

ПРОТОКОЛЫ ЭКСПЕРИМЕНТОВ, ПОЛЕЗНЫЕ МОДЕЛИ, ПРОГРАММЫ И СЕРВИСЫ

pIPredict ВЕРСИЯ 2: НОВЫЕ ВОЗМОЖНОСТИ И РАБОТА С PTM

В.С. Скворцов, Н.Н. Алексейчук, Ю.В. Мирошниченко, А.В. Рыбина*

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

Программа pIPredict, созданная как инструмент предсказания изоэлектрической точки пептидов и белков, служит для генерации карт виртуального двумерного (2D) электрофореза. Метод предсказания pI основан на использовании уравнения Хендерсона-Хассельбаха. В программе использованы несколько шкал значений pKa, включая вновь добавленную по сравнению с первой версией (ProMoST), а также новую версию собственной шкалы. Добавлены модули коррекции электрофоретического сдвига по молекулярному весу и вариативного предсказания pI для заданного набора модификаций. Функция предсказания pI для PTM может быть использована для предсказания положения модифицированных форм белков на карте виртуального 2D электрофореза, а также для предсказания конкретно наблюдаемой в эксперименте протеоформы. В программе также реализованы возможности предсказания по ряду других широко распространённых шкал с учётом некоторых посттрансляционных модификаций (PTM) и добавления собственных вариантов шкалы. Программа написана на языке JAVA и свободно доступна по адресу <http://www.ibmc.msk.ru/LPCIT/pIPredict>.

Ключевые слова: белок; пептид; изоэлектрическая точка; посттрансляционные модификации; предсказание свойств

DOI: 10.18097/BMCRM00009

ВВЕДЕНИЕ

Программа pIPredict [1] была создана как инструмент предсказания изоэлектрической точки пептидов и белков, она также может быть использована для генерации простейшей карты виртуального 2D электрофореза. Метод предсказания pI основан на использовании уравнения Хендерсона-Хассельбаха [2]. Несмотря на то, что его использование даёт хорошие результаты, само уравнение является аппроксимацией, а многие факторы, влияющие на значение pI (например, зависимости от температуры и ионной силы раствора), традиционно игнорируются. Как правило, различные варианты отличаются только шкалой констант диссоциации (pKa) соответствующих групп аминокислотных остатков, "рассчитанных" различными способами. Это может быть титрование модельных пептидов [3], квантово-химические расчёты или подбор значений pKa на большом массиве данных об экспериментальных значениях pI для известных белков и/или пептидов [1, 4]. Применяют также и другие методы (не связанные с уравнением Хендерсона-Хассельбаха), например, метод опорных векторов (SVM, Support Vector Machines) [5] или искусственные нейронные сети (ANN; Artificial Neural Networks) [1]. Однако, если последовательность сильно отличается по размеру от тех, что использовали при обучении, SVM и ANN не дают удовлетворительного результата. Так как ранее встроенный в программу pIPredict метод с использованием ANN был обучен на небольших пептидах и плохо работает на белках, от него было решено отказаться. В основном, новые и обновлённые функции программы pIPredict связаны с работой с посттрансляционными модификациями (PTM). Кроме того, в программу также добавлена возможность коррекции электрофоретического сдвига по молекулярному весу по методу, описанному нами ранее [6].

ОПИСАНИЕ ПРОГРАММЫ

Программа pIPredict 2.0 написана на языке JAVA, распространяется в виде исполняемого jar-архива и доступна по адресу <http://www.ibmc.msk.ru/LPCIT/pIPredict>. Программа имеет развитый графический интерфейс пользователя (рис. 1) и следующий набор возможностей:

1. Модуль загрузки последовательностей белков и пептидов из файлов в формате FASTA или простом текстовом формате с указанием модификаций аминокислотных остатков, список которых доступен в руководстве пользователя. Формат позволяет определить произвольное число PTM, описание которых может редактироваться пользователем.

