

ПРОТОКОЛЫ ЭКСПЕРИМЕНТОВ, ПОЛЕЗНЫЕ МОДЕЛИ, ПРОГРАММЫ И СЕРВИСЫ

АЛГОРИТМ ОПРЕДЕЛЕНИЯ ПОСЛЕДОВАТЕЛЬНОСТИ БЕЛКА С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИИ МЕТОДОВ ДЕГРАДАЦИИ ПО ЭДМАНУ И ПАНОРАМНОГО МАСС-СПЕКТРОМЕТРИЧЕСКОГО АНАЛИЗА

В.С. Скворцов*, А.В. Микурова, Н.Э. Вавилов, В.Г. Згода

Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича, 119121, Москва, Погодинская ул. 10; *эл. почта: vladlen@ibmh.msk.su

В работе рассмотрен алгоритм, позволяющий объединить преимущества таких методов как деградация по Эдману и масс-спектрометрическое *de novo* секвенирование. В качестве тестового белка в работе использовали белок-реагент для кожного теста “Диаскинтест”. Белок подвергали расщеплению трипсином и для смеси пептидов последовательно проводили 5 шагов деградации по Эдману. На каждом этапе полученную смесь подвергали протеомному панорамному анализу. Результаты анализировали как известными программами *de novo* секвенирования Novog и PepNovo+, так и собственной программой, кластеризующей отдельные спектры по С-концевой сигнатуре, образованной Y-ионами. Показано, что использование данного подхода позволяет уверенно определить аминокислотную последовательность N-концевой части пептидов, полученных при гидролизе белка.

Ключевые слова: деградация по Эдману; панорамный масс-спектрометрический анализ; *de novo* секвенирование

DOI: 10.18097/BMCRM00087

ВВЕДЕНИЕ

Несмотря на стремительный рост числа расшифрованных геномов, задача расшифровки белковой последовательности (или секвенирование) непосредственно по имеющемуся белку не потеряла своей актуальности. Это особенно важно при определении аминокислотной последовательности переменных фрагментов антител, а также белков организмов с неизвестным геномом, последовательности различных биологически активных пептидов и т.д.

Один из старейших методов, используемых для этого – деградация по Эдману [1]. В ходе секвенирования проводят пошаговое (по одной аминокислоте) химическое отщепление меченого присоединением фенилизотиоцианата (ФИТЦ) N-концевого аминокислотного остатка (рис. 1) и идентифицируют его при помощи хроматографии. Основными проблемами в этом методе являются невысокая скорость анализа и нарастающая вероятность ошибок при проведении большого числа итераций, так как, несмотря на высокий выход реакции, не все N-концевые остатки модифицируются, и постепенно достоверность определения отщепленного остатка падает. Другим способом расшифровки первичной структуры белков и пептидов

является масс-спектрометрический анализ пептидов, полученных при расщеплении белка с последующей идентификацией пептидов масс-спектров процедурой *de novo* секвенирования. Существует достаточное число алгоритмов и программ, позволяющих провести подобный анализ [2]. Однако, по понятным причинам результаты работы таких программ очень сильно зависят от качества спектров и часто не дают однозначного результата. Кроме того, при учёте всех возможных масс-спектрометрических артефактов время счёта критически возрастает, а точность идентификации существенно падает.

В настоящей работе рассмотрен алгоритм, позволяющий объединить преимущества таких методов как деградация по Эдману и масс-спектрометрическое *de novo* секвенирование.

МАТЕРИАЛЫ И МЕТОДЫ

В качестве тестового белка в работе использовали белок-реагент для кожного теста “Диаскинтест” (DSTP) [3], применяемый в диагностике туберкулёза. Это рекомбинантный белок размером 209 аминокислотных остатков. Экспериментальную часть выполняли по следующей схеме:

