминокислотной последовательности с последовательностями SNF-эндонуклеаз. В то же время Par_DSN проявляла уникальную субстратную специфичность, расщепляя только ДНК в составе дц молекул ДНК-ДНК и ДНК-РНК. Кроме того, Par_DSN обладала выраженной термостабильностью, не характерной для SNF. В литературных источниках не было описания подобной активности. Механизм, лежащий в основе особых свойств Par_DSN, требовал дальнейшего изучения.

В 2006 г. появились данные об аминокислотной последовательности еще одной нуклеазы, обладающей сходной с Par_DSN субстратной специфичностью – *CuquEndo* из семиполосного комара (Calvo E. et al., 2006). Интересно, что *CuquEndo* также была отнесена авторами к SNF нуклеазам.

Мы предположили, что сравнительный анализ Par_DSN, нуклеазы креветки *M. japonicus* и *CuquEndo* с «классическими» членами SNF позволит выявить особенности этих ферментов, обуславливающие их уникальную субстратную специфичность. Так как точность множественного выравнивания последовательностей тем выше, чем больше используется родственных последовательностей, было принято решение о клонировании гомологов Par_DSN из других ракообразных и проведении дополнительного поиска родственных последовательностей в доступных базах данных.

С помощью технологии, которую мы использовали для клонирования нуклеотидной последовательности Par_DSN, нами были получены полные кодирующие последовательности потенциальных нуклеаз трех видов ракообразных: Gam_Nuc из рачка *Gammarus* sp., Pal_Nuc из креветки *Palaemonidae* sp. и Uca_Nuc из краба *Uca crassipes*. Эти последовательности были депонированы в базу данных GenBank: последовательность нуклеотидов Gam_Nuc DQ862539, Pal_Nuc — DQ862538, Uca_Nuc — DQ862540; последовательности аминокислот — ABI18973, ABI18972 и ABI18974 соответственно. Предсказанные аминокислотные последовательности Pal_Nuc и Uca_Nuc имели 50-65% идентичности с последовательностью Par_DSN, в то время как последовательность Gam_Nuc —

только 32% идентичности. С помощью ПО SignalP-3.0/5.0 во всех предсказанных белках ракообразных (включая идентифицированные нами нуклеазы и нуклеазу креветки Mj_Nuc) был обнаружен сигнальный пептид, указывающий на возможный секреторный характер их экспрессии. Однако точность выявления сигнального пептида у Gam_Nuc оказалась существенно ниже, чем у других исследуемых нуклеаз. Согласно анализу, выполненному с помощью ПО SMART, Pal_Nuc, Uca_Nuc и Gam_Nuc имеют доменную организацию сходную с организацией доменов Par_DSN и Mj_Nuc, содержат сигнальный пептид и отделенный от него линкерной последовательностью нуклеазный домен, гомологичный домену эндонуклеаз SNF. В структуре Gam_Nuc в линкерной последовательности был также выявлен трансмембранный helix регион.

Дополнительно в доступных базах данных с помощью пакета программ NCBI BLAST был выявлен ряд последовательностей гомологов нуклеаз ракообразных. Большая часть найденных аминокислотных последовательностей относилась к неохарактеризованным белкам, предсказанным на основе данных геномного секвенирования насекомых. Для всех экстрагированных из GenBank последовательностей была исследована структурная организация полипептидной цепи. По данным ПО SMART все они содержали нуклеазный домен, характерный для SNF. При анализе с помощью ПО SignalP-5.0 сигнальный пептид был обнаружен у большинства белков, кроме группы митохондриальных эндонуклеаз (CaenEndoG, Hum_EndoG, DmEndoG, Calb_Nuc), а также белков AGAP007845, Cun_NucC1A, Dm_CG3819, Sr_Nuc1 и AGAP011048. Последовательность AGAP011048 была предсказана в ходе аннотирования генома *A. gambitae* и содержала неполные N- и C-концы. Дополнительный поиск «потерянных» экзонов для AGAP011048 был осуществлён с помощью ПО FGESH и позволил обнаружить некорректно выявленные при аннотировании места вырезания интронов и недостающие 3' и 5'-концевые экзоны. В дальнейшем сравнении мы использовали уточненный вариант последовательности. Результат множественного выравнивания последовательностей приведен на рисунке 11, где Par_DSN – камчатский краб *P. camtschaticus*, AAN86143; Gam_Nuc – гаммарус *Gammarus* sp., ABI18973; Mj_Nuc – креветка кумура *M. Japonicus*, CAB55635; Pal_Nuc – креветка *Palaemonidae* sp., ABI18972; Uca_Nuc – краб-скрипач *Uca crassipes*, ABI18974; Gm_Tsal1 – муха цеце *Glossina morsitans morsitans*, AAF82097; Gm_Tsal2 – муха цеце *G. morsitans morsitans*, AAF82098; Cmae_Nuc – зеленый краб *Carcinus maenas*; Hame_Nuc – лобстер *Homarus americanus*; Dm_CG14120 – плодовая муха *D. melanogaster*, CG3819; Dm_CG3819 – *D. melanogaster*, CG3819; LOC4812298 – *D. pseudoobscura*, XP_0013529 14.1; Dm_CG6839 – *D. melanogaster*, NP_649076; AGAP011047 – комар *Anopheles gambiae*, XP_308089; AGAP011048 – *A. gambiae*, XP_308088; AGAP007845 – *A. gambiae*, XP_317654; Dm_CG14118 – *D. melanogaster*, NP_648612; Dm_IP10440 – *D. melanogaster*, AAY55811; Dm_LP23408p – *D. melanogaster*, AAS77426; DmEndoG – *D. melanogaster*, NP_610737; CuquEndo – москит *Culex quinquefasciatus*, AAR18449; Dm_CG9989 – *D. melanogaster*, AAF56806; Lundep – *Lutzomyia longipalpis*, AAS16916; SmNucA – *erratia marcescens*, P13717; AnNucA – *Anabaena variabilis*, P38446; Sr_Nuc1 – *Syncephalastrum racemosum*, P81204; Calb_Nuc – *Candida albicans*, XP_716805; Cun_NucC1A – *Cunninghamella echinulate*, AAC78769; Sac_Nuc –

Saccharomyces cerevisiae, NP_012327; HumEndoG – Homo sapiens, NP_005098;
CaenEndoG – Caenorhabditis elegans, NP_491371.

| | | | | | | | | |
|------------|---|---|----|-------|-----|-------|-----|-------|
| | | * | 20 | * | 40 | * | 60 | |
| SmNucA | : | ----- | | ----- | | ----- | | ----- |
| AnNucA | : | ----- | | ----- | | ----- | | ----- |
| Hum_EndoG | : | ----- | | ----- | | ----- | | ----- |
| CaenEndoG | : | ----- | | ----- | | ----- | | ----- |
| AGAP007845 | : | ----- | | ----- | | ----- | | ----- |
| DmEndoG | : | ----- | | ----- | | ----- | | ----- |
| Calb_Nuc | : | ----- | | ----- | | ----- | | ----- |
| Sac_Nuc | : | ----- | | ----- | | ----- | | ----- |
| Sr_Nuc1 | : | ----- | | ----- | | ----- | | ----- |
| Cun_NucC1A | : | ----- | | ----- | | ----- | | ----- |
| Gam_Nuc | : | ----- | | ----- | | ----- | | ----- |
| CuquEndo | : | ----- | | ----- | | ----- | | ----- |
| Lundep | : | ----- | | ----- | | ----- | | ----- |
| Gm_Tsall | : | ----- | | ----- | | ----- | | ----- |
| Gm_Tsal2 | : | ----- | | ----- | | ----- | | ----- |
| Dm_CG9989 | : | ----- | | ----- | | ----- | | ----- |
| Dm_IP10440 | : | ----- | | ----- | | ----- | | ----- |
| Dm_LP23408 | : | ----- | | ----- | | ----- | | ----- |
| Dm_CG14118 | : | ----- | | ----- | | ----- | | ----- |
| Dm_CG14120 | : | ----- | | ----- | | ----- | | ----- |
| Dm_CG3819 | : | ----- | | ----- | | ----- | | ----- |
| AGAP011048 | : | ----- | | ----- | | ----- | | ----- |
| AGAP011047 | : | ----- | | ----- | | ----- | | ----- |
| LOC4812298 | : | ----- | | ----- | | ----- | | ----- |
| Dm_CG6839 | : | ----- | | ----- | | ----- | | ----- |
| Par_DSN | : | ----- | | ----- | | ----- | | ----- |
| Uca_NUC | : | ----- | | ----- | | ----- | | ----- |
| Mj_Nuc | : | ----- | | ----- | | ----- | | ----- |
| Pal_Nuc | : | ----- | | ----- | | ----- | | ----- |
| Cmae_Nuc | : | ----- | | ----- | | ----- | | ----- |
| Hame_Nuc | : | ----- | | ----- | | ----- | | ----- |
| | | * | 80 | * | 100 | * | 120 | |
| SmNucA | : | ----- | | ----- | | ----- | | ----- |
| AnNucA | : | ----- | | ----- | | ----- | | ----- |
| Hum_EndoG | : | ----- | | ----- | | ----- | | ----- |
| CaenEndoG | : | ----- | | ----- | | ----- | | ----- |
| AGAP007845 | : | ----- | | ----- | | ----- | | ----- |
| DmEndoG | : | ----- | | ----- | | ----- | | ----- |
| Calb_Nuc | : | ----- | | ----- | | ----- | | ----- |
| Sac_Nuc | : | ----- | | ----- | | ----- | | ----- |
| Sr_Nuc1 | : | ----- | | ----- | | ----- | | ----- |
| Cun_NucC1A | : | ----- | | ----- | | ----- | | ----- |
| Gam_Nuc | : | ALLLATEKFYESDDTRFIVQTTFCCTTIFYYFSCILLFEEKKVCATISTLLHNIV--- | | ----- | | ----- | | ----- |
| CuquEndo | : | --ISELQSERKVIITLNSDKRSLKQPSN---- | | ----- | | ----- | | ----- |
| Lundep | : | FPENVTAKKEPVYIKPSNDGSLSTPL-QPSGPFVSLKIGESLAFPCPGDGKD---- | | ----- | | ----- | | ----- |
| Gm_Tsall | : | IPESVENEKTPVIMVRKSLKTFEYDLFQPTGEVTNFPERTELLACTGAQNYF----- | | ----- | | ----- | | ----- |
| Gm_Tsal2 | : | IPDDLNGEEAPVVLVKTGNVVKLFRPE---EKTTTFPKGTELLACTGEGNGL----- | | ----- | | ----- | | ----- |
| Dm_CG9989 | : | -LAHLNANYVYLSQNNGVYDIQRS----- | | ----- | | ----- | | ----- |
| Dm_IP10440 | : | --QYMVEQSNGITFYRDASGSIQLQRL----- | | ----- | | ----- | | ----- |
| Dm_LP23408 | : | --RAVQGTNRIFMVRGQNRQLSLKRT----- | | ----- | | ----- | | ----- |
| Dm_CG14118 | : | -DLPRLDKVPYLRPGTDLYWLPNAY---- | | ----- | | ----- | | ----- |
| Dm_CG14120 | : | -VNGDLNDPAPLFSRHNSYEIIVPDPT----- | | ----- | | ----- | | ----- |
| Dm_CG3819 | : | -IKSDTSEIVG-FSDTGYVDVDKD----- | | ----- | | ----- | | ----- |
| AGAP011048 | : | -FADLPYEQPIILIPGTEKYVPLDE---TREILVPTGAPTELLACQGFRLF----- | | ----- | | ----- | | ----- |
| AGAP011047 | : | LATAALPQPQPLFLIPGTDQFKYPSTS---SGILTLSAGETLELACQNGFSLF----- | | ----- | | ----- | | ----- |
| LOC4812298 | : | IRGSELKDPQPLLIKPGTSDFFGYSDT---GVVDVEKDKTEFHCTSSLASP----- | | ----- | | ----- | | ----- |
| Dm_CG6839 | : | -IRGGLPSEPEVYIKTDSDFYFSDV----- | | ----- | | ----- | | ----- |
| Par_DSN | : | -KDTDFPED-PLIFDSNLELIRPVLEN--GKRIVSVPSGSLTLACSGSELIN----- | | ----- | | ----- | | ----- |
| Uca_NUC | : | -KDLDFPEHPPLIINNQLDFVLPVLEG---VNRVVRVAEGETVTLACSGSELVN----- | | ----- | | ----- | | ----- |
| Mj_Nuc | : | -KDTDFPEYPIILDASLEIVRPVAEG--EARVVRVSAGAKTTLACPGSELVN----- | | ----- | | ----- | | ----- |
| Pal_Nuc | : | -KDAFLPTPPLLLDSSLKMIYPVLEG---SLRMVRVAAGSTITVACSGTTISC----- | | ----- | | ----- | | ----- |
| Cmae_Nuc | : | -KDNDFPPEEVLIDGHLKLLLPVKKD--STREVRIPSDNNVTLACPGKGNV----- | | ----- | | ----- | | ----- |
| Hame_Nuc | : | -KDENYPPNPIILDDSHIVRPVLEG--EDRIVRLAAKSKVTLACPGTKISN----- | | ----- | | ----- | | ----- |

Рис. 11. Начало

```

          *           140           *           160           *           180
SmNucA   : -----MRFNKMLLAALLFAAQASADTLESI-DN : 29
AnNucA   : -----MGICGKLGVAALVALVICSPVQSQVPPLETLSPSISVHLLLG : 42
Hum_EndoG : KSIASRLRGSRRFLSGFVAGAVVGAAGALAAQFFRSQGAEGALTGKQPDGSAEKA-VL : 62
CaenEndoG : IAGISFLAGKYSNDLPIFRNVQSATNVPMNQIQVSEPMTVKPSALNADAMGPSRSA-EI : 69
AGAP007845 : LERKRLTIASDYINTGKAAESFCSKPGLPIFGITVSAATPIPAD-SPPSAMVNVSRVG-QI : 84
DmEndoG   : GAFYLGTHVERERQHNGSTSGLPRLPGLPTFTGTVSAASLIPAQENNVSLTATPSRIG-QI : 76
Calb_Nuc  : MSKLLVNTLSLGTIGGTALYFWGKSSIPTTPVQSPNLPAPITNPSGGVDFPSAINPQ-GF : 59
Sac_Nuc   : MCSRILLSGLVGLGAGTGLTYLLLNKHSPTQIETPYPTQKPNNSIQSHSFNVDPG-GF : 59
Sr_Nuc1   : -----MPIIRRTSRATASAEPPSSLHRLMNPIRRNTLSARASS-DI : 39
Cun_NucC1A : -----SPNSE-SI : 7
Gam_Nuc   : ----LYRQLGNTSLERDSEYLSLQKPVENFCESSIQYNVLPPLGGNEEDEEDSIG--GV : 138
CuquEndo  : -----LTSPSS-----INAKRFDKAKPPDVTERDL-NRECNKGSS-----KT : 113
Lundep    : -----VETITCNTN-----FDLASYSNKSTSTDTIET-EEVCGGSG-----KV : 122
Gm_Tsal1  : -KNGEESVTLMCHNNEFD--DGQGNGLDLFTCVKTPTAELRKT-KERCSLGDL-----YV : 131
Gm_Tsal2  : KSNQGETTTLSCNGNQFE--SAAKEKIKDMSKSMKAVVEQT-TKRCMGNDF-----NL : 127
Dm_CG9989 : -----TTFLCRYDSVF-----SPALSSAAPPDPVVVKVP-DTSCSIPS-----AT : 110
Dm_IP10440 : ---SVFKETVQDNGQF-----SVPLPMLRLSPMQPVTKHIRDGDGAG-----NL : 116
Dm_LP23408 : ---RDIVATTQAGRVPAF---QPPLPMTVRAAPAAITTPVQDRRCPA-----TM : 106
Dm_CG14118 : LDAKLRIRVKVQDITTFEWMGAKIHFSDFVNRHSMPYTVTERL-DRSCGSDTPSPSTSYL : 170
Dm_CG14120 : PFVNRWQVTATCLQNKYFLVDDLIYPFANFSCATAWPIFTALRS-GKDCNGGT-----DL : 151
Dm_CG3819 : -----SVTAKVGGTTFKIDKEDHLSAIKCTSWPVFVGKKS-GSSCNGGT-----TL : 132
AGAP011048 : --PTQRSITIQCVTNSTFSFGSPYPMKSLACTSYWLSSAKTT-QARCHNES-----VI : 140
AGAP011047 : --PSETSIIVTQVLDDQFNYESKMYAFTEFGTANWRSVARRT-ANRCYNDA-----TI : 151
LOC4812298 : --LSGKSVTAKVEGTTFRIGDKDHDVSDIKCSSWPAFVGKKS-GSSCNGGT-----TL : 151
Dm_CG6839 : ---SENLLTASVSGTTFVSGGNSFEFKDLYKSWPGFKAVKS-GATCNGG-----IV : 157
Par_DSN   : --LGMEAVEAKCAGGVMLAIEGTEWEIWSLGSNHNKVTIRRN-LGTGCEADQG----DR : 134
Uca_NUC   : --LGEAEVQARCLSSGLLTIGDAEWDLASLGCSSDVKETIFRD-LGTGAGGVG----IL : 133
Mj_Nuc    : --LGTAVDVQCGGNLLVVDGTEWRMDELGCCKKDKESIHRN-LGSCGDGGVG----VF : 130
Pal_Nuc   : SRSRGCWREPVLAACLITGDGTDNAINELGCVTPHLRACRRTWVPVRCR--SW-----YL : 132
Cmae_Nuc  : AVFGEMVVEATCLGGIMLDVNGKNKSLKEILGCSQKIRSTIKKN-LGSCGAENLG----QL : 129
Hame_Nuc  : --LDVAVVEAVCAGDLLAVAGTEWQLFELGCTEKLSESIHRD-MGSCGSDMG-----VI : 131