2. Модуль предсказания pI с использованием нескольких шкал значений pKa. В настоящей версии к имеющимся ранее шкалам (Bjellqvist [7] и ещё 7 шкал) добавлены: новая версия шкалы pIPredict (рис. 2), рассчитанная на основе скорректированной обучающей выборки pIPredict версии 1.0, и шкала программы ProMoST [8], позволяющая предсказывать pI с учётом фосфорилирования аминокислотных остатков. Шкалы pIPredict и ProMoST имеют значения pKa для учёта модификаций, вносящих дополнительные диссоциирующие группы (например, фосфорилирование), в остальных шкалах могут учитываться только модификации, элиминирующие такие группы, либо модифицирующие одни аминокислотные остатки в другие (например, дезаминирование аспарагина). При предсказании pI имеется возможность учитывать облигатные химические модификации (например, карбамидметилирование цистеина, что важно при оценке pI пептидов в масс-спектрометрических экспериментах) и удаление N-концевого остатка метионина (важно для оценки pI белков при анализе данных электрофореза). Как и в случае фосфорилирования удаление N-концевого метионина



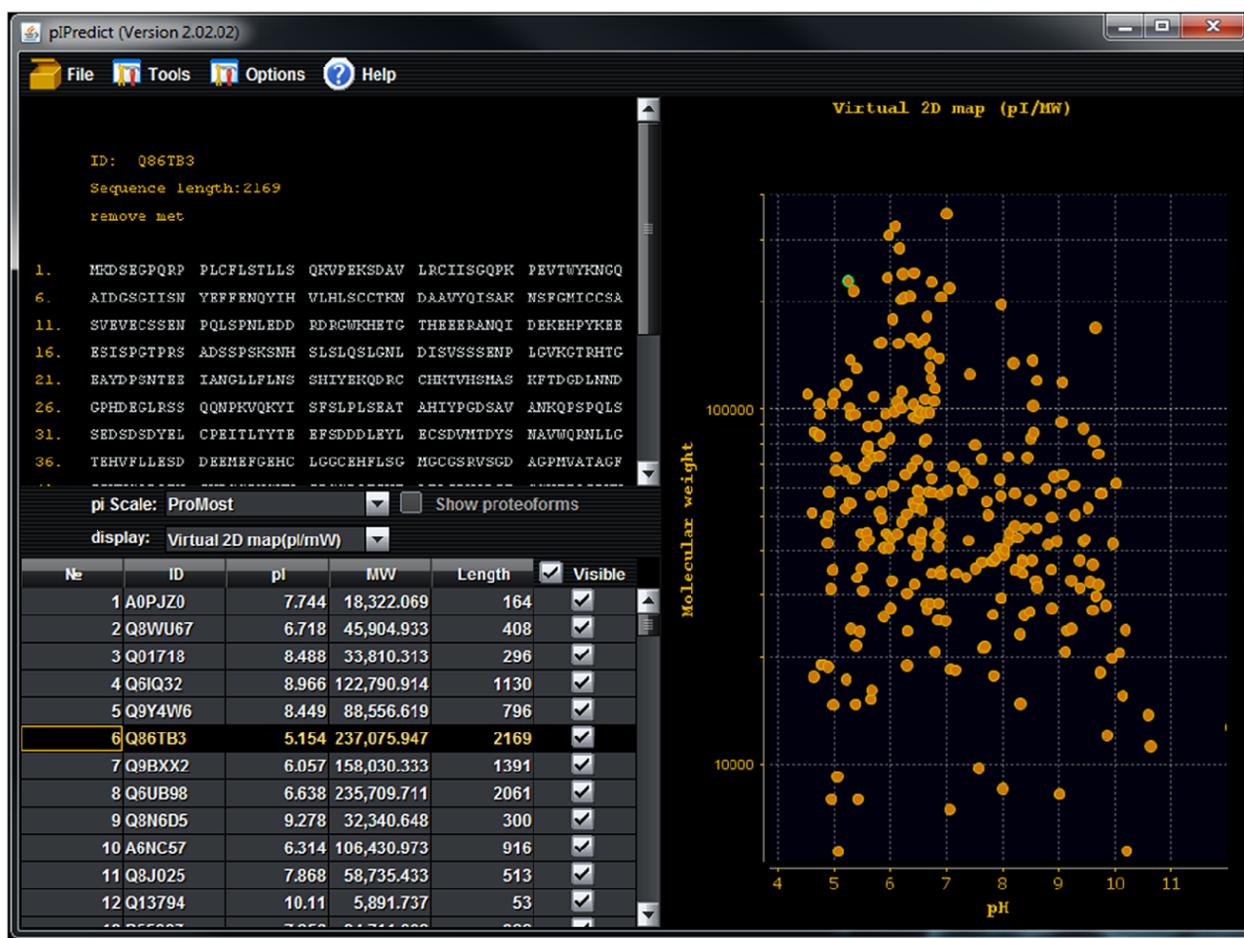


Рисунок 1. Пример пользовательского интерфейса программы piPredict 2. Загружено 277 белков XVIII хромосомы *Homo sapiens*. В левом верхнем фрейме - последовательность выбранного белка из списка (левый нижний фрейм). В правом фрейме находится виртуальная 2D карта.

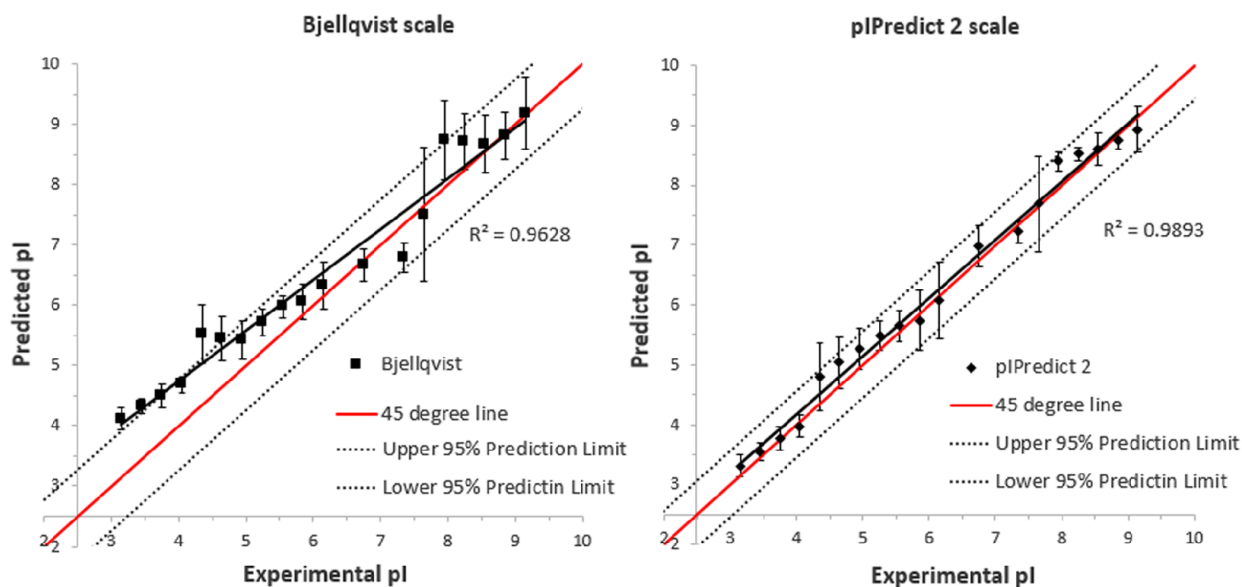


Рисунок 2. Сравнение экспериментальных и предсказанных значений pI для выборки из 7390 пептидов [5]. Поскольку данная выборка получена в результате анализа пептидов после изоэлектрического фокусирования, то экспериментальные значения определены с точностью до ± 0.15 единиц pH, или с "шагом" 0.3 единиц pH. Слева приведена шкала, идентичная шкале широко применяемого сервиса "Compute pI/Mw" (https://web.expasy.org/compute_pi/). Справа - новая шкала piPredict 2.