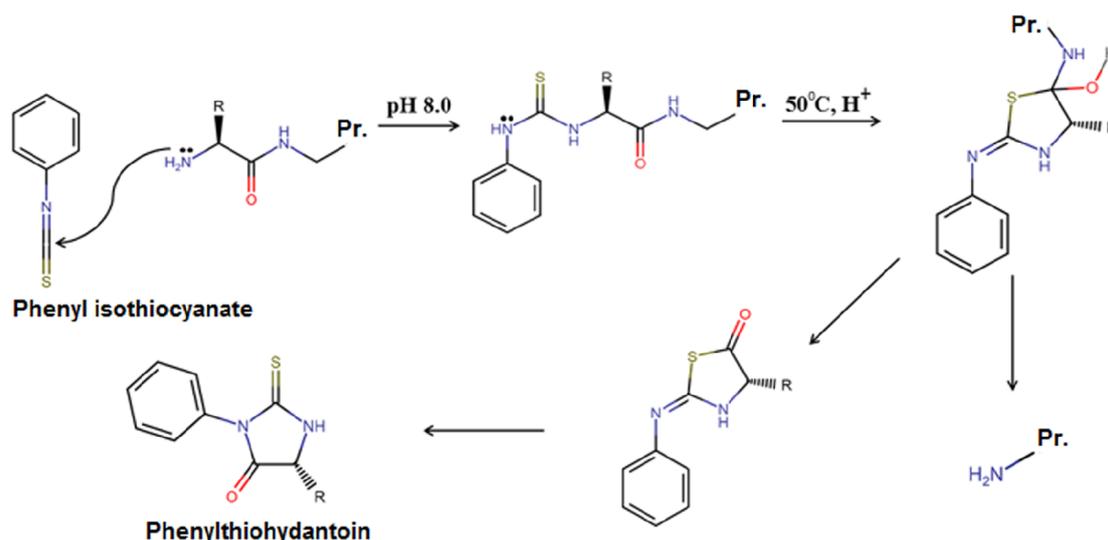


Рисунок 1. Схема реакции отщепления N-концевого аминокислотного остатка по Эдману.



белок предварительно подвергали расщеплению трипсином (метод описан ранее [4]), затем последовательно проводили 5 шагов деградации по Эдману. Исходную смесь пептидов и смесь, полученную после каждого из шагов деградации по Эдману, подвергали протеомному панорамному анализу.

Пробу пептидов, полученных при ферментативном гидролизе белка трипсином, предварительно выпаривали в вакуумном концентраторе Concentrator Plus ("Eppendorf", Германия). Реакцию присоединения ФИТЦ к N-концевой α -аминогруппе пептида проводили в буфере, содержащем пиридин:ФИТЦ:ddH₂O, pH 8.0 в соотношении 2.5:0.5:1. Реакцию проводили при температуре 50°C в течение 10 мин. По окончании реакции проводили отмывку от побочных продуктов реакции раствором, содержащим гексан:этилацетат в соотношении 10:1. Для проведения реакции кислотного гидролиза по первой пептидной связи использовали трифторуксусную кислоту, реакцию проводили при температуре 50°C в течение 3 мин. По окончании реакции отбирали аликвоту для LC-MS/MS анализа. На этом этапе цикл начинали сначала.

Перед аналитическим разделением пептиды наносили на обогащающую колонку Accalaim Precolumn (размер частиц 0.5 мм × 3 мм, 5 мкм) ("Thermo Scientific", США) при скорости потока 10 мкл/мин в течение 5 мин в изократическом режиме подвижной фазы Б (2% ацетонитрил, 0.1% муравьиная кислота).

Протеомный панорамный анализ пептидов осуществляли с использованием хроматографической ВЭЖХ системы Ultimate 3000 RSLCnano ("Thermo Scientific"). Пептиды разделяли на ВЭЖХ колонке Acclaim Pepmap® C18 (размер частиц 75 мкм × 150 мм, 2 мкм) ("Thermo Scientific") в градиентном режиме элюирования. Градиент формировали подвижной фазой А (0.1% муравьиная кислота) и подвижной фазой Б: (80% ацетонитрил, 0.1% водный раствор муравьиной кислоты) при скорости потока 0.3 мкл/мин. Колонку промывали 2% подвижной фазой Б в течение 10 мин, после чего линейно увеличивали концентрацию подвижной фазы Б до 37% за 45 мин, затем линейно увеличивали концентрацию фазы Б до 90%, после этого аналитическую колонку промывали для уравнивания 2% фазой Б в течение 10 мин. Общая длительность анализа составляла 65 мин.