```

```

          *           200           *           220           *           240
SmNucA   : CAVGCPGT--GSSNVSIVRHATLNNNSTTKFANWVAYHITKDTP---ASGKTR---NWK : 81
AnNucA   : NPSGATPTKLTDPDNYLMVKNQVAFSYNNSKGTANWVAVQLNSSW----GNAERQ--DNFR : 97
Hum_EndoG : EQFGFPII---GTEARCYTHALSYDQAKRVPRWVLEHISKSKL---MGDADRKHCKFK : 115
CaenEndoG : MKHGYPGF----TNVRTYEDVLSYDYKTRTAHWVCEHLTPERL--KHAEGVDRKLCEFK : 123
AGAP007845 : MKYGFPGI----DNVRSFDDVLSYDRRTRVAHWVFEHLTPATV--KHNDAVDRAKCDFK : 138
DmEndoG   : MKYGFPGI----DHVRSHSDVLSYDRRNRVPHWVFEHLTAEV--AKNDAVDRSKCDFK : 130
Calb_Nuc  : FKFGFPGP----IHDLANRTEVSCYDRSTRNPYVWIEHITKASLAKAESNGDRKRSVFK : 115
Sac_Nuc   : FKYGFPGP----IHDLQNRREHISCYNRQTQNPYVWLEHITPESL--AARNADRKNSFFK : 113
Sr_Nuc1   : LKLGNGPGP----VSDLLERSGYLISYNRRDRLAHWVGEHLTSASL--QAGQGVDRDKSNFQ : 94
Cun_NucC1A : LQFGDPGT----ARDFLERESVLSYNRRDRVASWTGEHLTADSL--KTGDGVDRDHSKFK : 62
Gam_Nuc   : SAFGYDIK-----ETFYTIKFEFDKTKMMSRVDDHILHGREL--LKAEPHRD--KFR : 188
CuquEndo  : FQMGFRAS-----GGLFVPYVLELCYLMDTCSVSHVKHRLLGASF--GGVSGTREK-SFQ : 164
Lundep    : YKVGFPPI---SGNFHSIYQTCFDKKNLTPLYSIHILNGQAV--GYHLKHTRG--SFR : 173
Gm_Tsal1  : YDVVFRIS----EEELVGPVYDLCYNEYSQKPAYSRNIINGAAV--NYRVPESETDLTV : 184
Gm_Tsal2  : YEAGYKVN-----GKFYGSVYDLCYNGKQSSGYTHNFIYGRTW--KYKLPEKPYEHYS : 179
Dm_CG9989 : FAVGFSFN-----GRFELVYRNFQDGYSLA--FQ--HSIYKAYR--YVNTVPRNPNTWQ : 158
Dm_IP10440 : YAVGYTID-----GKDLELVRTCFDCGGRLVYSQSDVYKTF----FPKRFVVEFV : 164
Dm_LP23408 : YRVGYNVG-----NNQFLELYRACFDTRAVRAIFVEHRVYKPF-----YITRCPVQFS : 155
Dm_CG14118 : YRVGYDTG-----DGRFVATMELCHDPNQLRTHYAHHQLTPANV--HFQKKLKR--RFS : 221
Dm_CG14120 : VQVGFVEVE-----DGGFLQSYELCHDAEAEATRYVHHVLYPSSY--DYQHGVARV--NFL : 202
Dm_CG3819 : IKVGFELSGSR-----FATQVEVCFNEDEEVTRYVYHRLEPGNN--YYATGVDR--TFG : 183
AGAP011048 : VKVGFELA-----DARWVNVEDVCYDEQLYHTEVVRHYMNRANG--GYQSGNPR--GWY : 191
AGAP011047 : VETGFEMG-----ARFPMKIMDVCHDEVTYDNHYLVHEFTPANA--GFQTGVPRP--GWI : 201
LOC4812298 : IRVGFELD-----GGRFATEVCFNEDEEVTRYVYHTLAPGNN--YYATGVDR--TFG : 202
Dm_CG6839 : IRVGFELIT-----SSRFAEQMQVCFNEEEVTRYTRHKLEPGSN--YYETGVARI--TFQ : 208
Par_DSN   : HSI GF EY Y----GGSIY YELISVCFGPVSETTLRTEHVHLHGANA AAKDIETS RP--SFK : 187
Uca_NUC   : NGIGFQIF----SLNYDKVIVNCFEAASETTLFTHILHGADIAAKDVEASRP--SFK : 186
Mj_Nuc    : EGI GF E I F----GSDS F YELIRVCFEPKAETTLYSEHVLHGANA AAKDIDSS RP--SFK : 183
Pal_Nuc   : HAVGFNIA-----TTGSFHEFESTCFDHAETTLYTKHTLHGANA IAKDVDPSRP--PFK : 185
Cmae_Nuc  : HII GF QFA----NKFEYQVLVCFDHDKQATLYTRHLHGANA IAKDIDSS RP--SFK : 180
Hame_Nuc  : HSI GF E I F----GPFDFYELIRVCFDPKAETTLYSENVILHGANA IAKDKDNS RP--GFK : 184

```

Рис. 11. Продолжение

```

                *          260                    *          280                    *          300
SmNucA       : TDPAL-NPADTLAPADYITGANAA-----LKVDRGHQAFLASL : 117
AnNucA       : PKTLTPAGWVVRTPSMYSGSG-----YDRGHIAFSADR : 130
Hum_EndoG    : PDPI-NPTFSAFNEDYVGS-----GWSRGMHAFAGNN : 147
CaenEndoG    : PDITF-PQKFLSQNTDYKCS-----GFDRGHLLAAGNH : 155
AGAP007845  : PDES-IHPFFRSLNTDYKGS-----GFDRGHLLAAGNH : 170
DmEndoG      : QDESI-HPFFRSQNTDYRRS-----GYDRGHMAAAGNH : 162
Calb_Nuc     : DEEAI-PLKFRNRLRDYFRS-----GFDRGHQAFAADA : 147
Sac_Nuc      : DEEVI-PEKFRGKLRDYFRS-----GYDRGHQAFAADA : 145
Sr_Nuc1      : EDTDI-PEMFRAHLKDYVSS-----GYDRGHQAFAADD : 126
Cun_NucC1A   : EDPDV-PSLFRSTLADYSGS-----GFDRGMHAFAGDA : 94
Gam_Nuc      : SAVLPWGSITKMKCYSNQNDAVLQHFQDQG-----SNEKRKYFARGHLAAANADF : 240
CuquEndo     : SFVCP----GLRYNTHYTQASQKQDPDR-----FYFQRGHLEFDRDA : 202
Lundep       : TNGIY---GKVNIDKLYKTQITKFNKFLGPKQTFFRRPL-----NFLSRGHLSFVDF : 223
Gm_Tsal1     : SLAKD--FTTRDVNNYFKLENQKRFQGYTIDGKPLVDDKN-----FFAFQQLAFDTSM : 236
Gm_Tsal2     : SRDPQ---AGKDLKLYKEQKRFRKNTKVNGKPLLDDEH-----YFTEGQTLTSDTSI : 228
Dm_CG9989    : SGLQS-----GGFVNAYEGRATQACLLTLNLG---AVQPQ----CKFDRGHMTPASAF : 203
Dm_IP10440   : ADEM-----SPQEAAYMKSNIYFAFKCIYGDDQSYLQYQNY----LVINRGMHVASADF : 216
Dm_LP23408   : SJGVI----SGADEASYTVRNIGHGTFRRLFGNNQNYIPNNRD----VIINRGHLSASADF : 207
Dm_CG14118   : TAGHF---TGFDARIYSPPKSEKLMVPLIDVKS-----GLFLARGHLTAKADL : 268
Dm_CG14120   : ELDFY---GGRDVTNRYTQVQNTITISNILGLD---ASPYYFNFSDR-RILSRGHMIAKTDQ : 256
Dm_CG3819    : AGGYF---AGKNDKLYTQAVQKETIDKELDMD---SSRFFDSAKN-IFLARGHMGAKADF : 237
AGAP011048   : QGAYY---TGVNINTLYTVNRRETITATILNSQARADVLVQNTNGIYMARGHIAARADF : 248
AGAP011047   : QGNFY---PGVTVNTLYTVNMORETIATILDSQPRADELVQNTNNGIYMARGHIAARADF : 258
LOC4812298   : AGGYF---AGKNDKLYTQVTKETIDKELDMD---SARYFDSAKN-VFLARGHMGAKADF : 256
Dm_CG6839    : TAGF-----DGNVDKLYTQATLETINNELGGD---AEKYFDSSSN-VYLRARGHLAGADF : 262
Par_DSN      : TSTGF---FVSVMSTVYSQASLQQLMTDILGSDLANIIDPSQQ-LYFAKGHMSFDADF : 243
Uca_NUC      : TSTGF---FVSVMSTVYSQNSLQQLMSTILGDEDPANTIIDPSKQ-LYFAKGHMSFDAGF : 242
Mj_Nuc       : SSTGF---FTVSMSTCYTQNSLASMKILLGDDDLANAIINPHEQ-YFPAKGHLSFDADF : 239
Pal_Nuc      : PTGF---FTVEVNTVYTQASLALMEQLLGDSALANQIINPDQE-LFMSRGHLSFDADH : 241
Cmae_Nuc     : TSSGF---FNVSMSTVYSQNSLQQLMSTILGDEDPANTIIDPSKQ-LYFAKGHMSFDADF : 235
Hame_Nuc     : SSTGF---FTVSMKHYVTEKSLNLMIDLGLGEDLAHKIIDVHKQ-FYLRARGHLSDADF : 233

                *          320                    *          340                    *          360
SmNucA       : AG-VSDWESLNLSNITPCK-----SDLNQ--GAWARLEDQBRKIIDR----- : 157
AnNucA       : TKTTEDNAATELMTNMMPQTP-----DNNR--NTWGNLEDYCRELVSQ----- : 171
Hum_EndoG    : KFSSKAMAE TFYLSNIVPOD-----FDNNS--GYWNRIEMYCRELETER----- : 188
CaenEndoG    : RKSQAVDQTFYLSNMSPOVG-----RGFNR--DKWNDLEMHCRRVAK----- : 196
AGAP007845  : RMEQKHCDETFYLTNMAPQAREMDWFLGFNR--DKWNDLERYVRKTK----- : 216
DmEndoG      : RLHQKHCDETFYLSNMAPQVG-----QGFNR--DAWNTLEAHVRRTK----- : 203
Calb_Nuc     : KFNQLAMDETFYLTNMCPQVG-----DGfNR--DYWSHFEDFVRRTN----- : 188
Sac_Nuc      : KFSQQAMDDTFYLSNMCPQVG-----EGFNR--DYWAHLEYFCRGTK----- : 186
Sr_Nuc1      : ISSQEAMDETFLLSNMAPQVG-----VGFNR--HYWAYLEGFMRDLTON----- : 168
Cun_NucC1A   : VATQPAMDQTFYLSNMSPOVG-----IGFNR--HYWAYLEGFORSTTKK----- : 136
Gam_Nuc      : VS-QDEQKASYSFANVAPQW-----QAFNN--GNWKKLENRIETLAKM----- : 280
CuquEndo     : PV-FAWKLATYSGNCVPOW-----NTINSEGGNWWTLEYLVFKYAQQT----- : 245
Lundep       : TF-RREQHATEMYINTAPOY-----QSINQ--GNWLRVENHVRDLAKV----- : 263
Gm_Tsal1     : IT-PADKLSYDYANIAPQY-----KTvyD--GNVWRVENITRDLAVN----- : 276
Gm_Tsal2     : IT-GADKLSYDYANIAPLF-----KDIYD--GNIWRYENMTQELADQ----- : 268
Dm_CG9989    : IS-TELKKS TFRYINAIPQY-----RGVNR--CKWKAVETWVNNVVRGLYDNPIINNVO : 254
Dm_IP10440   : LF-TDQMGSTFRYLNVVPOF-----KSIND--GNWEKIERWVRSQIPK----- : 256
Dm_LP23408   : FF-GDQLCATFKYVNAVPOF-----KSIND--GNWETIERFVRSVVTG----- : 247
Dm_CG14118   : IY-ASQOKSSFNYMNAVPOW-----QSFNG--GQWSKLEESTROYVAR----- : 308
Dm_CG14120   : IF-GAAQHTFEFINVAPQW-----QTFNG--GNWEKVETSVRKVFAD----- : 296
Dm_CG3819    : VFAPE-QRATFEFINAAPQW-----QTFNA--GNWARVEDGVRRAWAKENKHVECWTGv- : 288
AGAP011048   : VY-GTEQNA TFWFINAAPQW-----QNFNG--VNWERVESSVRDFVVGK----- : 288
AGAP011047   : IY-ATQQNATFWFINAAPQW-----QNFNA--GNWERIESSVKSFVAW----- : 298
LOC4812298   : VY-APEQRATFEFINAAPQW-----QTFNA--GNWARVEDGVRRAWVSK----- : 296
Dm_CG6839    : DY-APEQRATFEFINAAPQW-----QTFNA--GNWARVEDGTLRAWVSK----- : 302
Par_DSN      : VT-VAEQDATYFINALPOW-----QAFNN--GNWKYLEYATRDLAES----- : 283
Uca_NUC      : VT-IASQDATYFINALPOW-----QAFNN--GNWKYLETNTFNLAMK----- : 282
Mj_Nuc       : VT-EAEQDATYYYINAVPOW-----QAFNN--GNWKYLEFATRDLAES----- : 279
Pal_Nuc      : VL-IAEQDATYYYINAVPOW-----QAFNN--GNWKYLEFAGRDLAVA----- : 281
Cmae_Nuc     : VT-LAEQDATYYYINAVPOW-----AAFNN--GNWKY----- : 264
Hame_Nuc     : ----- : -

```

Рис. 11. Продолжение

| | | | | | | | | | |
|------------|---|--------------|------------|----------|---------|---------|-------------|------------------|-------|
| | | * | 380 | * | 400 | * | 420 | | |
| SmNucA | : | --ADISSVYTV | GPLYERDMG | ----- | KLPGTQ | ----- | KAHTIPSA | YWKVIF : 196 | |
| AnNucA | : | ----GKELYIV | AGPNGSLGK | ----- | PLKG | ----- | KVTVPKST | WKIVV : 204 | |
| Hum_EndoG | : | ----FEDVWVVS | GPLTLPQTR | --G | DCK | --- | KIVSYQVIGE | ----- | |
| CaenEndoG | : | ---KMINSYII | IIGPLYPKLE | --G | DCK | --- | KYIKYQVIGDN | ----- | |
| AGAP007845 | : | ---QYPNVYCC | TGPLYLP | PRRE | --D | DCK | --- | LYVKYQVIGSN | ----- |
| DmEndoG | : | ---TYSNVYVCT | GPLYLPHKE | --D | DCK | --- | SYVKYEVIGAN | ----- | |
| Calb_Nuc | : | ---KYDDVRVM | IIGPLYL | PKLG | --A | DCK | --- | YRVTYEVIGSP | ----- |
| Sac_Nuc | : | ---KYKSVRIV | TGPLYL | PKKD | -PI | NK | --- | FRVNYEVIGNP | ----- |
| Sr_Nuc1 | : | ----FTDVYVY | TGPLFLPSAA | --S | T | R | --- | KNPAYSEIYPFLGATT | PNV |
| Cun_NucC1A | : | ----FSDVYVFI | TGPLFLPTK | --S | D | C | --- | YTVTYNVLQG | ----- |
| Gam_Nuc | : | ---KEATLRVVI | GTSTGLLKL | NSSS | Q | LENVCL | CDGE | ----- | |
| CuquEndo | : | ---PNTEYTI | YDGVV | YDPR | --- | NQYLAK | QEK | ----- | |
| Lundep | : | ---LQKDI | TVVIGILG | LRLKSK | KIEK | --E | IYLGDD | ----- | |
| Gm_Tsall | : | ---RQAKFEVY | IGGYERL | TAL | --H | NCKEEE | IFLSAK | Y | |
| Gm_Tsall2 | : | ---RQATFEY | YIGGFYS | YEVE | --- | KWKPIGL | DNAEFP | ----- | |
| Dm_CG9989 | : | I | PRTYDV | LKVCIG | ALGVHRL | RHNT | NMNIPI | YLLDNN | |
| Dm_IP10440 | : | ---SSYFRVK | SGGIGL | TLP | -D | T | R | CFLOSA | |
| Dm_LP23408 | : | ---NNFVNV | RITGARG | LSLP | -S | G | R | PKNVFL | |
| Dm_CG14118 | : | ---SGITAT | VYIGTYGEM | KVA | --- | GSKVLH | MTTANN | NI | |
| Dm_CG14120 | : | ---RNLTTD | CYIGTWG | VSTLP | -D | V | D | GI | |
| Dm_CG3819 | : | --WGVTTLP | PNKNGE | QRQLYLS | -H | D | NGNGLI | PVPKLYF | |
| AGAP011048 | : | ---RDLE | ITVYSGTYG | VQKLA | -D | G | N | CDG | |
| AGAP011047 | : | ---RNIRVR | VYGGTYG | VQTQA | -D | G | N | DHREIFLDFDPN | |
| LOC4812298 | : | ---EKKH | VECWIGV | WVTTLP | -N | D | C | K | |
| Dm_CG6839 | : | ---NKLNV | NCYIGVY | VTTLP | -N | K | D | C | |
| Par_DSN | : | ---HGS | DLRVYSG | WSLQLD | -D | I | N | C | |
| Uca_NUC | : | ---KGR | DLRVYSG | WDLQLD | -D | I | N | C | |
| Mj_Nuc | : | ---HST | DLTIYIG | GWGLTLD | -D | I | N | C | |
| Pal_Nuc | : | ---HGT | DLTIYD | CGWGLQLD | -D | I | N | C | |
| Cmae_Nuc | : | ----- | ----- | ----- | ----- | ----- | ----- | ----- | |
| Hame_Nuc | : | ----- | ----- | ----- | ----- | ----- | ----- | ----- | |

| | | | | | | | | |
|------------|---|---------|---------|----------|--------|-------|--------|-----------|
| | | * | 440 | * | 460 | * | 480 | |
| SmNucA | : | INNS | ----- | PAVNHYAA | ELF | -D | QNTPKG | ADF |
| AnNucA | : | VLDSPG | SGLEGIT | ANTRVIA | VNI | P | N | DEL |
| Hum_EndoG | : | ARRS | ---SV | STEPLAL | GAFV | V | P | NEAIGFQ |
| CaenEndoG | : | FEVT | ----- | PGKFELES | YILP | N | AVIEDT | VEI |
| AGAP007845 | : | METS | ----- | DGKLEMES | YVLP | N | Q | KIDNGTPI |
| DmEndoG | : | GESA | ----- | DHKLHMES | YVMP | N | Q | VISNDTPI |
| Calb_Nuc | : | GETQ | --- | GSEKISCA | AFVLP | N | D | VIDNSTPL |
| Sac_Nuc | : | AEAPTAN | PARE | DI | AVAA | F | V | L |
| Sr_Nuc1 | : | TTTA | --- | SSEYAL | GAFVLP | N | Q | AIDSSSTPL |
| Cun_NucC1A | : | VPQG | --- | DNKYAY | GAFVLP | N | Q | AIDTKTPL |
| Gam_Nuc | : | ALR | ----- | QALMPN | H | P | A | AGISAR |
| CuquEndo | : | NQTH | ----- | VLKVFE | VSHNL | K | G | T |
| Lundep | : | DPQK | ----- | QEATVE | VSSNN | P | H | V |
| Gm_Tsall | : | DKEN | ----- | DAGIVE | F | T | L | N |
| Gm_Tsall2 | : | DTES | ----- | KDGIVE | V | T | L | N |
| Dm_CG9989 | : | SHLS | ----- | GDKWV | M | L | T | N |
| Dm_IP10440 | : | DATG | ----- | NGLYV | F | L | T | N |
| Dm_LP23408 | : | NANN | ----- | QPIVA | F | L | T | N |
| Dm_CG14118 | : | DEGH | ----- | PTRGIA | L | V | G | N |
| Dm_CG14120 | : | DRVT | ----- | REGIVL | I | G | I | N |
| Dm_CG3819 | : | ----- | ----- | LIGVNN | P | H | L | S |
| AGAP011048 | : | DEAS | ----- | NAGIAL | V | G | N | N |
| AGAP011047 | : | NEAQ | ----- | NSGIVL | I | G | N | N |
| LOC4812298 | : | EPST | ----- | KKGIVL | V | G | N | N |
| Dm_CG6839 | : | DPSS | ----- | HRGIV | V | G | N | N |
| Par_DSN | : | EESS | ----- | SKAAA | I | V | G | I |
| Uca_NUC | : | DEST | ----- | NCAVA | V | V | G | N |
| Mj_Nuc | : | EESS | ----- | SRAVG | V | V | G | N |
| Pal_Nuc | : | EEST | ----- | NRAAA | V | I | G | I |
| Cmae_Nuc | : | ----- | ----- | ----- | ----- | ----- | ----- | ----- |
| Hame_Nuc | : | ----- | ----- | ----- | ----- | ----- | ----- | ----- |