Масс-спектрометрический анализ проводили на гибридном орбитальном масс-спектрометре Orbitrap Q-exactive ("Thermo Scientific") в режиме положительной ионизации в источнике NESI ("Thermo Scientific"). Напряжение на эмиттере 2.1 кВ. Панорамное сканирование проводили в диапазоне масс от 400 m/z до 2000 m/z , тандемное сканирование фрагментных ионов от нижней границы 110 m/z до верхней границы, определяемой зарядным состоянием прекурсорного иона. Для тандемного сканирования учитывали только ионы от $z = 2+$ до $z = 6+$ по зарядному состоянию. Максимальное число разрешённых для изоляции ионов в режиме MS2 было установлено не более 20. Максимальное время накопления для прекурсорных ионов составило 50 мс, для фрагментных ионов – 110 мс.

Полученные после панорамного сканирования MS/MS спектры анализировали как собственным программным обеспечением, описанным ниже, так и двумя программами *de novo* секвенирования Novog [5] и PepNovo+ [6]. Для всех программ использовали сходный набор параметров: точность определения массы первичного иона 5 ppm, точность определения массы фрагментов 0.02 Да; в качестве возможных модификаций рассматривали карбамидметилирование цистеина, дезаминирование глутамина и аспарагина, а также окисление метионина. Программы обрабатывали все 6 проб одновременно.

Для анализа покрытия использовали программу ProteoCat [7].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Первым этапом работы была попытка использовать уже существующие программы *de novo* секвенирования. На рисунке 2 представлено покрытие последовательности белка, полученное в результате работы программ Novog и PepNovo+. Общий процент покрытия составил 24% для программы Novog и 39% для PepNovo+. Однако, в отличие от PepNovo+ программа Novog всегда пытается дать единственное, но полноценное решение, тогда как PepNovo+ может выдавать достоверную часть последовательности (TAG) и остаточные массы (с N и/или C концов). Если рассмотреть только фрагменты пептидов, предсказанных программой Novog, для которых поаминокислотный счёт больше 50 при минимальной длине фрагмента в 6 остатков, степень покрытия возрастает до 43%. Тем не менее, даже при таком допущении часть возможных пептидов, наличие которых ожидаемо после трипсинолиза, не идентифицируется процедурой *de novo* секвенирования.

Если более подробно рассмотреть предсказанные пептиды, то иногда видны следы того, что деградация по Эдману проводилась. Наиболее яркий пример, полученный программой PepNovo+, представлен в таблице 1 на пептиде QKQELDELSTNIR (напомним, что пару I/L программы *de novo* секвенирования не различают). Видно, что N-концевой фрагмент QKQE легко восстанавливается. Однако, та же таблица демонстрирует и имеющуюся проблему. Пептиды 4, 5, 6 и 14 заведомо не могут пройти барьер в 5 ppm, определённый как точность измерения массы первичного иона. Пептид 11 также не пройдёт этот барьер, но данный случай, скорее всего, действительно связан либо с ошибкой измерения, либо с особенностями работы программы PepNovo+, в то время как остальные, вероятнее всего, обусловлены наличием масс-спектрометрических артефактов, либо неучтённых модификаций. Например, в исследуемой выборке можно предположить наличие следующих модификаций помимо 4-х, указанных выше: формилирование, цитрулирование аргинина, аддукт формальдегида, ацетилирование. Иногда модификации накладываются друг на друга, и предположить, что именно вызывает изменение массы пептида, затруднительно.

Последующий алгоритм определения аминокислотной последовательности белка включает следующие процедуры. Для сравнения спектров между собой и объединения их в кластеры для каждого из спектров вычисляют сигнатуру из пиков, которые могут соответствовать Y-ионам (масса пика равна сумме масс комбинации из 20 аминокислотных остатков с учётом возможных модификаций плюс 19.0184). Затем рассматривают все последовательности таких пиков таким образом, чтобы расстояние между ними соответствовало массе одного остатка, либо их комбинации (в текущем варианте комбинация не могла иметь массы более 300), при условии, что отклонения укладываются в заданную точность (в данной работе 0.02 Да). В качестве сигнатуры отбирают последовательность пиков с минимальным RMSD от "идеального" и максимальной длины. Сравнение сигнатуры идёт от пика наименьшей массы в сторону увеличения. Пики считаются совпадающими, если их массы различаются в пределах установленной точности измерения. При этом допускается пропуск 1-2 пиков, но не больше чем на дельту масс в 300 Да. Совпадающими считают сигнатуры, у которых число совпавших пиков более установленного порога (в данной работе 7). Сигнатура может быть представлена в виде аминокислотной последовательности, но, если промежуток между пиками соответствует больше чем одному аминокислотному остатку, критериев выбора их очередности в программе нет.

Данная процедура позволяет отнести к конкретному кластеру значительно большее число спектров, чем полноценная процедура *de novo* секвенирования. Так, если

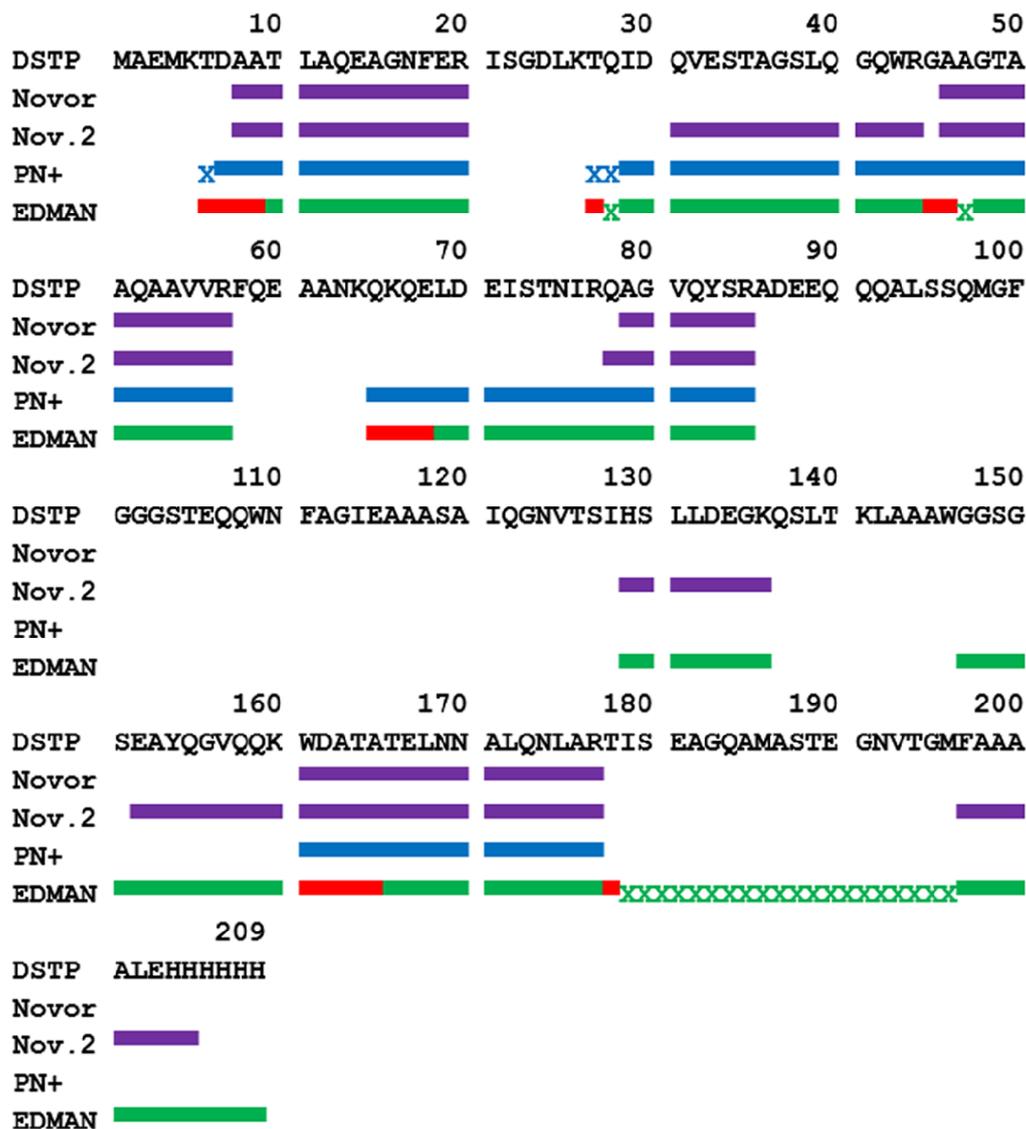


Рисунок 2. Покрытие последовательности DSTP пептидами, идентифицированным программами *de novo* секвенирования Novor и PepNovo+ (PN+). Nov.2 – используется только часть пептида, предсказанного Novor, с оценкой по-аминокислотной достоверности >50 (минимальная длина 6 остатков). У PepNovo+ значком “X” показана совпадающая остаточная масса при правильно идентифицированном TAG. EDMAN: красным цветом отмечен N-концевой фрагмент, рассчитанный по разности масс пептидов; зелёным – C-концевой; “X” – фрагмент для которого известна масса.

Таблица 1. Идентификация программой PepNovo+ пептида QKQELDELSTNIR и пептидов, полученных в результате его деградации по Эдману. Красным цветом выделены остатки, идентифицируемые по разнице масс пептидов (последовательность QKQE с N-конца пептида). Числа – остаточная масса за вычетом массы известных аминокислотных остатков до данной позиции.

Peptid	MH+ (Da)	res. 1	res. 2	res. 3	res. 4	res. 5	res. 6	res. 7	res. 8	res. 9	res. 10	res. 11	res. 12	res. 13
P1	1626.73	52.91	180.97	309.07	E	L	D	E	L	S	T	N	L	R
P2	1626.73	52.91	180.97	309.07	437.13	L	D	E	L	S	T	N	L	R
P3	1626.73	52.91	180.97	309.06	437.12	566.17	D	E	L	S	T	N	L	R
P4	1573.82	Q	K	Q	E	L	D	E	L	S	T	N	L	R
P5	1573.81	-0.003	128.06	Q	E	L	D	E	L	S	402.25	288.21	175.12	19.02
P6	1573.82	-0.003	128.06	Q	E	L	D	E	L	S	T	N	L	R
P7	1445.76		K	Q	E	L	D	E	L	503.29	402.24	288.2	175.12	19.01
P8	1445.76		K	Q	E	L	D	E	L	S	402.24	288.2	175.12	19.01
P9	1445.77		K	Q	E	L	D	E	L	S	T	288.21	175.12	19.02
P10	1445.76		K	Q	E	L	D	E	L	S	T	N	L	R
P11	1317.62			Q	E	L	D	E	L	S	401.79	287.75	174.66	18.56
P12	1189.61				E	L	D	E	L	S	T	288.21	175.12	19.02
P13	1060.56					L	D	E	L	S	T	N	175.12	19.02
P14	1047.6					-12.962	102.07	E	L	S	T	N	L	R