Рис. 11. Продолжение

| | * | 500 | * | | | | | |
|------------|---|-------------|------------|------------|---------------|-------|----------|-------|
| SmNucA | : | EKRTGLIIWAG | PDDVQASLKS | PGVLP | PELMGCKN | -- | : 266 | |
| AnNucA | : | ESLTGYDFLSN | SPNIQTSI | ESKVDN | ----- | | : 273 | |
| Hum_EndoG | : | EKLSVLVFFPH | DRTSDIRN | ICSVDTC | CKLLDFQEF | -- | : 307 | |
| Caen_EndoG | : | ERSAGLEIFAR | DPKSIVK | ENGAKK | ----- | GGLLW | -- : 308 | |
| AGAP007845 | : | ERAAGLLFFDK | A---RNKL | AKINGKKV | ----- | | : 323 | |
| Dm_EndoG | : | ERSAGLLFFDQ | N---RKQL | TTINGKKV | ----- | | : 310 | |
| Calb_Nuc | : | ERSSGLDLLQK | PYYNKQD | LCKQVKE | IIIVREFPKQV | | : 309 | |
| Sac_Nuc | : | ERSTGLELLQK | PPSKKKAL | CKEVN | CQIVVRDFSNA | A | : 312 | |
| Sr_Nuc1 | : | EKAAGLVFFDK | DRSKFAD | LCSKTTCQVR | ----- | | : 284 | |
| Cun_NucC1A | : | EKASGLTFFDK | DVSTLGD | LCAATTCAV | SSSSGGGD | -- | : 251 | |
| Gam_Nuc | : | --TRGALY | CLTPGD | L-CGVEPK | ACGYKDLM | ----- | : 393 | |
| CuquEndo | : | EASRGITHCC | YWKDLAR | KNPLPIE | VQREINNKN | --- | : 364 | |
| Lundep | : | NYSIGLTICCK | EEFVKRN | KIILPK | EVNKNKNTK | KL | : 387 | |
| Gm_Tsall | : | DWTQGYTFCC | TWNNVKD | YVRALPQ | DIENNLLKFN | - | : 399 | |
| Gm_Tsal2 | : | TQSAGSYGDF | GNRIRTL | PKDIYVK | ----- | GLLKY | -- : 388 | |
| Dm_CG9989 | : | TKDVGH | TVC | DPYRFIT | INAPHLTGVC | ----- | : 370 | |
| Dm_IP10440 | : | -PNDGYTFCC | DPKRFPY | ----- | | | : 356 | |
| Dm_LP23408 | : | -AQAGFSFC | NPATFRP | ----- | | | : 346 | |
| Dm_CG14118 | : | NLKNGLYACS | ANLARAV | GHLPRP | LLEVD | ELLT | --- | : 439 |
| Dm_CG14120 | : | DLHEGYSYACS | VEDFIEV | VKDLPLE | DLHTNGILGL | -- | : 426 | |
| Dm_CG3819 | : | DITAGYSYACE | VPEFR-- | KKVTHL | PEFSVSGLLV | --- | : 404 | |
| AGAP011048 | : | NLTIGYCYACE | NAFNDA | IGRPHQ | LNVA-AKLLTSSG | | : 420 | |
| AGAP011047 | : | NLSRGYSYACE | NEFNRT | GHLPLQ | LNVA--SLLI | --- | : 425 | |
| LOC4812298 | : | DISAGYSYACD | VAEFRKK | VHDHLPQ | FAVS--GLLV | --- | : 423 | |
| Dm_CG6839 | : | DIKAGWSYACE | ADFLKT | VKHLPAL | TAKG-GLLV | --- | : 430 | |
| Par_DSN | : | DLAHGYTYCCA | VDDL-RQ | AIPYIP | DLGNV-GLLTN | -- | : 406 | |
| Uca_NUC | : | SLASGYTYRCS | VAEL-RAS | VPHVPD | LGNV-CLLTD | -- | : 405 | |
| Mj_Nuc | : | DLGHGYTYCCT | VDDL-RAA | IPHV | DLGSV-GLLDK | -- | : 402 | |
| Pal_Nuc | : | DLFRGFTYCCT | VDDL-RAA | IPHV | DLGNV-GLLDS | -- | : 404 | |
| Cmae_Nuc | : | ----- | | | | | : - | |
| Hame_Nuc | : | ----- | | | | | : - | |

Рисунок 11. Множественное выравнивание аминокислотных последовательностей Par_DSN и ее гомологов из других членистоногих с нуклеазами SNF. Множественные выравнивания последовательностей были подготовлены с помощью программного обеспечения Clustal X с визуализацией в программе GeneDoc. Консервативные остатки выделены черным цветом — процент консервативности выше 80%, темно серым — 50%, светло серым — 35%. Подчеркнут участок, который использовали для построения филогенетического дерева.

Высококонсервативный участок нуклеазных доменов использовали для построения дерева филогенетического родства с помощью метода максимального правдоподобия.

На филогенетическом древе (рисунок 12) изучаемые белки разделились на две большие клады. Первая включала в себя «классические» SNF нуклеазы. Вторая клада, названная нами подгруппой DSN-подобных нуклеаз, состояла из нуклеаз насекомых и ракообразных. В свою очередь, эта подгруппа делилась на несколько подклад, при этом нуклеазы ракообразных, за исключением нуклеазы рачка бокоплава Gam_Nuc, группировались на дендрограмме вместе.

Хотя большая часть нуклеаз насекомых оказались в одной кладе с Par_DSN, предсказанные белки DmEndoG и AGAP007845 вошли в кладу «классических» SNF нуклеаз, что характерно для митохондриальных эндонуклеаз G. Интересно, что нуклеаза CuquEndo, аминокислотная последовательность которой имеет чуть большее сходство с SmNucA (26% идентичности и 42% сходства), чем с Par_DSN (25% идентичности и 40% сходства), оказалась в одной подгруппе с DSN-подобными нуклеазами. На основании анализа филогенетического древа и данных об известной субстратной специфичности некоторых нуклеаз нами было выдвинуто предположение, что DSN-подобные нуклеазы и SNF нуклеазы эволюционировали по

дивергентному пути, о чем свидетельствуют их различные биологические функции и субстратные специфичности.

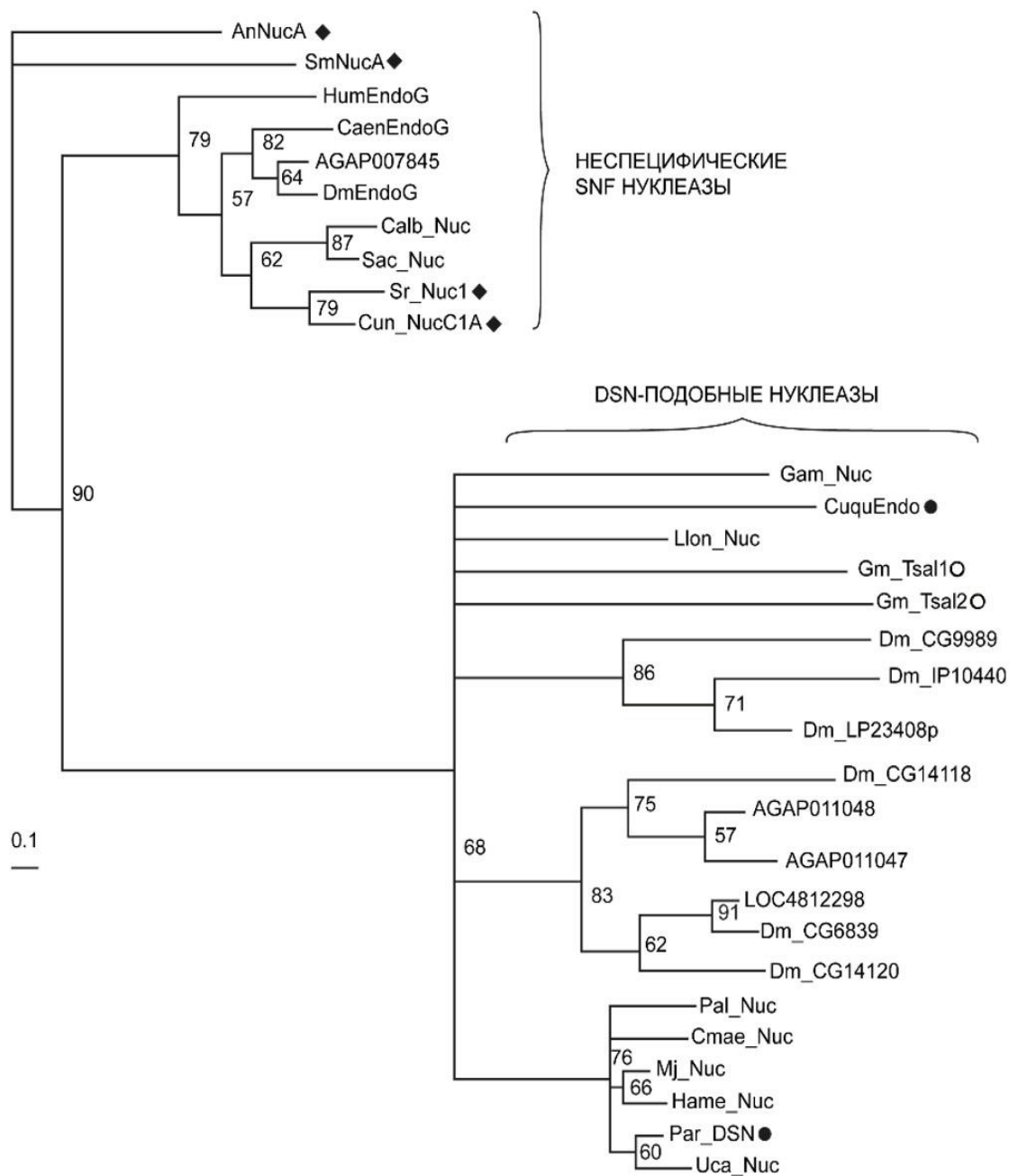
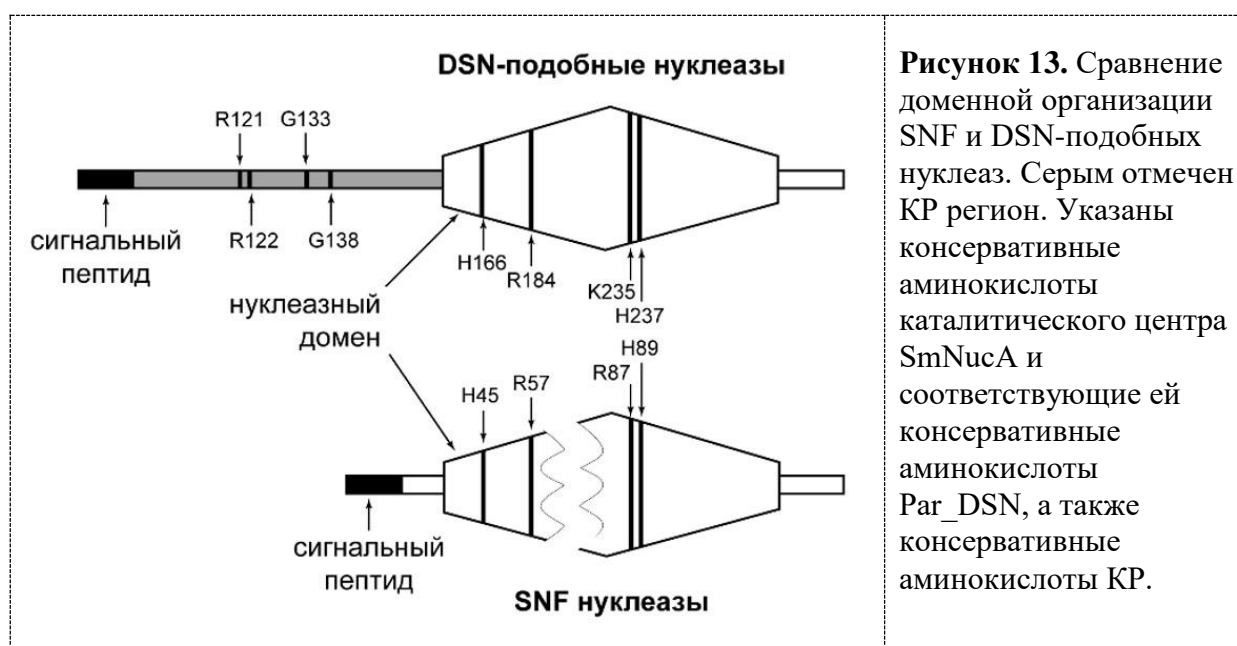


Рисунок 12. Филогенетическое древо максимального правдоподобия, построенное на основе множественных выравниваний нуклеаз (рисунок 11) с использованием алгоритма «quartet puzzling» с 10000 шагов. Черными кружками отмечены нуклеазы, имеющие подтвержденную субстратную избирательность к дц ДНК, белыми — нуклеазы, для которых эта специфичность была подтверждена позднее. Ромбами отмечены SNF нуклеазы, расцепляющие дц и оц ДНК и РНК. Цифрами показана вероятностная поддержка каждой ветви.

Сравнение доменной организации DSN-подобных и SNF нуклеаз по данным множественного выравнивания показано на рис. 13.

Было обнаружено, что у всех DSN-подобных нуклеаз нуклеазный домен более чем на 50 аминокислотных остатков длиннее нуклеазного домена, характерного для

SNF и содержит «вставку» из, по крайней мере, шестнадцати аминокислотных остатков перед консервативным RGH-мотивом. Также у всех DSN-подобных нуклеаз между сигнальным пептидом и нуклеазным доменом находится последовательность длиной 120-180 аминокислотных остатков (КР-консервативный регион) с неизвестной функцией. Родственных ей последовательностей в доступных базах данных обнаружить не удалось. Высококонсервативные участки у DSN-подобных нуклеаз оказались расположены не только внутри нуклеазного домена, определенного множественным выравниванием, но и по всей длине КР. Интересно, что характерной особенностью этих белков оказались консервативные остатки цистеина, распределенные по всей длине выравнивания, которые в том числе были найдены и внутри КР. Мы предположили, что нуклеазный домен DSN-подобных нуклеаз может иметь иную точку начала, нежели нуклеазный домен у SNF.



Для определения возможных границ нуклеазного домена у DSN-подобных нуклеаз был проведен анализ информационной структуры белка по методу А.Н. Некрасова (Nekrasov A.N., 2002). Результат анализа информационной структуры Par_DSN и SmNucA показан на рисунке 14. Анализ информационной структуры Par_DSN выявил наличие шести информационных ассоциаций, первая из которых соответствовала области сигнального пептида. Вторая ассоциация была расположена между 35 и 114 аминокислотными остатками, третья – между 114 и 238, четвертая – между 238 и 260, пятая – между 265 и 325, а шестая начиналась с 336 остатка и заканчивалась в районе 376 аминокислотного остатка, в месте окончания нуклеазного домена.

На основании информационной структуры Par_DSN было выдвинуто предположение, что началу нуклеазного домена соответствует 114 аминокислотная позиция. Менее вероятной представлялась позиция 162, но ее нельзя было теоретически исключить. Для проверки этого предположения были клонированы и экспрессированы несколько укороченных с 5'-конца вариантов фермента.

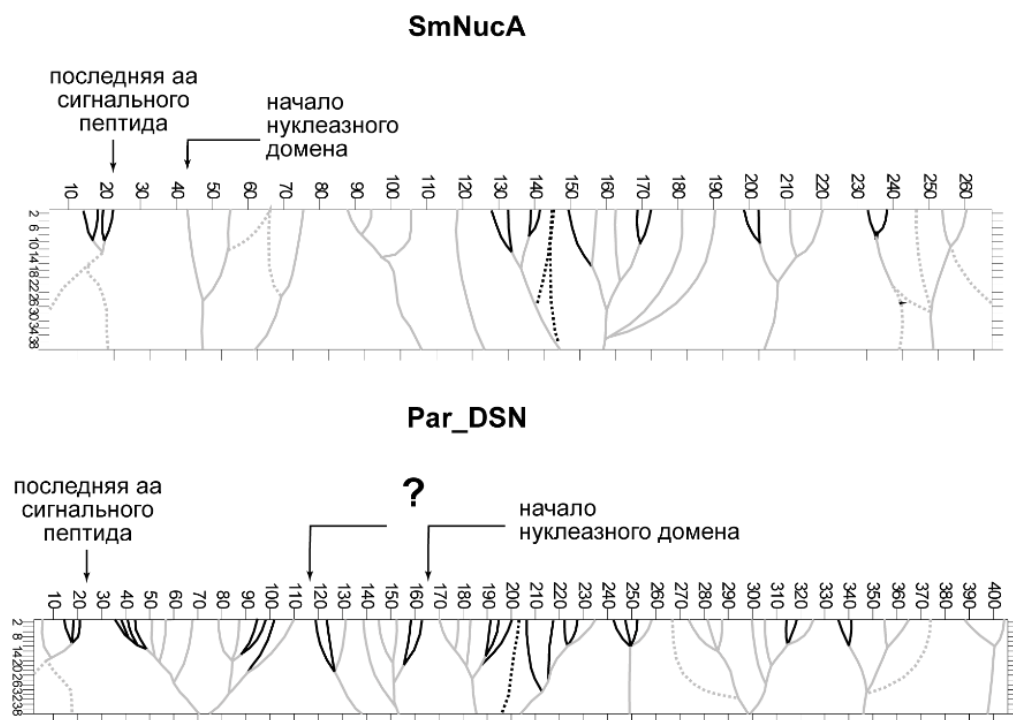


Рисунок 14. Анализ информационной структуры Par_DSN и SmNucA. Показаны пронумерованные аминокислотные остатки полипептидных цепей и образуемые ими информационные ассоциации. Аа — аминокислота.

Первый вариант (Par_DSN-T1) представлял собой полипептид, включающий в себя аминокислотные остатки 192–379 Par_DSN. Эта последовательность соответствовала нуклеазному домену, определенному с помощью программы SMART. Полипептид был очищен и ренатурирован. К сожалению, он оказался нерастворимым и неактивным.

Второй вариант, Par_DSN-T2, содержащий аминокислотные остатки с 149 по 406, соответствовал NUC домену, определенному с помощью множественного выравнивания. После ренатурации и очистки был получен растворимый белок, не проявляющий ДНКазную активность. Инкубация Par_DSN-T2 с протеиназой К приводила к его полному расщеплению.

Третий вариант, Par_DSN-T3, содержал остатки 162–406. Из его структуры был исключен неспаренный остаток цистеина. Как и предыдущий вариант Par_DSN-T3 был растворим, чувствителен к протеиназе К и не обладал ДНКазной активностью.

Наконец, вариант Par_DSN-T4, содержащий аминокислотные остатки с 114 по 406, оказался растворимым белком, способным гидролизовать ДНК субстрат. Однако Par_DSN-T4 обладал свойствами, сильно отличающимися его от Par_DSN дикого типа. Например, он проявлял значительно меньшую каталитическую активность (6 единиц Куница против 6000 единиц активности интактной Par_DSN), имел температурный оптимум при 37°C (против 60°C полноразмерной Par_DSN-T4) и пониженную температурную стабильность. Инкубация в течение 20 мин при 70°C приводила к полной инактивации этого белка, тогда как полноразмерная Par_DSN сохраняла при этих условиях около 60% активности. Кроме того, Par_DSN-T4 оказался чувствителен к обработке протеиназой К. Наконец, соотношение гидролитической активности дц к оц ДНК субстратам для этого варианта не превышало 10.

На основании полученных данных нами был сделан вывод, что нуклеазный домен Par_DSN, необходимый для проявления энзиматической активности, включает аминокислотные остатки с 114 по 406. Фрагмент с 28-го по 113-й аминокислотный остаток, по-видимому, вовлечен в модуляцию субстратной специфичности Par_DSN и важен для селективного связывания дц ДНК.

На основании множественного выравнивания DSN-подобных нуклеаз и представителей SNF было показано, что большинство DSN-подобных нуклеаз содержат консервативные Glu, Trp, Asn, His и Arg остатки, соответствующие Glu127, Trp123, Asn119, His89 и Arg57 SmNucA (рисунок 15). «Замены» в этих позициях были обнаружены только у Tsal белков слюнных желез *G. morsitans* (Haddow J.D. et al., 2002). Остатки аргинина, соответствующие Arg167 AnNucA, были обнаружены во всех DSN-подобных нуклеазах за исключением белков Dm_CG9989, CuquEndo и Tsal2. У Tsal белков не выявлялось аминокислотного остатка, соответствующего Arg122 нуклеазы AnNucA, а у четырех из шести нуклеаз ракообразных в этой позиции находился остаток лизина. Аминокислотный остаток в позиции, соответствующий Asp86 SmNucA, у DSN-подобных нуклеаз оказался неконсервативным. У всех нуклеаз ракообразных в этом месте находился остаток аланина.

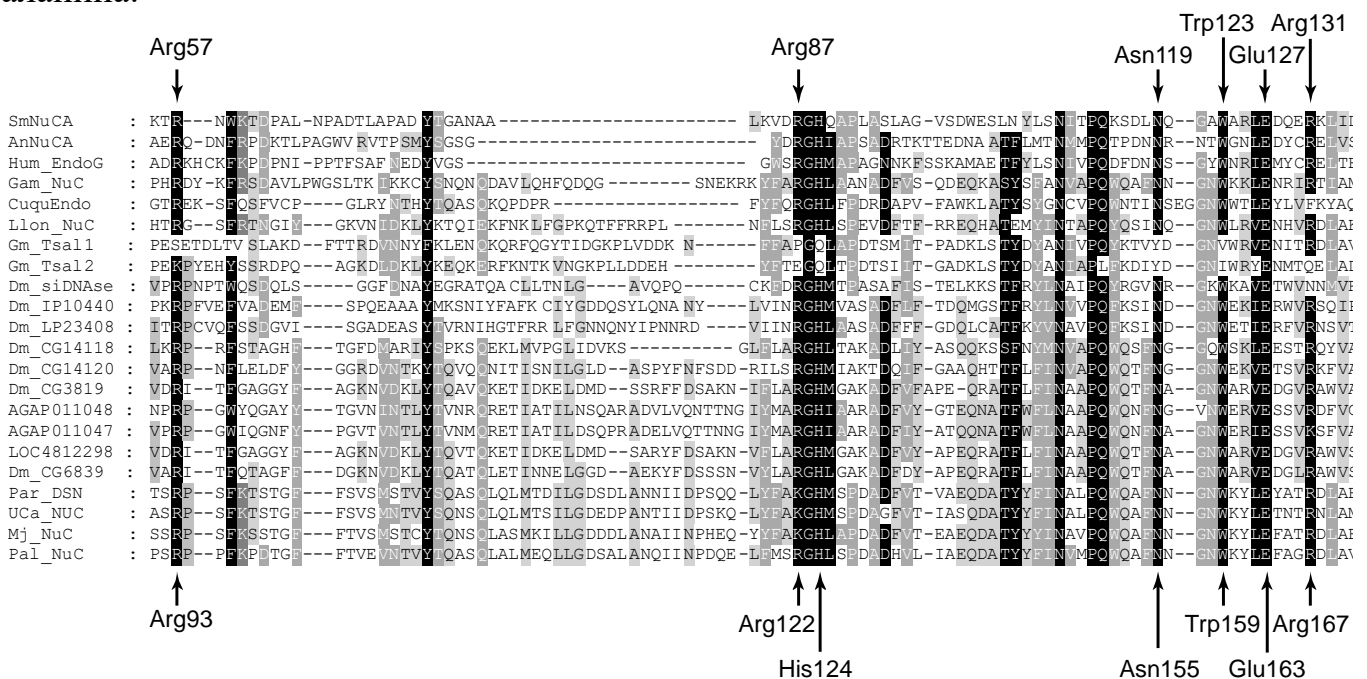


Рисунок 15. Расположение аминокислотных остатков активных центров SmNucA и AnNucA на фрагменте множественного выравнивания аминокислотных последовательностей нуклеаз (полное выравнивание см. на рис. 11). Сверху стрелками указаны аминокислотные остатки каталитического центра SmNucA, снизу — AnNucA.

Наиболее заметной особенностью, характерной для ракообразных, входящих в группу DSN-подобных нуклеаз, является замена в нуклеазном домене высококонсервативного мотива DRGH на AKGH. Предполагается, что в SmNucA остаток D86 косвенно участвует в связывании Mg^{2+} , координируя N119 (Miller M.D. et al, 1999). Было показано, что мутация D86A в SmNucA приводит к увеличению потребности нуклеазы в магнии, при этом без подавления ее ферментной активности (Friedhoff P. et al., 1996). Возможно, замена аспарагиновой кислоты на аланин у

нуклеаз ракообразных определяет особенности металлозависимости, сходной с мутантным вариантом SmNucA.

Для идентификации аминокислотных остатков, важных для проявления энзиматической активности Par_DSN, а также участвующих в формировании ее трехмерной структуры, нами был осуществлен сайт-направленный мутагенез с последующим анализом активности рекомбинантных мутантных белков. Так как устойчивость к протеолизу свидетельствует о правильном формировании дисульфидных связей в белке, нами была изучена устойчивость мутантных белков к протеиназе К. Энзиматическая активность полученных рекомбинантных белков оценивалась по способности гидролизовать геномную дц ДНК и оц ДНК-субстраты (синтетические олигонуклеотиды).

Как и ожидалось, замена His237 (соответствующего His89 SmNucA) на Ala привела к потере энзиматической активности Par_DSN, что говорит о ключевой роли этого аминокислотного остатка в катализе. Замена Lys235, соответствующего вовлеченному в связывание субстрата Arg87 SmNucA, на Ala привела к резкому снижению нуклеазной активности. Такой же эффект был продемонстрирован ранее при замене Arg87Ala для SmNucA (Friedhoff P. et al., 1996). Следует отметить, что мутант Lys235Arg обладал нуклеазной активностью соизмеримой с активностью Par_DSN дикого типа. Все эти замены не влияли на устойчивость фермента к протеолизу протеиназой К. На основе полученных результатов, нами был сделан вывод, что His237 и Lys235 Par_DSN и родственных ей нуклеаз играют ту же роль в работе каталитического центра, что и His89 и Arg87 у SmNucA.

По результатам множественного выравнивания позиция Arg184 Par_DSN соответствует Arg57 SmNucA. Ранее было показано, что у SmNucA замена этого аминокислотного остатка на остатки лизина или аланина приводит к падению активности фермента на 94 и 99 %, соответственно (Friedhoff P. et al., 1996). В отличие от мутантов SmNucA, мутант Par_DSN, содержащий замену Arg184Lys, обладал активностью, сравнимой с активностью Par_DSN дикого типа, а замена Arg184Ala приводила только к 3-кратному падению активности. Обе замены не влияли на устойчивость фермента к протеолизу протеиназой К. Иными словами, в отличие от Arg57 SmNucA, Arg184 Par_DSN, по-видимому, не является существенным для функционирования каталитического центра. Мутанты Par_DSN, содержащие замены His168, соответствующего высококонсервативному у SNF нуклеаз остатку His45 (нумерация по SmNucA), на остатки аланина и лизина, оказались чувствительными к протеолизу протеиназой К и не обладали гидролитической нуклеазной активностью. К такому же эффекту приводили замены His171Asp, Gly139Val и Gly133Val. Был сделан вывод, что эти остатки могут играть важную роль в формировании правильной трехмерной структуры белка. Интересно, что His171 консервативен только у Par_DSN-подобных нуклеаз ракообразных, Gly139 – у всех изученных нуклеаз членистоногих, за исключением Tsall, а Gly133 не консервативен. Дополнительно, была продемонстрирована важная для функционирования фермента роль консервативных аминокислотных остатков в KP области. Мутантный белок 121Ala-122Ala оказался полностью неактивен, тогда как мутанты Arg121Ala и Arg122Ala обладали высокой гидролитической активностью, соизмеримой с Par_DSN дикого типа. Мы предположили, что, по крайней мере, один

из остатков аргинина каким-то образом участвует либо в формировании, либо регуляции каталитической активности активного центра Par_DSN. При замене одного из остатков аргинина второй берет на себя эту функцию. Следует заметить, что большинство Par_DSN-подобных нуклеаз содержат остаток аргинина или лизина по крайней мере в одной из этих двух позиций.

Таким образом, Par_DSN и ее гомологи содержат в своем составе нуклеазные домены, которые длиннее соответствующего домена нуклеаз SNF, а их каталитический центр, хотя и сходен с каталитическим центром SNF нуклеаз, имеет ряд принципиальных отличий. Для нуклеаз ракообразных, в частности, характерна замена консервативного мотива DRGH на AKGH. Между нуклеазным доменом и сигнальным пептидом Par_DSN-подобные нуклеазы содержат регион длиной около 90 аминокислотных остатков, который, по-видимому, важен для определения субстратной специфичности и стабильности фермента. На основании этих отличий, нами было предложено выделить Par_DSN-подобные нуклеазы в новое семейство внутри суперсемейства His-Me нуклеаз. Исходя из определенного сходства членов этого семейства с нуклеазами SNF, мы предположили их дивергентное расхождение в эволюции и постулировали, что нуклеазы с дц-ДНКазной активностью могут быть найдены у других членистоногих. Нами было показано, что нуклеазный домен Par_DSN отделенный от КР области сохраняет каталитические свойства, но в значительной степени утрачивает избирательность по отношению к дц ДНК. Таким образом, среди Par_DSN-подобных нуклеаз можно ожидать как неспецифической активности по отношению к дц и оц ДНК и РНК, так и дуплекс-специфической активности.

Наше предположение было в дальнейшем подтверждено рядом работ независимых лабораторий, которым удалось продемонстрировать активность некоторых ферментов, отнесенных нами к семейству Par_DSN-подобных нуклеаз.

Методы анализа мутаций и выявления ДНК-мишеней в сложных смесях нуклеиновых кислот. Исследование активности Par_DSN против полностью комплементарных ДНК-ДНК дуплексов и дуплексов, содержащих одну нуклеотидную замену, показало, что Par_DSN расщепляет полностью комплементарные молекулы ДНК с заметно более высокой эффективностью. Различия в активности в отношении таких субстратов наблюдались при размерах дуплекса от 8 до 12 п.о., с наибольшей разницей при длине дуплекса 10 п.о. С помощью коллекции 18-звеньевых олигонуклеотидов и 10-звеньевых флуоресцентно-меченого зонда было показано, что тип и положение точечной замены не влияет на эффективность гидролиза субстрата нуклеазой (данные не приведены). Способность Par_DSN дискриминировать полностью комплементарные и частично комплементарные ДНК-содержащие дуплексы позволила использовать ее для разработки метода анализа мутаций (ДСНП-анализ, от англ. duplex-specific nuclease preference). Схема метода показана на рисунке 16.

Для выявления мутантной последовательности используется два 10-звеньевых зонда (далее ДСНП-зонды), каждый из которых содержит флуорофор и гаситель флуоресценции. После расщепления зонда происходит высвобождение флуорофора и появление специфического флуоресцентного сигнала. Один зонд содержит

последовательность нуклеотидов, которая полностью комплементарна ДНК последовательности дикого типа, второй – последовательность нуклеотидов, комплементарную мутантной последовательности.

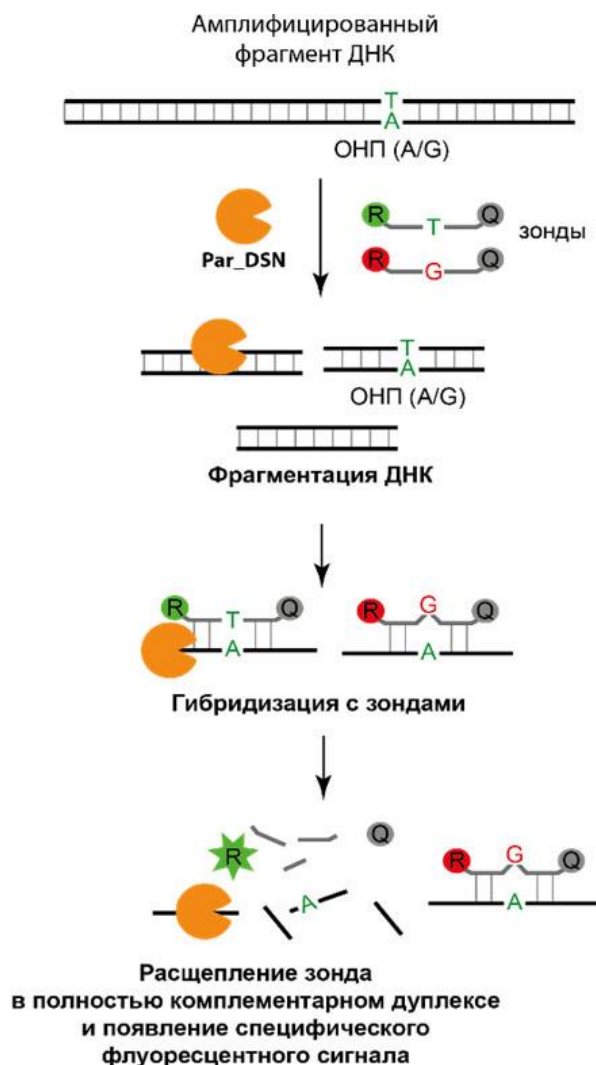


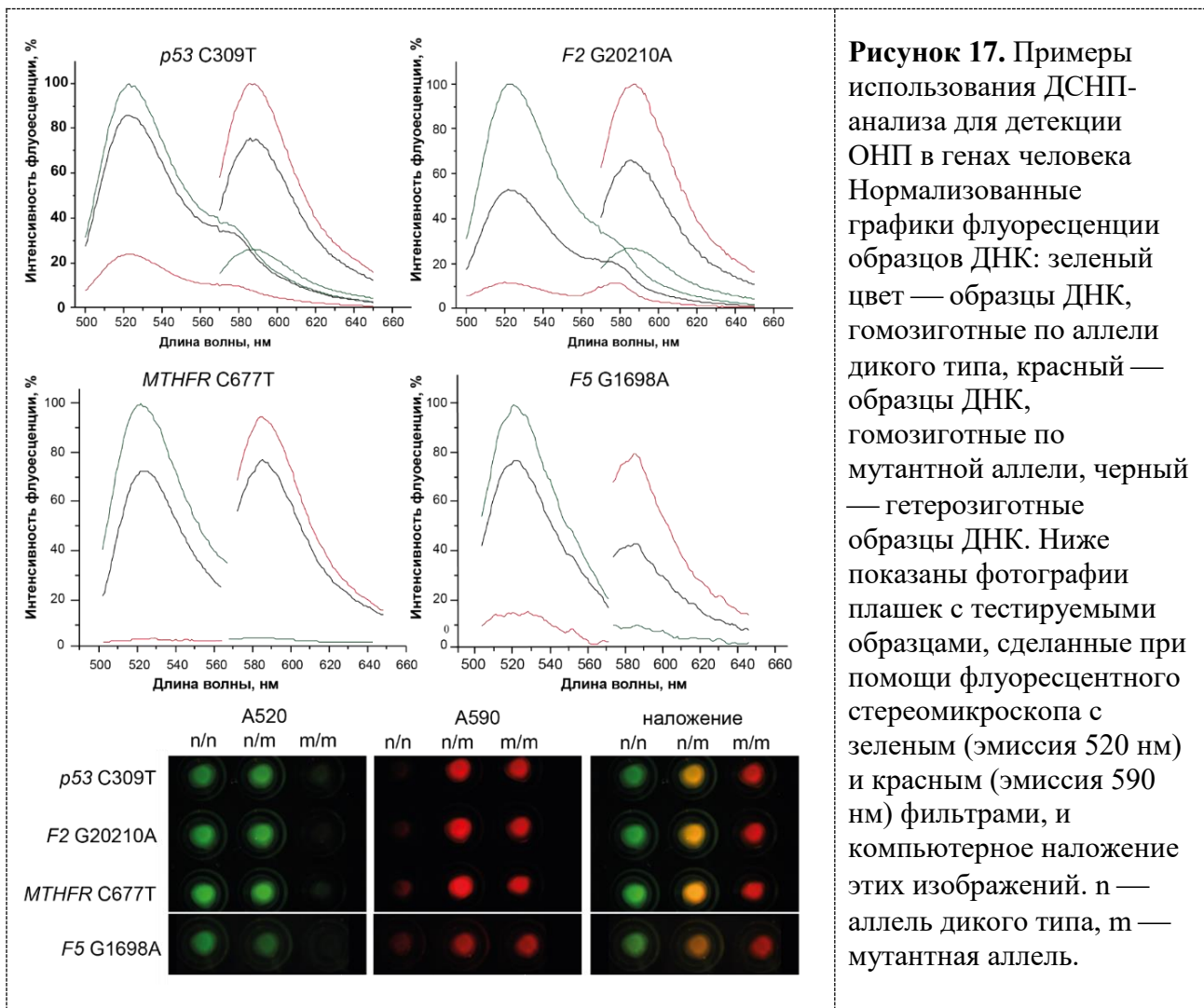
Рисунок 16. Схема выявления мутантной последовательности в биологических образцах с помощью Par_DSN. R — донор флуоресценции, Q — гаситель флуоресценции.