Кроме того, некоторые небольшие пептиды были полностью идентифицированы только за счёт Y-ионов. Также видно, что идентифицированные фрагменты пептидов 9 и 11 не триптические. Вероятно, это связано с наличием побочной неспецифичной активности у фермента [8]. Следует помнить, что задачи достоверной и доказательной идентификации C-концевого фрагмента или полного пептида перед нашей программой не стояло. В первую очередь она призвана выявлять последовательность N-концевого фрагмента с помощью метода деградации по Эдману. Но даже несмотря на то, что процедура идентификации C-концевого фрагмента может быть существенно улучшена, покрытие DSTP увеличивается до 57.5% (рис. 2) без учёта фрагмента 179-206, для которого вычислена масса, но сам он не идентифицирован.

ЗАКЛЮЧЕНИЕ

С помощью предлагаемого алгоритма метод деградации по Эдману можно с успехом применять для разрешения последовательности неизвестного белка или идентификации смеси пептидов, особенно, если комбинировать его с другими методами *de novo* секвенирования. Из приведённых данных видно, что неразрешёнными остаются очень короткие и очень протяжённые фрагменты белковой цепи. Если первый вариант имеет смысл решать с использованием протеазы со специфичностью отличной от трипсина, то второй – путём увеличения числа шагов деградации по Эдману, так как пептид при этом будет укорачиваться, и вероятность его детекции возрастёт. В отличие от классического секвенирования методом деградации по Эдману, то, что по мере увеличения числа шагов увеличивается число непрореагировавшего белка, и идентифицировать уходящий аминокислотный остаток становится затруднительно, в данном варианте не принципиально. Кроме того, на каждом последующем шаге количество всё более и более редуцированного пептида будет уменьшаться, но при достаточно большом количестве исходного белка это также не должно быть проблемой, пока оставшаяся длина пептида достаточна для идентификации C-концевого фрагмента.