При добавлении ДСНП-зондов и Par_DSN в продукт ПЦР, содержащий амплифицированный целевой фрагмент с полиморфным сайтом, на первом этапе происходит расщепления ПЦР-продукта с появлением коротких фрагментов ДНК, которые способны образовывать дуплексы с зондами. Полностью комплементарные дуплексы также подвергаются атаке Par_DSN, что приводит к появлению флуоресценции определенной длины волны от соответствующего флуоресцентного донора, освобожденного от гасителя флуоресценции. Несовершенные дуплексы остаются интактными. Было проведено исследование длины фрагментов ДНК, которые могут быть использованы для анализа ОНП. На примере ДНК, содержащей ОНП С7028Т гена COX1, мы показали, что метод применим при использовании фрагментов ДНК в диапазоне до 4000 п.о.

Метод ДСНП-анализа был опробован в модельных экспериментах на образцах геномной ДНК человека с различным аллельным статусом по исследуемому ОНП,

ассоциированных с рядом заболеваний. Аллельный статус исследуемых образцов подтверждали с помощью секвенирования. В каждом случае с помощью пары ген-специфических праймеров осуществляли амплификацию фрагмента ДНК, содержащего ОНП, смешивали неочищенный продукт с двумя ДСНП-зондами, обеспечивающими однозначное выявление аллели дикого типа и/или мутантной аллели. Для зонда, комплементарного аллели дикого типа, использовали пару «флуорофор и гаситель флуоресценции» FAM-DABCYL, для зонда, комплементарного мутантной аллели, - пару TAMRA-DABCYL. Таким образом, расщепление зонда, специфического для аллели дикого типа, приводило к возникновению флуоресценции с максимумом эмиссии при 520 нм (зеленый цвет), а зонда, специфического для мутантной аллели, - с максимумом эмиссии при 590 нм (красный цвет).

Результаты экспериментов представлены на рисунке 17. Во всех случаях с помощью ДСНП-анализа удалось достоверно установить аллельный статус исследуемого образца.



На примере полиморфизма 5382insC в гене BRCA1 также была продемонстрирована применимость анализа для обнаружения однонуклеотидных делеций и инсерций (данные не приведены).

Следует отметить, что для выявления различных мутаций применяли стандартные условия реакции, что позволяет использовать метод ДСНП для широкого спектра мутаций. Так ДСНП-анализ был использован в экспериментальной работе для анализа ОНП rs165599 гена *COMT*. Этот ген кодирует катехоламин-О-метилтрансферазу – фермент, вовлеченный в определение болевой чувствительности у млекопитающих. В рамках этой работы были идентифицированы три генетических гаплотипа гена *COMT*, которые определяли низкую, среднюю и высокую болевую чувствительность. Эти гаплотипы охватывают 96% человеческой популяции, и пять комбинаций этих гаплотипов тесно связаны ($P = 0,0004$) с изменением чувствительности к экспериментальной боли.

Нами был проведен анализ ОНП rs165599 гена *COMT* в образцах геномной ДНК из периферической крови здоровых женщин-добровольцев, принявших участие в 3-летнем проспективном когортном исследовании для выявления факторов риска развития миогенного поражения височно-нижнечелюстного сустава.

ОНП rs165599 расположен в самом конце 3'-некодирующей транскрибируемой области (UTR) гена *COMT*; по литературным данным его аллель G связан с шизофренией (Shifman S. et al., 2002). Все определенные с помощью ДСНП-анализа аллельные статусы были подтверждены с помощью рестрикционного анализа.

Было показано, что ОНП rs165599 не оказывает существенного влияния на болевую чувствительность, что вместе с аналогичным результатом, полученным для ОНП, расположенным в 5'-UTR этого гена, позволило сделать вывод о низком вкладе UTR в модуляцию активности гена *COMT*, связанной с болевой чувствительностью.

В описанном выше варианте ДСНП-анализа для каждой анализируемой последовательности необходим синтез своего ДСНП-зонда, что в определенной степени удорожает проведение и усложняет дизайн эксперимента. В ряде исследований была постулирована необходимость разработки методов ОНП-анализа с универсальными зондами, позволяющими избежать подобного дорогостоящего синтеза.

Также одной из часто возникающих проблем ПЦР является образование нецелевых продуктов в ходе реакции при использовании в качестве матрицы ДНК высокой сложности (например, образцов геномной ДНК или суммарной кДНК). Содержание целевой последовательности в таких образцах часто довольно низкое, и поэтому высока вероятность получения в ходе ПЦР нецелевых продуктов, в том числе нецелевых продуктов той же длины, что и целевая амплифицируемая последовательность. Для решения этих проблем нами была предложена модификация метода ДСНП-анализа с использованием ген-специфического олигонуклеотида (ГСО) и 10-звеньевое универсальное зонда, содержащего пару «флуорофор и гаситель флуоресценции». ГСО включает последовательность, комплементарную последовательности целевого продукта, и ковалентно связанную с ней последовательность, идентичную последовательности универсального зонда. Схематически метод, получивший название ДСН-детекция, представлен на рисунке 18.

Свежий продукт ПЦР инкубируют короткое время в присутствии ГСО, зонда и Par_DSN при 60°C. Во время инкубации за счет нуклеазной активности Par_DSN происходит расщепление ДНК и гибридизация ГСО с этой фрагментированной и

частично денатурированной ДНК. ДНК-полимераза, оставшаяся в реакционной смеси, используя в качестве матрицы ГСО, образует продукт синтеза, содержащий последовательность, комплементарную универсальному зонду. На следующем этапе температуру гибридизации понижают, обеспечивая условия для гибридизации зонда с комплементарной ему последовательностью. В результате расщепления зонда в образовавшихся ДНК-ДНК дуплексах нуклеазой Par_DSN происходит высвобождение флуорофора и появляется флуоресценция специфической длины волны, которая может быть детектирована любым доступным способом. В отсутствие целевой ДНК для ГСО не находится комплементарной последовательности, в результате не появляется места для отжига зонда и флуоресцентный сигнал не формируется.

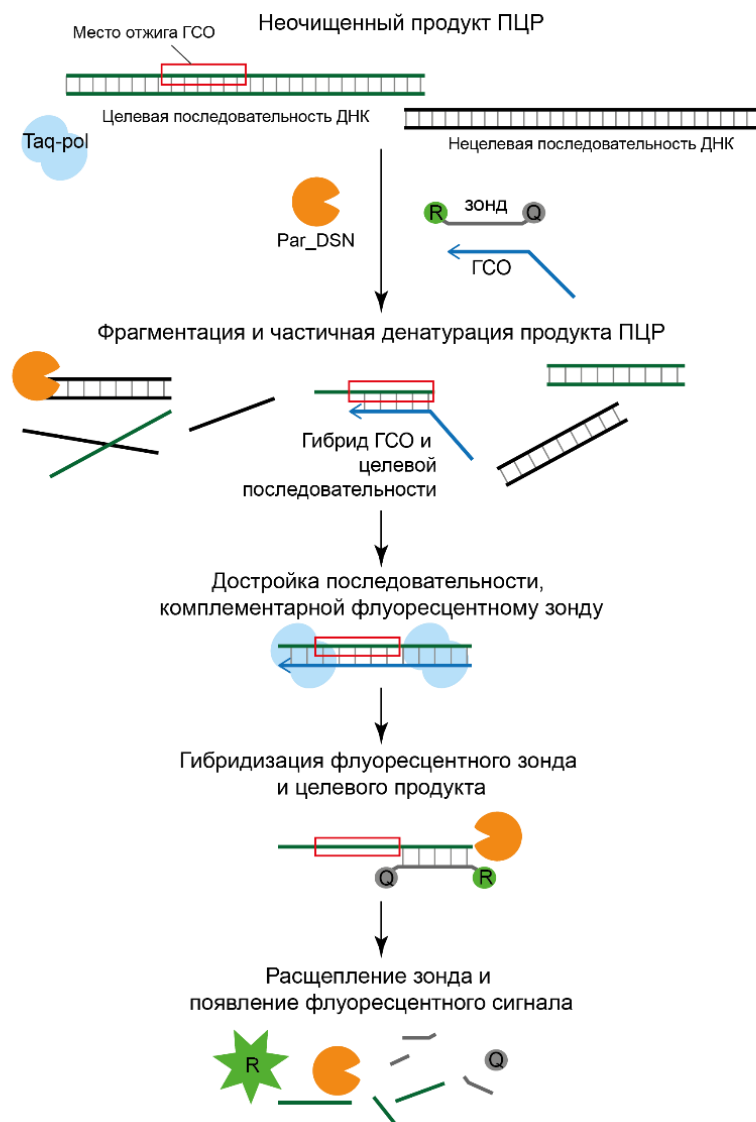


Рисунок 18. Способ выявления целевой последовательности с использованием Par_DSN. R — флуорофор, Q — гаситель флуоресценции, TaqPol — ДНК-полимераза, ГСО — ген-специфический олигонуклеотид.

Метод был протестирован в модельной системе для выявления гипервариабельных мотивов вариабельных доменов β -субъединицы Т-клеточного рецептора (ТКР). Анализ этих последовательностей востребован при исследованиях

аутоиммунных, инфекционных и онкологических заболеваний (Zmievskaia E. et al., 2021). Продукт амплификации фрагмента варибельного домена ТКР восемнадцатого семейства (ТКР18 β), полученный с использованием универсального праймера (TCR-C-Uni) и встречного праймера, специфического для последовательностей этого семейства (BV18), выглядит при электрофоретическом разделении как одна полоса. Однако при анализе библиотеки, полученной путем клонирования этого ПЦР-продукта, было обнаружено, что более 60% вставок представлены нецелевыми последовательностями. Эта библиотека была выбрана в качестве модельного объекта для ДСН-детекции. Следующим этапом работы стало тестирование клонированного продукта ПЦР, полученного при амплификации кДНК человека с использованием праймеров TCR-C-Uni и BV18. Для детекции использовали ГСО с геноспецифической последовательностью длиной 17 н.о. и флуоресцентный зонд. Результаты, полученные с помощью ДСН детекции верифицировали секвенированием. Во всех случаях секвенирование подтвердило результаты ДСН-детекции.

Метод ДСН-детекции был также использован для тестирования в модельной системе двухнуклеотидной инверсии 166_167invCT в третьем экзоне гена ДКС1 (GenBank NM_001363). Эта мутация была выявлена нами при исследовании случая врожденного дискератоza (Kurnikova M. et al., 2009). Для анализа использовали два ГСО, один включал ген-специфическую последовательность длиной 17 н.о., полностью комплементарную варианту ДКС1 дикого типа (ДКС1-Norm), второй – мутантному варианту (ДКС1-Mut). Оба ГСО содержали на 5'-конце последовательность UZ-Fam, соответствующую флуоресцентному зонду. ДНК для анализа получали путем ПЦР с плазмидной ДНК, содержащей целевые вставки. Было продемонстрировано, что ДСН-детекция позволяет уверенно выявлять последовательности ДНК, отличающиеся двухнуклеотидной инверсией.

Для исследования возможности различать последовательности, отличающиеся однонуклеотидной заменой, был проведен анализ девяти ранее охарактеризованных образцов ДНК, содержащих ОНП rs165599 гена СОМТ. В работе использовали два ГСО: СОМТ-А1, содержащий геноспецифическую последовательность длиной 16 н.о., совпадающую с последовательностью «аллеля А», и СОМТ-С, содержащий геноспецифическую последовательность длиной 15 н.о., комплементарную последовательности «аллеля G». Фрагменты, содержащие целевую последовательность ДНК, амплифицировали в ПЦР с использованием двух встречных праймеров и использовали в ДСН-детекции образцов гомо- и гетерозиготной ДНК. Было продемонстрировано, что метод позволяет уверенно различать аллельные варианты в образцах геномной ДНК.

Наконец, мы совместили технологию ДСН-детекции с аллель-специфической ПЦР (схема метода не приведена). Метод был апробирован на модели ОНП rs165599 гена СОМТ с использованием 30 ранее охарактеризованных образцов. Во всех случаях результаты ДСН-детекции совпали с полученными ранее данными.

Таким образом, на основе Par_DSN нами был разработан комплекс методов для выявления целевых последовательностей в образцах ДНК, обладающий высокой специфичностью и позволяющий различать нуклеиновые кислоты, отличающиеся всего одной нуклеотидной заменой. Высокая адаптивность предложенного

технологий и отсутствие трудоемких процедур позволяет легко адаптировать их под самые различные нужды исследователей и различный формат экспериментов.

Нормализация кДНК. Удаление дц ДНК фракции является основной проблемой методов нормализации образцов НК, предназначенных для исчерпывающего секвенирования или для приготовления библиотек. Par_DSN, обладающая крайне высокой специфичностью к дц ДНК и термостабильностью, позволила решить эту проблему. Общий принцип нормализации с использованием Par_DSN (далее ДСН-нормализация) амплифицированной кДНК схематически представлен на рисунке 19. Метод основан на использовании кинетики гибридизации НК и отличается от альтернативных технологий способом удаления дц ДНК фракции.

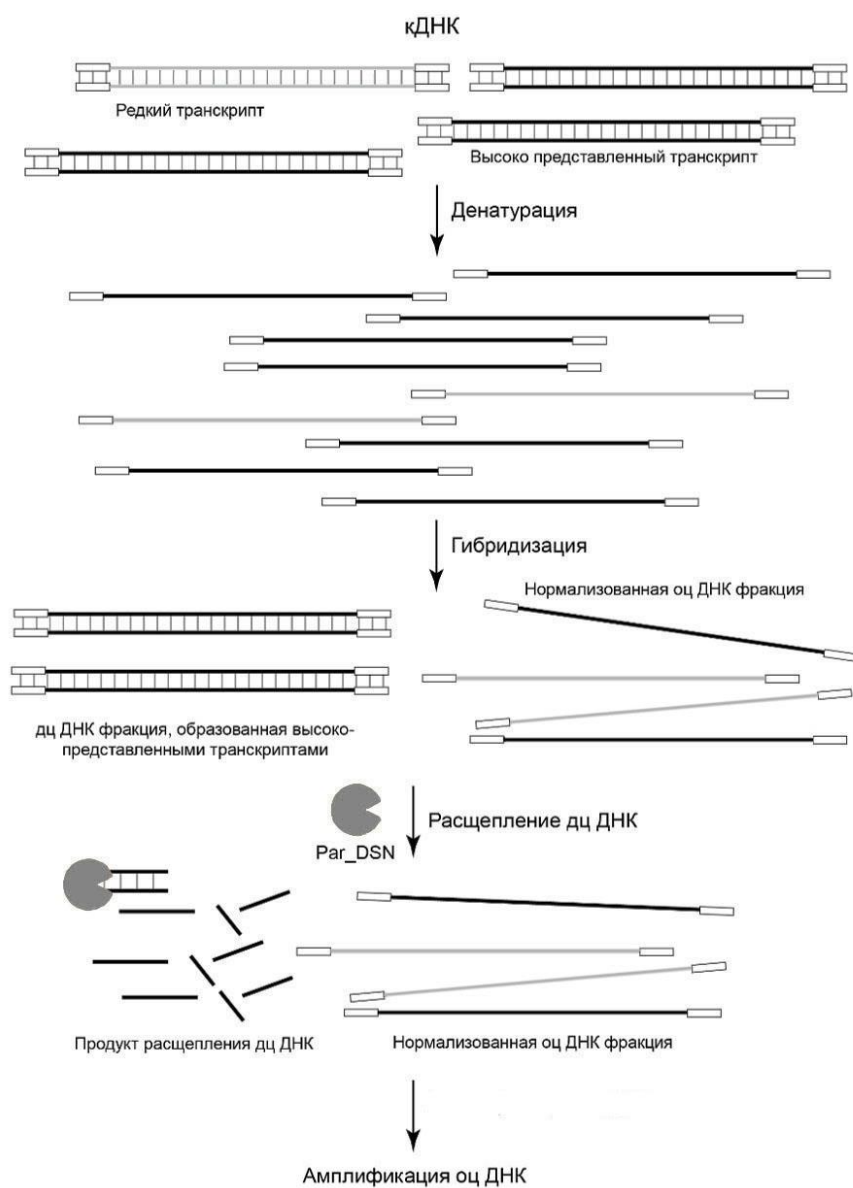


Рисунок 19. Схема нормализации кДНК с использованием Par_DSN. Прямоугольниками обозначены последовательности адаптеров. Редкий транскрипт обозначен серой линией, высокопредставленный — черной линией.

В ходе ДСН-нормализации дц ДНК денатурируют и гибридизуют в условиях, обеспечивающих реассоциацию комплементарных молекул ДНК. Во время реассоциации высокопредставленные в образце молекулы образуют дц фракцию, подлежащую удалению, а в оц фракции происходит выравнивание концентраций различных транскриптов. По окончании гибридизации для удаления нецелевой дц кДНК фракции реакционную смесь обрабатывают Par_DSN. Оставшуюся нормализованную оц фракцию кДНК далее амплифицируют с помощью ПЦР с адаптерных праймеров. Так как Par_DSN сохраняет высокую активность при температуре 70°C, удаление дц ДНК из смеси НК может осуществляться при температуре гибридизации. Благодаря этому во время нуклеазной обработки не происходит потери транскриптов, склонных к образованию вторичных структур.

Результаты разделения на гель-электрофорезе продуктов нормализации амплифицированной кДНК, полученной на основе суммарной РНК из некоторых тканей человека показаны на рисунке 20. Рядом показаны образцы ненормализованной кДНК. Во всех случаях после нормализации наблюдалось исчезновение полос, соответствующих высокопредставленным, характерным для данной ткани, транскриптам.