БЛАГОДАРНОСТИ

Работа выполнена в рамках Программы фундаментальных научных исследований государственных академий наук на 2013-2020 годы. Масс-спектрометрические данные были получены в ИБМХ на оборудовании ЦКП “Протеом человека”, поддержанного Минобрнауки России в рамках выполнения соглашения №14.621.21.0017 (уникальный идентификатор работ RFMEFI62117X0017).

ЛИТЕРАТУРА

1. Edman, P. (1950) Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand*, 4(7), 283-293. DOI: 10.3891/acta.chem.scand.04-0283
2. Vyatkina, K.V. (2018) *De novo* секвенирование белков и пептидов: алгоритмы, приложения, перспективы. *Biomedical Chemistry: Research and Methods*, 1(1), e00005. DOI: 10.18097/BMCRM00005
3. Kiselev, V.I., Baranovskii, P.M., Rudykh, I.V., Shuster, A.M., Mart'ianov, V.A., Mednikov, B.L., ... & Slogotskaia, L.V. (2009). Clinical trials of the new skin test Diaskintest for the diagnosis of tuberculosis. *Problemy tuberkuleza i boleznei legkikh*, (2), 11-16.
4. Mikurova A.V., Novikova S.E., Skvortsov V.S., Alekseychuk N.N., Rybina A.V., Miroshnichenko Yu.V. (2017) The sequence coverage in different methods of mass spectrometry data analysis obtained on model proteins, *Biomeditsinskaya Khimiya*, 63(5), 397-404. DOI: 10.18097/PBMC20176305397
5. Ma, B. (2015) Novor: real-time peptide *de novo* sequencing software. *J. Am. Soc. Mass Spectrom.*, 26(11), 1885-1894. DOI: 10.1007/s13361-015-1204-0
6. Frank, A., & Pevzner, P. (2005) PepNovo: *de novo* peptide sequencing via probabilistic network modeling. *Anal. Chem.*, 77(4), 964-973. DOI: 10.1021/ac048788h
7. Skvortsov, V.S., Alekseychuk, N.N., Miroshnichenko, Y.V., & Rybina, A.V. (2018) pIPredict Version 2: New Features and PTM Analysis. *Biomedical Chemistry: Research and Methods*, 1(2), e00009. DOI: 10.18097/BMCRM00009
8. Keil, B. (2012) Specificity of proteolysis. Springer Science & Business Media.

Поступила: 03. 12. 2018.
Принята к публикации: 15. 12. 2018.

ALGORITHM OF PROTEIN SEQUENCE DETERMINATION BY COMBINATION OF THE EDMAN DEGRADATION METHOD AND SHOTGUN MASS SPECTROMETRY

V.S. Skvortsov*, A.V. Mikurova, N.E. Vavilov, V.G. Zgoda

Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow, 119121 Russia; *e-mail: vladlen@ibmh.msk.su

An algorithm combining advantages of the Edman degradation method and *de novo* mass-spectrometric sequencing was developed. The protein from the “Diaskintest” diagnostic test was used for analysis. The protein was digested with trypsin and 5 steps of Edman degradation were carried out sequentially for the mixture of peptides. At each stage, the resulting mixture was analyzed by shotgun mass spectrometry analysis. The results of mass-spectrometry were analyzed both by the well-known *de novo* sequencing programs Novor and PepNovo+, and by own program that clustered individual spectra with a C-terminal signature formed by Y-ions. This approach allows us to determine confidently the amino acid sequence of the N-terminal part of the peptides obtained after the protein hydrolysis by trypsin.

Key words: Edman degradation; shotgun mass spectrometry; *de novo* sequencing

ACKNOWLEDGMENTS

This work was performed within the framework of the Program for Basic Research of State Academies of Sciences for 2013-2020. Mass spectrometric measurements were performed using the equipment of the “Human Proteome” Core Facility at the Institute of Biomedical Chemistry supported by the Ministry of Education and Science of the Russian Federation (agreement 14.621.21.0017, Project ID RFMEFI62117X0017).