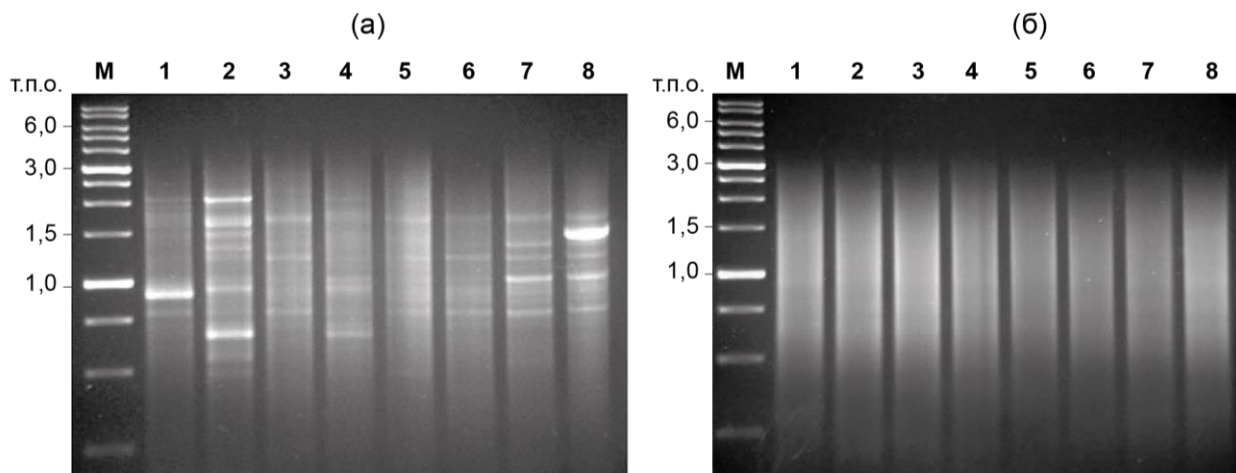


Рисунок 20. Результат нормализации образцов кДНК из тканей человека.

Результат электрофореза в агарозном геле образцов ненормализованной (а) и нормализованной (б) кДНК, приготовленной на основе суммарной РНК из плаценты (дорожка 1), эмбриональной печени (дорожка 2), мозга (дорожка 3), эмбриональных легких (дорожка 4), яичек (дорожка 5), почек (дорожка 6), тонкой кишки (дорожка 7) и желудка (дорожка 8). М — маркер длин ДНК, 1 kb ladder (Сибэнзим, Россия).

Уменьшение относительной концентрации высокопредставленных транскриптов было подтверждено с помощью виртуального Нозерн блот анализа (рисунок 21).

Изменение концентраций различных транскриптов и сохранение репрезентативности библиотек кДНК было проверено на примере нормализованного и ненормализованного образцов кДНК из эмбриональной печени человека путем их гибридизации на макрочипе с 183 ПЦР-пробами фрагментов генов человека. Результаты гибридизации показаны на рисунке 22. Было продемонстрировано выравнивание концентраций разнопредставленных транскриптов в образце нормализованной кДНК.

При гибридизации на микрочипе с 183 пробами в ненормализованной библиотеке кДНК было выявлено 130 последовательностей, комплементарных нанесенным на микрочип зондам. В нормализованной библиотеке – 151 последовательность. Концентрация для 26 транскриптов была повышена в ходе нормализации до детектируемого с помощью гибридизации уровня. При этом было обнаружено падение в ходе нормализации концентрации до недетектируемого с помощью гибридизации уровня для пяти транскриптов.

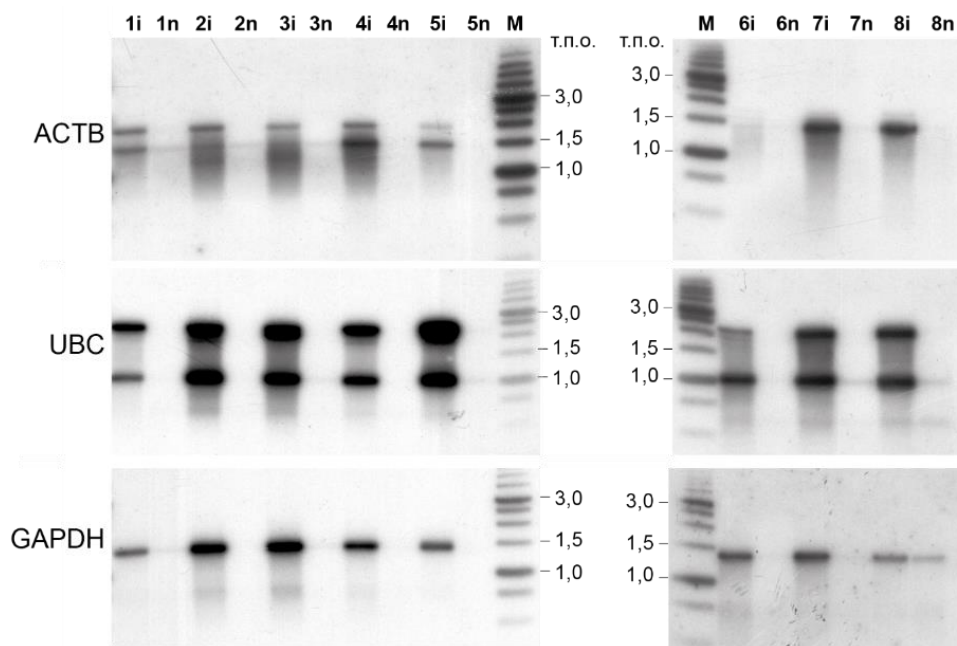


Рисунок 21. Виртуальный Нозерн блот анализ нормализованной (n) и ненормализованной (i) кДНК с P^{32} -мечеными пробами, комплементарными высокопредставленным генам. Образцы, использованные для гибридизации, описаны на рис. 21. ACTB (GenBank NM_001101), UBC (GenBank NM_021009), GAPDH (GenBank NM_002046). М — маркер длин ДНК, 1 kb ladder (Сибэнзим, Россия)

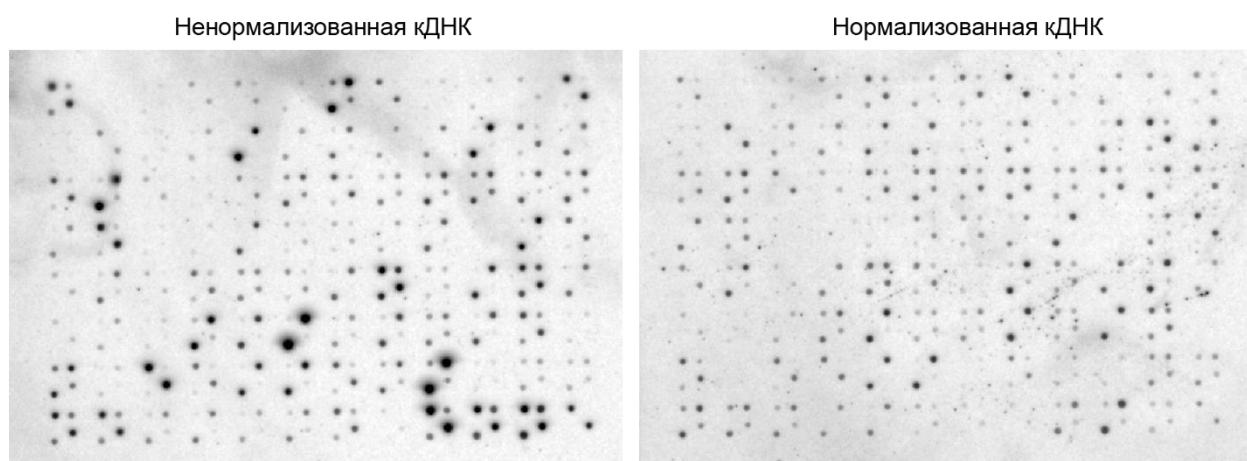


Рисунок 22. Результат гибридизации нормализованного и ненормализованного образцов кДНК из эмбриональной печени с микрочипом, содержащим 183 пробы

Аналогичный эксперимент был также проведен с использованием микрочипа, содержащего 2300 олигонуклеотидных проб (не показано). В ходе этого

эксперимента, в ненормализованной библиотеке кДНК было выявлено 388 транскрипта, в нормализованной библиотеке – 771 (почти двукратный прирост). Падение в ходе нормализации концентрации транскриптов до недетектируемого уровня было обнаружено для 51 последовательности. Пять последовательностей, уровень которых в ходе нормализации упал ниже ожидаемого, были проанализированы с помощью ПЦР. Четыре из них были обнаружены в нормализованном образце после 30-32 циклов ПЦР, одна последовательность так и не была выявлена.

Также нами была разработана модификация метода, позволяющая нормализовать первую цепь кДНК против мРНК, выступающей в качестве драйвера. Схематически эта модификация ДСН-нормализации показана на рисунке 23.

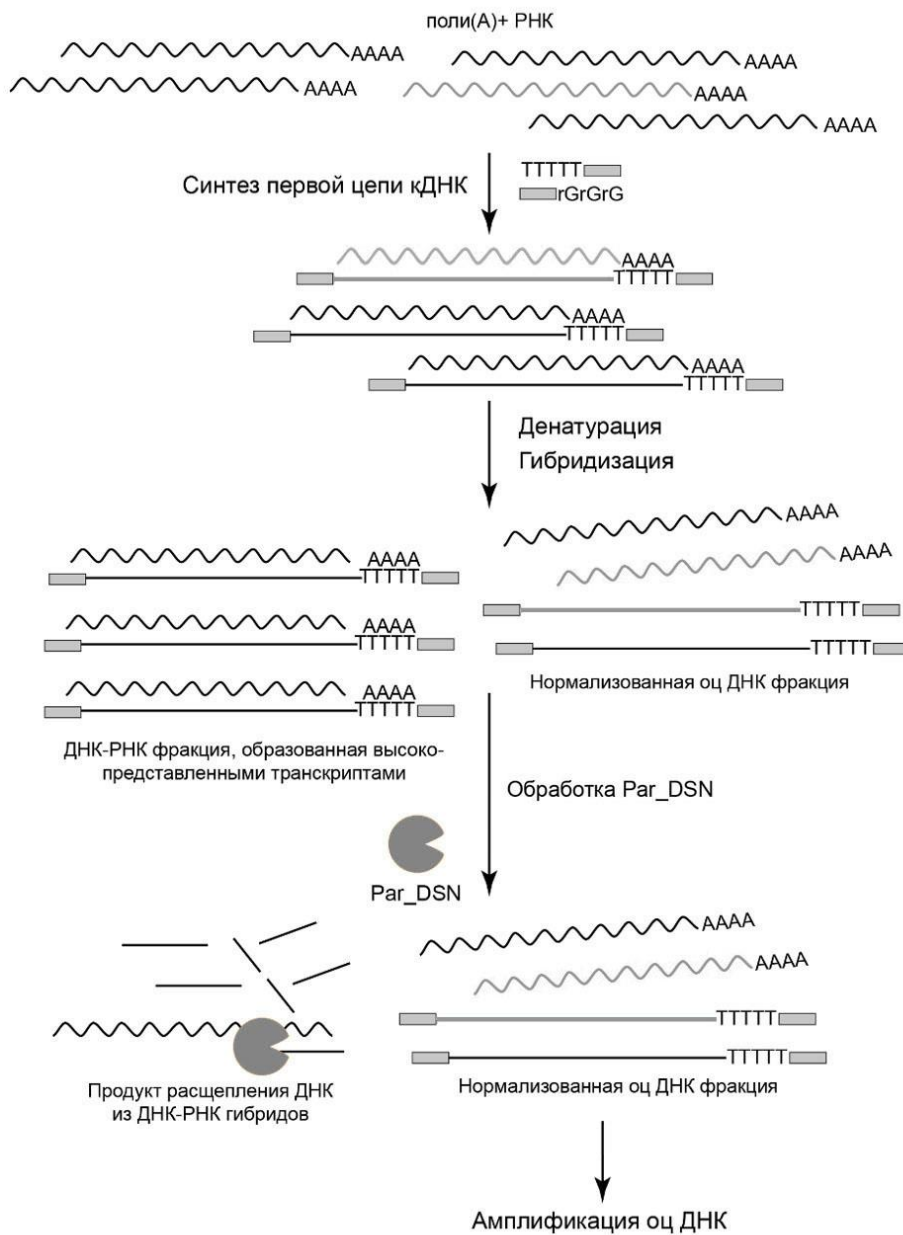


Рисунок 23. Схема ДСН-нормализации первой цепи кДНК. Прямоугольниками обозначены последовательности адаптеров. Редкий транскрипт обозначен серой линией, высокопредставленный — черной линией.

Поли(А)+ РНК используется для синтеза первой цепи кДНК, затем продукты реакции денатурируют и подвергают реассоциации. В ходе реассоциации первая цепь кДНК высокопредставленных транскриптов образует гибриды с комплементарной ей РНК, тогда как первая цепь кДНК редких транскриптов остается в оц форме.

Далее в ходе инкубации реакционной смеси с Par_DSN происходит расщепление ДНК в ДНК-РНК гибридах с сохранением нормализованной оц-фракции кДНК, которую амплифицируют с адаптерных праймеров.

Для апробации модификации метода ДСН-нормализации первой цепи кДНК была использована поли(А)+ РНК из скелетной мышцы человека. На ее основе были получены образцы амплифицированной нормализованной кДНК с использованием в качестве стартового материала для нормализации как первой цепи кДНК (образец FSN), так и амплифицированной кДНК (образец AN). Эти образцы сравнивали с образцом ненормализованной контрольной кДНК (образец AI).

На рисунке 24 показан результат электрофореза этих образцов в агарозном геле (а) и результаты виртуального Нозерн блот анализа указанных образцов с P^{32} -мечеными пробами, комплементарными транскриптам с различной представленностью (б). Согласно данным виртуального Нозерн блот анализа в ходе нормализации наблюдается увеличение концентрации редких и уменьшения концентрации высокопредставленных транскриптов.

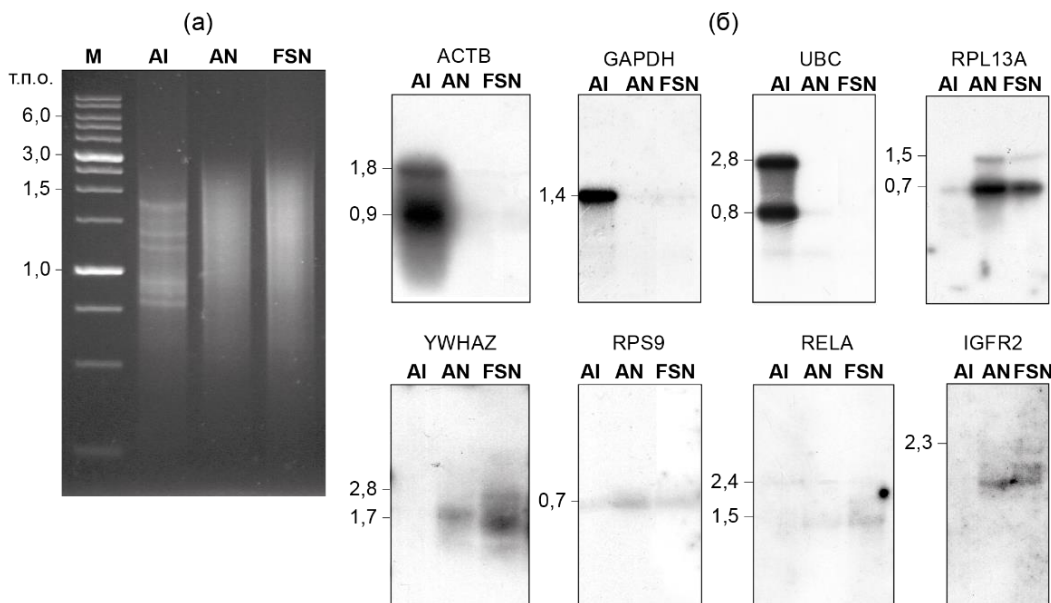


Рисунок 24. Результат нормализации образцов кДНК из скелетной мышцы человека. (а) Результат электрофореза в агарозном геле образцов ненормализованной (AI), нормализованной амплифицированной (AN) и нормализованной первой цепи (FSN) кДНК. М — маркер длин ДНК, 1 kb ladder (Сибэнзим, Россия). (б) Результат виртуального Нозерн-блот анализа этих образцов с P^{32} -мечеными пробами, комплементарными генам ACTB (GenBank NM_001101), GAPDH (GenBank NM_002046), UBC (GenBank NM_021009), YWHAZ (GenBank NM_003406), RPS9 (NM_001013), RPL13A (GenBank NM_012423), IGF2R (GenBank NM_001329128).

Образцы AI, AN и FSN были использованы для конструирования библиотек кДНК, с дальнейшей гибридизацией этих библиотек с набором радиоактивно

меченых проб к ряду генов человека. Было обнаружено заметное уменьшение числа клонов, несущих вставки с последовательностями высокопредставленных транскриптов (ACTB, GAPD, UBC) в библиотеках AN и FSN. Число клонов, содержащих вставки последовательностей генов средней представленности (RPL13A, RPS9, YWHAZ) осталось практически неизменным в AN библиотеке по сравнению с AI библиотекой и несколько снизилось в FSN библиотеке. В FSN библиотеке также наблюдалось 3-4-кратное увеличение, по сравнению с AI библиотекой, числа клонов, содержащих вставки последовательностей редких транскриптов (IGF2R, JUNC, NFkB1). В AN библиотеке концентрация этих транскриптов не изменилась.

Ранее было продемонстрировано, что если в образце ненормализованной кДНК присутствует ~10000 разных типов молекул, то при его нормализации концентрация каждого типа молекул кДНК будет стремиться к значению 0.01% (Moreno-Palanques R.F. et al., 1994). В этом случае, при исчерпывающей нормализации концентрация транскриптов со средней представленностью должна снизиться в несколько раз или остаться неизменной. Концентрация высокопредставленных транскриптов снизится на один-два порядка. Концентрация же редких транскриптов должна вырасти, но не более чем в три раза. Результаты анализа библиотек кДНК, полученных с помощью ДСН-нормализации, хорошо согласуются с расчетными данными, подтверждая эффективность разработанного нами метода. Примечательно, что в библиотеке FSN количество клонов, соответствующих высокопредставленным транскриптам, оказалось меньше, чем количество клонов, соответствующих редким транскриптам. Подобная “сверхнормализация” может объясняться тем, что при гидролизе кДНК-РНК гибридов с помощью *Par_DSN* происходит расщепление только цепи кДНК, а высвобождаемые молекулы РНК образуют новые гибриды с молекулами кДНК оц-фракции. Фракция молекул кДНК во вновь образованных ДНК/РНК гибридах также подвергается расщеплению. Исходя из полученных данных можно сделать вывод, что нормализация первой цепи кДНК является более эффективной по сравнению с нормализацией амплифицированной кДНК. Тем не менее, ДСН-нормализация амплифицированной кДНК также ведет к значительному снижению в образце высокопредставленных транскриптов, что позволяет значимо повысить эффективность поиска редких последовательностей в ситуации, когда доступна только суммарная РНК.

ДСН-нормализация была использована при исследовании аберрантно-слитых транскриптов в клеточной линии клеток рака молочной железы MCF-7. Полученные нами данные показывают, что транскрипты злокачественных опухолей человека могут содержать большое количество транскриптов слияния. А применение ДСН-нормализации сделало возможным при разумном объеме секвенирования выявление крайне редких транскриптов слияния. Их идентификация подтвердила функциональное значение геномных перестроек в онкогенезе клеток MCF-7, а также сделало возможным изучение выявленных транскриптов слияния в качестве возможных мишеней терапевтического воздействия и/или маркеров опухолевого развития.

Метод ДСН-нормализации кДНК был также использован для создания коллекции генетических маркеров, определяющих продолжительность жизни

Nothobranchius furzeri. Было доказано, что определение продолжительности жизни у *N. furzeri* является полигенным, и что обнаружение генов-кандидатов может быть осуществлено с помощью межвидового анализа.

Таким образом, нами была предложена технология нормализации кДНК, основанная на кинетике гибридизации НК и использовании уникальных свойств Par_DSN. ДСН-нормализация отличается от аналогичных методов простотой исполнения и решает задачу быстрого и эффективного выделения нормализованной оц кДНК фракции из гибридизационной смеси. ДСН-нормализация применима как образцам амплифицированной кДНК, приготовленным из суммарной РНК, так и к первой цепи кДНК, приготовленной на основе поли(А)+ РНК, и адаптирована для работы с образцами, обогащенными полноразмерными последовательностями. Благодаря простоте исполнения и доступности технологии ДСН-нормализация быстро стала самым популярным методом нормализации кДНК перед высокопроизводительным секвенированием. Разработанный нами метод был использован более чем в двух сотнях работ. В рамках этих работ было подтверждено, что технология в целом сохраняет исходное разнообразие библиотеки кДНК и позволяет эффективно идентифицировать редкие и крайне редкие транскрипты, а также улучшает эффективность секвенирования и качество данных из образцов кДНК, полученной из деградированной РНК или из малых количеств РНК.

Нормализация геномной ДНК. В связи с бурным развитием технологий высокопроизводительного секвенирования полногеномное секвенирование стало популярным подходом для получения генетической информации об организмах. Однако значительную часть геномной ДНК составляют высокомолекулярные повторяющиеся элементы. Их присутствие не только увеличивает число прочтений, необходимое для секвенирования генома, но затрудняет, а иногда делает невозможным, сборку генома в ходе биоинформатической обработки.

Мы исследовали возможность применения технологии ДСН-нормализации для удаления высокомолекулярных повторяющихся элементов из эукариотической геномной ДНК (схема не приведена).

Геномную ДНК фрагментируют и лигируют на концы адаптерные последовательности. После денатурации и последующей ренатурации фрагментированной геномной ДНК дц ДНК фракцию повторяющихся последовательностей подвергают гидролизу под действием Par_DSN. Оц фракцию геномной ДНК, обогащенную низкокопийными последовательностями, амплифицируют в ПЦР с праймером, комплементарным последовательности адаптера.

Степень желаемой нормализации может зависеть от потребностей исследовательского коллектива. В частности, модулировать ее можно, добавляя в реакционную смесь для гибридизации фракции C0t.

В эксперименте мы использовали геномную ДНК человека. Хотя содержание повторяющихся элементов в геноме человека значительно ниже, чем в геномах высших растений, геномная ДНК человека представляется более удобной для апробации метода, так как для нее известен точный состав и частота встречаемости повторяющихся элементов.

В модельном эксперименте нормализацию осуществляли с добавлением фракции C0t-1 ДНК человека, которая представляет собой последовательности, содержащие высокую концентрацию повторов LINE и SINE, и традиционно используется для блокирования неспецифического связывания богатой повторами ДНК с зондами гибридизации. При проведении гибридизации важно учитывать, что оптимальные результаты могут быть достигнуты и с другими повторяющимися фракциями ДНК от C0t 2 до C0t 10 (Trifonov V.A. et al., 2009).

Нормализованный и контрольный (ненормализованный) образцы ДНК секвенировали с использованием автоматизированной системы секвенирования 454 GS FLX (Roche). Было получено 29240 прочтений для нормализованного и 31789 для ненормализованного образцов, состоящих из 6269460 и 6643277 нуклеотидов соответственно. Анализ уровня встречаемости различных повторяющихся элементов по результатам секвенирования показал, что в результате нормализации происходит снижение процентного содержания повторяющихся элементов примерно с 40% до 25%. При этом в образце нормализованной геномной ДНК можно отметить существенное снижение уровней Alu, LINE L1P, ERV-K и ERV1 повторов, а также сателлитных последовательностей. В то же время концентрация других повторов не претерпевает существенных изменений. Анализ репрезентативности повторов с различной степенью дивергенции выявил, что в ненормализованном образце ДНК примерно 10% последовательностей имеют 100-91% идентичности; 20% – 90-71%, а остальные последовательности – менее 70% идентичности. В результате нормализации произошло 15-кратное снижение концентрации повторяющихся элементов с высокой степенью нуклеотидной идентичности (100-91%) и 2-кратное – элементов, имеющих 90-81% идентичности. Концентрация последовательностей, идентичность которых ниже 80% практически не изменилась. Иными словами, наблюдается корреляция эффективности нормализации со степенью дивергенции последовательностей семейства: чем ниже дивергенция последовательностей, тем эффективнее они нормализуются.

Для проверки сохранности в ходе нормализации уникальных последовательностей ДНК фрагменты 11 различных генов были проанализированы с помощью ПЦР в режиме реального времени в образцах геномной ДНК до и после нормализации. Для всех 11 фрагментов не было выявлено значимых изменений в уровнях представленности.

На основании полученных данных был сделан вывод, что ДСН-нормализация может быть полезной для снижения в образце геномной ДНК эволюционно молодых геномных повторяющихся элементов с низкой степенью дивергенции. Метод нашел широкое применение в целом ряде исследований при изучении растительных геномов и метагеномов бактерий.

Селективное удаление нецелевых последовательностей из образцов кДНК. Уникальные свойства Par_DSN легли в основу технологии удаления выбранных транскриптов из библиотек кДНК. Метод получил название ДСН-деплеция кДНК. Предложенная технология проста в исполнении и может быть применена как к фрагментированной кДНК, так и к кДНК, обогащенной полноразмерными последовательностями (рисунок 25).

Как и большинство известных на сегодняшний день методов вычитающей гибридизации и нормализации кДНК, ДСН-деплегция основана на кинетике гибридизации НК (Young BD et al., 1985). Глубину нормализации образца можно варьировать, меняя время гибридизации. Гидролиз дц ДНК фракции, происходящий за счет нуклеазной активности Par_DSN, является ключевым этапом ДСН-деплегции. Так же, как и в случае с ДСН-нормализацией, обработка Par_DSN происходит при высоких температурах, что позволяет избежать потери транскриптов из-за формирования вторичных структур и неспецифической гибридизации. В качестве трейсера используют образец дц кДНК, фланкированный последовательностями адаптера. Трейсер смешивают с избытком ДНК драйвера, представляющего собой фрагменты генов, последовательности которых нужно удалить из образца трейсера, денатурируют и оставляют реассоциировать. Во время гибридизации, молекулы драйвера формируют гибриды с комплементарными им последовательностями. Полученную фракцию дц кДНК гидролизуют, инкубируя с дуплекс-специфичной нуклеазой краба. Оставшуюся в смеси оц кДНК амплифицируют при помощи ПЦР. Процесс ДСН-деплегции сопровождается частичной нормализацией кДНК, поскольку в ходе гибридизации наиболее часто встречающиеся молекулы кДНК успевают перейти в двухцепочечную форму и затем подвергаются гидролизу.

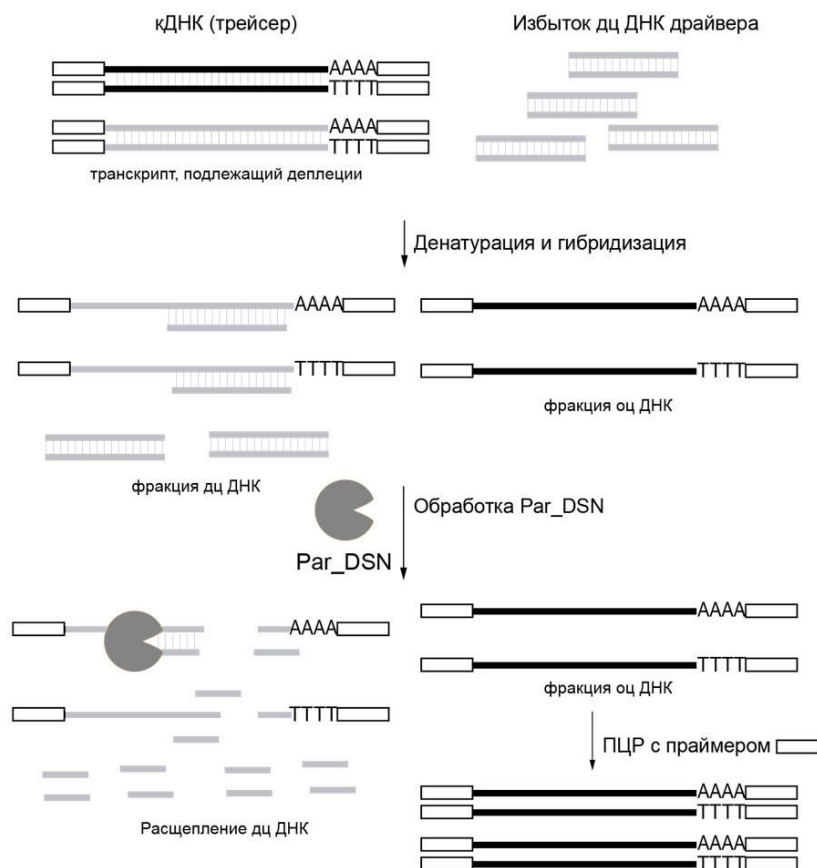


Рисунок 25. Схема ДСН-деплегции кДНК. Прямоугольниками обозначены последовательности адаптеров.

Для ДСН-деплегции может быть использована кДНК, полученная практически любым из известных методов. Единственное предъявляемое требование – наличие на

концах молекул трейсера известных адаптерных последовательностей для последующей амплификации. В модельных экспериментах в качестве драйвера нами были апробированы как амплифицированные фрагменты ДНК длиной примерно 100 п.о., так и синтетические 47- и 60-звеньевые олигонуклеотиды.

В модельном эксперименте в качестве трейсера использовали кДНК плаценты человека, из которой удаляли последовательности генов семейства фосфодиэстераз. Фосфодиэстеразы млекопитающих формируют суперсемейство, состоящее из 10 семейств, отличающихся по аминокислотным последовательностям, субстратной специфичности, чувствительности к ингибиторам, путям регуляции и тканевому распределению (Beavo J.A., 1995). С-концевой фрагмент кодирующей последовательности генов этого суперсемейства содержит высококонсервативный регион длиной ~250 аминокислотных остатков, формирующий каталитический домен (Manganiello V.C. et al., 1995). Для приготовления драйвера фрагменты, соответствующие последовательности каталитических доменов девяти генов фосфодиэстераз человека, были амплифицированы, смешаны в равных пропорциях, подвергнуты очистке на колонке, и использованы в качестве драйвера.

Драйвер (примерно 100-кратный избыток для каждого удаляемого гена) был добавлен к амплифицированной кДНК плаценты человека (трейсер). После этого образец денатурировали и гибридизовали при 68°C в течение 1 ч. Далее, гибризационную смесь инкубировали с Par_DSN для расщепления дц ДНК фракции. Оставшуюся оц кДНК амплифицировали с использованием праймеров, комплементарных адаптерным последовательностям, фланкирующим кДНК. Для оценки вклада нормализации в процесс ДСН-деплеции использовали кДНК трейсера, подвергнутую аналогичным процедурам в отсутствие драйвера.

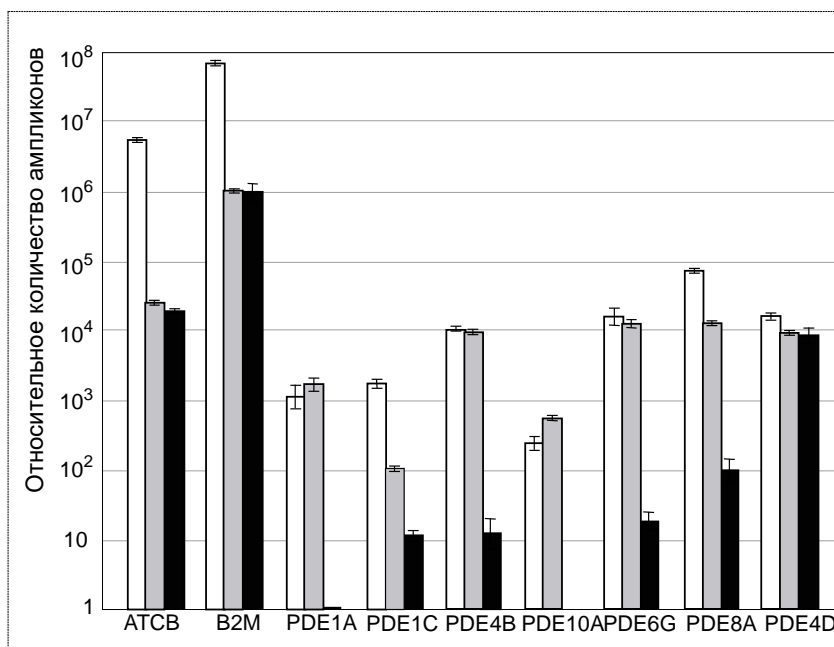


Рисунок 26. Влияние ДСН-деплеции кДНК на представленность различных транскриптов по результатам количественной ПЦР. Белые столбцы — образец С (контроль), серые столбцы — образец N (без добавления драйвера), черные — образец D (с добавлением драйвера).

Эффективность ДСН-деплеции была оценена с помощью количественной ПЦР с использованием TaqMan-проб, комплементарных 5'-последовательностям кодирующих областей генов фосфодиэстераз, лежащих выше консервативного каталитического домена. В качестве контроля использовали пробы к некоторым

высокопредставленным транскриптам и к гену фосфодиэстеразы PDE4D, который не подвергался деплеции. Результаты анализа представлены на рисунке 26.

Таким образом, ДСН-деплеция позволяет эффективно удалить из образца уже известные транскрипты, одновременно снижая концентрацию высокоповторяющихся кДНК. Это обеспечивает возможность более глубокого анализа экспрессионных библиотек кДНК методом функционального скрининга.

Удаление рибосомальной РНК перед полномасштабным секвенированием транскриптомов прокариот. Одним из важнейших применений Par_DSN стало удаление рРНК из образцов кДНК перед анализом бактериального транскриптома. При работе с прокариотами использование стандартных технологий анализа транскриптома сильно затруднено из-за невозможности применения традиционных подходов селекции мРНК. Использование олиго(dT)-праймера в качестве затравки при синтезе кДНК прокариот малоэффективно, так как мРНК бактерий по большей части неполиаденилирована. При этом кДНК, если ее синтезировать с использованием случайной затравки, более чем на 95% будет состоять из копий рРНК и тРНК (Brown T., 2018). Таким образом, актуальной задачей при разработке технологий анализа транскриптома прокариот является удаление из образцов суммарной кДНК бактерий избытка копий рРНК и тРНК при сохранении относительной представленности кодирующих последовательностей. Сходная задача стоит и при работе с образцами деградированной РНК из фиксированных тканей человека, для которых также не всегда применима селекция поли(А)+ РНК с помощью олиго(dT)-затравки (Kraus A.J. et al., 2019).

Мы разработали модификацию ДСН-деплеции для подготовки транскриптомов прокариот к полномасштабному секвенированию (далее – ДСН-деплеция рРНК) (схема не приведена). Для синтеза первой цепи кДНК на основе суммарной РНК прокариот мы адаптировали технологию SMART (Zhu Y.Y. et al. 2001), заменив 3'-адаптер, содержащий олиго(dT)-последовательность на адаптер, содержащий декамерную случайную последовательность на 3'-конце и общую с SMART Oligo 5'-часть. Первая цепь кДНК, получаемая в ходе синтеза, оказывается фланкирована последовательностью адаптера, которая используется для амплификации кДНК. После амплификации дц кДНК денатурируют и оставляют ренатурировать в гибридизационных условиях. Образовавшиеся дц ДНК молекулы удаляются из реакционной смеси с помощью обработки Par_DSN. В отличие от ДСН-нормализации, условия гибридации в данном случае подобраны таким образом, что дц ДНК фракцию образуют многокопийные молекулы рРНК и тРНК, а копии мРНК остаются одноцепочечными и сохраняют исходное распределение в образце. Эффективность предложенной процедуры истощения кДНК рРНК была подтверждена в модельном эксперименте на примере *E. coli* штамм XL1-blue. На рисунке 27 представлены результаты анализа профилей экспрессии различных генов в образцах ДСН-деплецированных и контрольных образцов кДНК. Из приведенных данных видно, что в ходе деплеции происходит обеднение образцов кДНК рибосомальной РНК при возрастании концентрация молекул мРНК. При этом профиль распределения мРНК в образце меняется незначительно. Очевидно, что увеличить глубину истощения рРНК можно дополнительно добавив в реакционную

смесь для гибридизации драйвер, содержащий фрагменты транскриптов, подлежащих деплеции, как это было описано выше.

Для дальнейшей проверки применимости технологии ДСН-деплеции рРНК к дифференциальному анализу экспрессии, было проведено сравнение с помощью супрессионной вычитающей гибридизации (Diatchenko L. et al., 1996) деплецированных транскриптомов двух близких штаммов *E. coli* штамм XL1-blue и BL 21.DE3. В качестве контроля использовали образцы кДНК, не подвергавшиеся ДСН-деплеции рРНК.

Полученные после вычитающей гибридизации образцы клонировали и анализировали полученные библиотеки методом дифференциального скрининга. Выявленные клоны секвенировали и подтверждали дифференциальный характер их экспрессии с помощью ПЦР в реальном времени. Все клоны из экспериментальных библиотек подтвердили дифференциальный характер экспрессии, тогда как клоны из контрольных библиотек оказались ложнопозитивными.

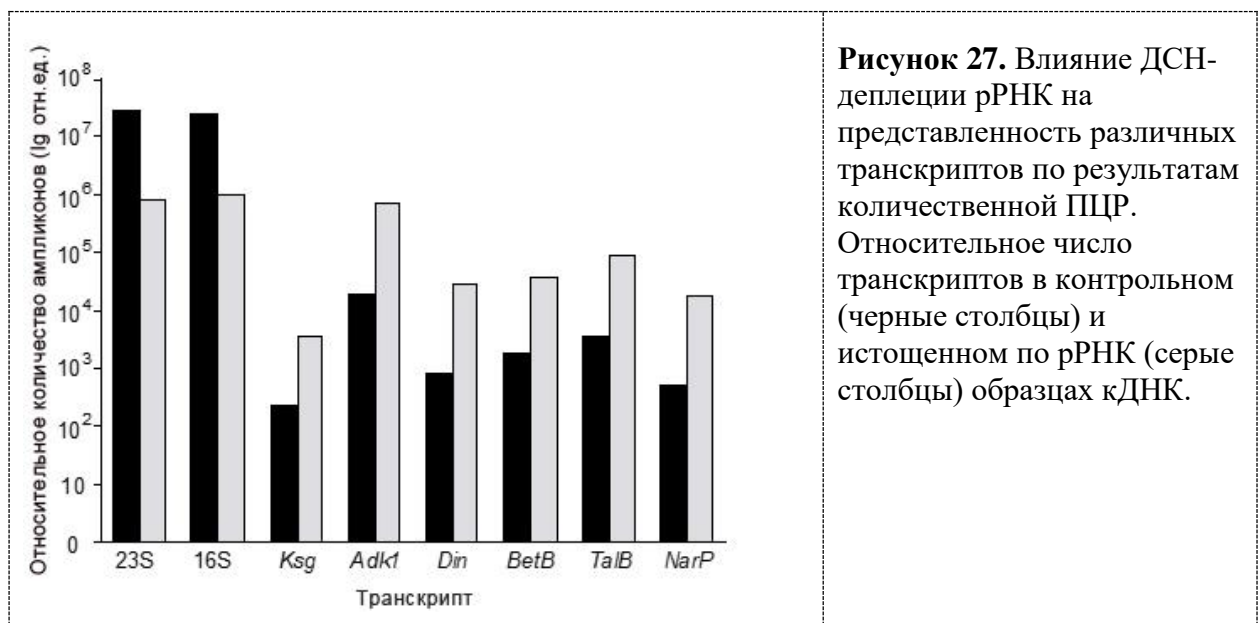


Рисунок 27. Влияние ДСН-деплеции рРНК на представленность различных транскриптов по результатам количественной ПЦР. Относительное число транскриптов в контрольном (черные столбцы) и истощенном по рРНК (серые столбцы) образцах кДНК.

Таким образом, метод ДСН-деплеции рРНК позволяет эффективно уменьшать в суммарной кДНК прокариот содержание сверхпредставленных транскриптов (копии рРНК), сохраняя относительную представленность копий мРНК. При этом деплецированные образцы кДНК могут быть использованы для количественного сравнения уровней экспрессии генов, а также для секвенирования транскриптома.

Разработанные нами метод ДСН-деплеции был апробирован во множестве лабораторий, в основном для удаления рРНК из различных образцов кДНК микроорганизмов.

ЗАКЛЮЧЕНИЕ

Значение нуклеаз для процессов, происходящих в живых организмах, сложно переоценить. Не менее важным является практическое использование нуклеаз в биотехнологии и биомедицине. Поэтому поиск и изучение нуклеаз с новыми свойствами представляет собой важную научную и практическую задачи. Несмотря

на широкий спектр нуклеаз, используемых в молекулярно-биологических исследованиях на момент начала данной работы, среди них отсутствовали клонированные термостабильные ферменты, способные избирательно гидролизовать двухцепочечную ДНК, оставляя одноцепочечную ДНК интактной. В то же время ферменты с такой активностью были необходимы для создания технологий анализа сложных смесей НК.

В рамках настоящей работы была клонирована полноразмерная кодирующая последовательность нового фермента из гепатопанкреаса камчатского краба – дуплекс-специфическая нуклеаза (Par_DSN). Анализ нуклеотидной последовательности Par_DSN выявил открытую рамку считывания длиной в 1221 п.о., кодирующую белок, состоящий из 407 аминокислот. Предсказанная аминокислотная последовательность нуклеазы краба включала в себя сигнальный пептид и нуклеазный домен. Получить активный рекомбинантный белок в гетерологической бактериальной системе экспрессии не удалось – весь продукт обнаруживался в нерастворимой фракции в тельцах включения. Поэтому была предложена и реализована методика очистки природного фермента, состоящая из комбинации нескольких последовательных стадий хроматографии. Идентичность очищенного природного белка нуклеазы из камчатского краба клонированному нами ферменту Par_DSN была подтверждена с помощью поликлональных антител (полученных к рекомбинантному белку), а также при помощи масс-спектрометрического анализа. Был проведен детальный анализ физико-химических свойств выделенного природного фермента, включая зависимость нуклеазы от катионов двухвалентных металлов, температуры, pH, мочевины, ионной силы раствора и т.д. Было показано, что нуклеаза проявляет ферментативную активность только в присутствии двухвалентных катионов металлов и ингибируется ЭДТА, то есть является металл-зависимым ферментом. Было обнаружено, что Par_DSN проявляет активность по разрушению дц ДНК в широком диапазоне температур. При использовании в качестве субстрата дц ДНК температурный оптимум наблюдался при 60°C, а 5% активности фермента сохранялось даже после его инкубации в течение 30 мин при температуре 100°C. Было продемонстрировано, что фермент обладает селективной избирательностью по отношению к дц ДНК, оставляя оц ДНК интактной. Нуклеаза расщепляла любые дц ДНК субстраты (геномную ДНК, фаговую ДНК, суперскрученную плазмидную ДНК), но не проявляла заметной активности против оц ДНК. Подсчитанное соотношение гидролитической активности фермента к дц и оц ДНК субстрату составило не менее 1000:1, что значительно выше, чем для всех известных нуклеаз. Было показано, что Par_DSN практически неактивна по отношению к РНК субстрату, но проявляет свойства металл-зависимой ДНКазы по отношению к ДНК в ДНК-РНК гибридах. Была определена минимальная длина дуплекса ДНК, разрушаемого нуклеазой краба, составившая 8 п.о. Также продемонстрирована преимущественная активность фермента к полностью комплементарным дуплексам ДНК, по сравнению с несущими в своем составе хотя бы одну нуклеотидную замену в одной из цепей ДНК.

Был проведен сравнительный анализ последовательностей Par_DSN и ее гомологов из других членистоногих (в том числе клонированных нами полноразмерных кодирующих последовательности нуклеаз еще из трех видов

ракообразных) с нуклеазами SNF. Анализ филогенетического древа выявил две большие клады. Первая включала в себя «классические» SNF нуклеазы. Вторая клада, названная нами подгруппой DSN-подобных нуклеаз, состояла из нуклеаз насекомых и ракообразных. На основании анализа филогенетического древа и данных об известной субстратной специфичности некоторых нуклеаз нами было выдвинуто предположение, что DSN-подобные нуклеазы и SNF нуклеазы эволюционировали по дивергентному пути, о чем свидетельствуют их различные биологические функции и субстратные специфичности. Анализ нуклеазной активности сконструированных укороченных форм Par_DSN, а также рекомбинантных мутантных белков, полученных сайт-направленным мутагенезом, позволил сделать вывод, что Par_DSN и ее гомологи содержат в своем составе нуклеазные домены, которые длиннее соответствующего домена нуклеаз SNF, а их каталитический центр, хотя и сходен с каталитическим центром SNF нуклеаз, имеет ряд принципиальных отличий. Кроме того, между нуклеазным доменом и сигнальным пептидом DSN-подобные нуклеазы содержат регион длиной около 90 аминокислотных остатков, важный для определения субстратной специфичности и стабильности фермента. На основании этих отличий, нами было предложено выделить DSN-подобные нуклеазы в новое семейство – Дуплекс специфических нуклеаз (ДСН) внутри суперсемейства His-Me нуклеаз. Исходя из определенного сходства членов этого семейства с нуклеазами SNF, мы предположили их дивергентное расхождение в эволюции и постулировали, что нуклеазы с дц ДНКазной активностью могут быть найдены у других членистоногих, что и было в дальнейшем подтверждено рядом работ независимых лабораторий.

Таким образом, нами впервые в мире был клонирован и охарактеризован термостабильный фермент, обладающий непревзойденной селективностью по отношению к дц ДНК – дуплекс-специфическая нуклеаза Par_DSN.

На основе уникальных свойств Par_DSN нами были разработаны новые высокоэффективные технологии: анализа ОНП в генах эукариот, выявления целевых ДНК-мишеней в комплексной смеси нуклеиновых кислот, нормализации кДНК и геномной ДНК, селективного удаления нецелевых транскриптов из популяций кДНК эукариотических организмов и удаления последовательностей рРНК при создании библиотек кДНК прокариотических организмов. Разработанные нами технологии нашли свое применение в совместных работах в рамках международного научного консорциума по исследованию хронической боли, международной программы по исследованию рака молочной железы, в рамках проекта по изучению механизмов старения, и были высоко оценены зарубежными коллегами.

Данные настоящей работы легли в основу создания ряда коммерчески доступных продуктов, что привело к широкому внедрению полученных в исследовании результатов в мировую лабораторную практику. Методы нормализации и деплеции с использованием Par_DSN стали методами выбора при подготовке биологических образцов для высокопроизводительного секвенирования. В настоящее время число статей, упоминающих дуплекс-специфическую нуклеазу камчатского краба, в базе данных Google Scholar превышает 2000.

ВЫВОДЫ

1. Открыто новое семейство нуклеаз, специфически разрушающих двухцепочечную ДНК, и получившее название дуплекс специфические нуклеазы (ДСН).
2. Проведен филогенетический анализ белков данного семейства, а также исследованы функциональные и структурные особенности ферментов данного семейства на примере дуплекс специфической нуклеазы Par_DSN из камчатского краба (*Paralithodes camtschaticus*).
3. Определены границы нуклеазного домена Par_DSN, а также проведен анализ сайтов, ответственных за каталитическую активность данного фермента.
4. На основе уникальной способности Par_DSN расщеплять полностью комплементарные короткие ДНК-ДНК дуплексы со значительно более высокой эффективностью, чем дуплексы, содержащие одну и более нуклеотидных замен, разработана группа технологий анализа мутаций в генах эукариот, а также технология выявления целевых ДНК-мишеней в комплексной смеси нуклеиновых кислот.
5. На основе уникальных особенностей Par_DSN – специфичности по отношению к дц ДНК, а также термостабильности фермента – разработана группа высокоэффективных технологий выделения целевой фракции из сложных смесей нуклеиновых кислот, включая технологию создания нормализованных библиотек кДНК из клеток и тканей эукариотических организмов, технологию создания нормализованных библиотек геномной ДНК из клеток и тканей эукариотических организмов, технологию селективного удаления нецелевых транскриптов из популяций кДНК эукариотических организмов, технологию удаления последовательностей рРНК при создании библиотек кДНК прокариотических организмов.

СПИСОК СОКРАЩЕНИЙ

| | | | |
|-------------|--|---------|---|
| ГСО | ген-специфический олигонуклеотид | рРНК | рибосомальная рибонуклеиновая кислота |
| ДНК | дезоксирибонуклеиновая кислота | т.п.о. | тысяча пар оснований |
| ДСН | дуплекс-специфическая нуклеаза | тРНК | транспортная рибонуклеиновая кислота |
| ДСНП-анализ | анализ полиморфизмов с помощью ДСН | ЭДТА | этилендиаминтетрауксусная кислота |
| дц | двухцепочечный | С0t | произведение концентрации нуклеотидов (C0) в молях на литр на время реассоциации в секундах (t) |
| кДНК | комплементарная дезоксирибонуклеиновая кислота | NGS | next generation sequencing (секвенирование нового поколения) |
| мРНК | матричная рибонуклеиновая кислота | Par_DSN | дуплекс-специфическая нуклеаза из камчатского краба <i>Paralithodes camtschaticus</i> |
| НК | нуклеиновая кислота | SDS | sodium dodecyl sulfate |
| н.о. | нуклеотидное основание | SNF | <i>Serratia</i> family of nucleases |
| ОНП | однонуклеотидный полиморфизм | | |
| оц | одноцепочечный | | |
| ПА | поликлональные антитела | | |
| п.о. | пара оснований | | |
| ПЦР | полимеразная цепная реакция | | |
| РНК | рибонуклеиновая кислота | | |

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Обзоры

1. **Д.А.Шагин**, Д.В. Ребриков. Применение дуплекс-специфической нуклеазы из камчатского краба в методах молекулярной биологии // Вестник РГМУ. 2022. Т. 1. С. 5-11.

Статьи

1. Petzold A, Reichwald K, Groth M, Taudien S, Hartmann N, Priebe S, **Shagin D**, Englert C, Platzer M. The transcript catalogue of the short-lived fish *Nothobranchius furzeri* provides insights into age-dependent changes of mRNA levels // BMC genomics. 2013. Vol. 14. P. 1-16.

2. Kirschner J, Weber D, Neuschl C, Franke A, Böttger M, Zielke L, Powalsky E, Groth M, **Shagin D**, Petzold A, Hartmann N, Englert C, Brockmann GA, Platzer M, Cellerino A, Reichwald K. Mapping of quantitative trait loci controlling lifespan in the short-lived fish *Nothobranchius furzeri*- a new vertebrate model for age research // Aging Cell. 2012. Vol. 11, № 2. P. 252-261.

3. Богданова Е.А., И. А. Шагина, Ю. Г. Янушевич, Л.Л. Вагнер, С. А. Лукьянов, **Д. А. Шагин**. Подготовка кДНК прокариот для широкомасштабного анализа транскриптома // Биоорг. химия. 2011.Т. 37, № 6. С. 854-857.

4. Шагина И.А., Е. А. Богданова, И. М. Альтшулер, С. А. Лукьянов, **Д. А. Шагин**. Использование дуплекс-специфической нуклеазы краба для быстрого анализа однонуклеотидных полиморфизмов и выявления ДНК-мишеней в комплексном продукте ПЦР // Биоорг. химия. 2011. Т. 37, № 4. С. 522-529.

5. Shagina I, Bogdanova E, Mamedov IZ, Lebedev Y, Lukyanov S, **Shagin D**. Normalization of genomic DNA using duplex-specific nuclease // Biotechniques. 2010. Vol. 48, № 6. P. 455-459.

6. Bogdanova EA, Shagina IA, Mudrik E, Ivanov I, Amon P, Vagner LL, Lukyanov SA, **Shagin DA**. DSN depletion is a simple method to remove selected transcripts from cDNA populations // Mol Biotechnol. 2009. Vol. 41, № 3. P. 247-253.

7. Anisimova VE, Shcheglov AS, Bogdanova EA, Rebrikov DV, Nekrasov AN, Barsova EV, **Shagin DA**, Lukyanov SA. Is crab duplex-specific nuclease a member of the Serratia family of non-specific nucleases? // Gene. 2008. Vol. 418, № 1-2. P. 41-48.

8. Anisimova VE, Rebrikov DV, **Shagin DA**, Kozhemyako VB, Menzorova NI, Staroverov DB, Ziganshin R, Vagner LL, Rasskazov VA, Lukyanov SA, Shcheglov AS. Isolation, characterization and molecular cloning of Duplex-Specific Nuclease from the hepatopancreas of the Kamchatka crab // BMC biochemistry. 2008. Т. 9, №. 1. С. 1-12.

9. Bogdanova EA, **Shagin DA**, Lukyanov SA. Normalization of full-length enriched cDNA // Mol Biosyst. 2008. Vol. 4, № 3. P. 205-212.

10. Volik S, Raphael BJ, Huang G, Stratton MR, Bignel G, Murnane J, Brebner JH, Bajsarowicz K, Paris PL, Tao Q, Kowbel D, Lapuk A, **Shagin DA**, Shagina IA, Gray JW, Cheng JF, de Jong PJ, Pevzner P, Collins C. Decoding the fine-scale structure of a breast cancer genome and transcriptome // Genome Res. 2006. Vol. 16, № 3. P. 394-404.

11. Diatchenko L, Slade GD, Nackley AG, Bhalang K, Sigurdsson A, Belfer I, Goldman D, Xu K, Shabalina SA, **Shagin D**, Max MB, Makarov SS, Maixner W. Genetic basis for

individual variations in pain perception and the development of a chronic pain condition // Hum Mol Genet. 2005. Vol. 14, № 1. P. 135-143.

12. И. М. Альтшулер, П. А. Жулидов, Е. А. Богданова, Н.Н. Мудрик, **Д. А. Шагин**. Использование метода DSNP для изучения точечных мутаций генов человека // Биоорг. химия. 2005. Т. 31, № 6. С. 627-636.

13. Жулидов П.А., Е. А. Богданова, А. С. Щеглов, И. А. Шагина, Л. Л. Вагнер, Г. Л. Хазпеков, В. В. Кожемяко, С. А. Лукьянов, **Д. А. Шагин**. Метод создания нормализованных библиотек кДНК, обогащенных полноразмерными последовательностями // Биоорг. химия. 2005. Т. 31, № 2. С. 186-194.

14. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA, **Shagin DA**. Simple cDNA normalization using kamchatka crab duplex-specific nuclease // Nucleic Acids Res. 2004. Vol. 32, № 3. P. e37.

15. **Shagin DA**, Rebrikov DV, Kozhemyako VB, Altshuler IM, Shcheglov AS, Zhulidov PA, Bogdanova EA, Staroverov DB, Rasskazov VA, Lukyanov S. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas // Genome Res. 2002. Vol. 12, № 12. P. 1935-1942.

Патенты

1. Патент США № 7,435,794 14.10.2008. Lukyanov S., Rebrikov D., **Shagin D**. Methods and compositions for selectively cleaving DNA containing duplex nucleic acids in a complex nucleic acid mixture, and nuclease compositions for use in practicing the same.

2. Патент США № 7,803,922 28.09.2010, Lukyanov S., Rebrikov D., **Shagin D**. Methods and compositions for selectively cleaving DNA containing duplex nucleic acids in a complex nucleic acid mixture, and nuclease compositions for use in practicing the same.

Монографии, главы в книгах

1. Bogdanova EA, Barsova EV, Shagina IA, Scheglov A, Anisimova V, Vagner LL, Lukyanov SA, **Shagin DA**. Normalization of full-length-enriched cDNA // cDNA Libraries: Methods and Applications. 2011. P. 85-98.

2. Bogdanova EA, Shagina I, Barsova EV, Kelmanson I, **Shagin DA**, Lukyanov SA. Normalizing cDNA libraries // Current protocols in molecular biology. 2010. Vol. 90, № 1. P. 5.12. 1-5.12. 27.

3. Shcheglov A, Zhulidov P, Bogdanova E, **Shagin D**. Generation of normalized cDNA libraries // Nucleic Acids Hybridization Modern Applications. 2007. P. 97-124.

Тезисы докладов

1. Подготовка кДНК прокариот для широкомасштабного анализа транскриптома. **Шагин Д.А.**, Шагина И.А., Янушевич Ю.Г., Богданова Е.А., Молекулярная диагностика, 18-20 апреля 2017 г., Москва, Молекулярная диагностика 2017; Т2 – С.451

2. Метод нормализации геномной ДНК с использованием дуплекс-специфической нуклеазы камчатского краба. **Шагин Д.А.**, Шагина И.А., II Всероссийская конференция с международным участием «Высокопроизводительное секвенирование в геномике», 18-23 июня 2017 г Новосибирск, ActaNaturae спецвыпуск №1 2017 Т9 – С. 23