даёт эффект только при использовании шкал pIPredict и ProMoST. Как правило, этими изменениями можно пренебречь, так как изменения превышают среднюю ошибку предсказания примерно в 5% случаях (рис. 3); тем не менее, это может быть важно для каждого конкретного белка. Имеется возможность загрузки собственной шкалы рКа (или модификации любой из имеющихся шкал).

3. Модуль расчёта зависимости усреднённого заряда белков (пептидов) от pH для выбранных белков и модуль расчёта значений усреднённого заряда при заданном pH для каждой аминокислотной последовательности из загруженного набора белков.

4. Модуль коррекции электрофоретического сдвига по молекулярной массе [6].

5. Модуль вариативного предсказания pI для набора модификаций, полученных из мастерной аминокислотной последовательности белка по заданным правилам (рис. 4). Данная функция может быть использована как для предсказания положения модифицированной формы на карте виртуального

2D электрофореза, так и для формирования гипотезы, какая конкретно форма белка может наблюдаться в эксперименте. Например, (рис. 4) при анализе белков тканей почки *Bos Taurus* [9] для ряда белков были обнаружены несколько вариантов положения идентифицированных белков (выделены красным). Если поварьировать 4 возможных изменения (фосфорилирование остатков треонина или серина и спонтанное дезаминирование остатков глутамина или аспарагина), сделать предсказания для pI шкалами ProMoST (зелёный) и pIPredict (жёлтый) и сравнить изменения относительно варианта с крайним значением pI в основной области (в данном случае делается предположение, что данный вариант соответствует мастерной форме белка), видно, что в ряде случаев (O02691, O77834, P10096, P14568, P41976, Q28035, Q95KZ6) можно предсказать, какой форме будет соответствовать обнаруженная экспериментально протеоформа. Правда, следует отметить, что в настоящее время у авторов нет неопровержимых доказательств в пользу того, что это не случайные совпадения. Однако, по мере развития программы возможности такого рода предсказания будут исследоваться отдельно. Важным для данной функции являются точность предсказания pI (для шкалы pIPredict априорная средняя точность предсказания 0.1 значения величины pH), а также способ, которым проводится калибровка данных электрофореза. Часто полученные данные калибруются по предсказанным значениям pI для группы идентифицированных белков.

Все данные, полученные в результате работы программы pIPredict версии 2.0 могут быть экспортированы в виде текстовых файлов с разделителем, что позволяет легко использовать их в любой программе анализа электронных таблиц.

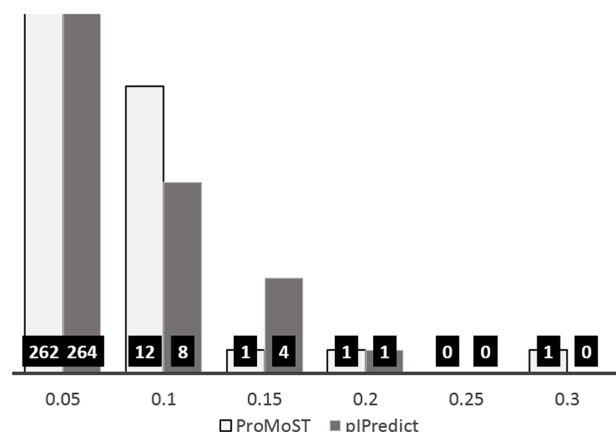


Рисунок 3. Распределение 277 белков XVIII хромосомы *Homo sapiens* по изменениям предсказанного pI (по модулю) при удалении N-концевого метионина. Используются шкалы ProMoST и pIPredict 2. Группа с незначимыми изменениями (от 0 до 0.05) обрезана по вертикали.

БЛАГОДАРНОСТИ

Работа выполнена при финансовой поддержке Программы фундаментальных научных исследований государственных академий наук на 2013–2020 годы.

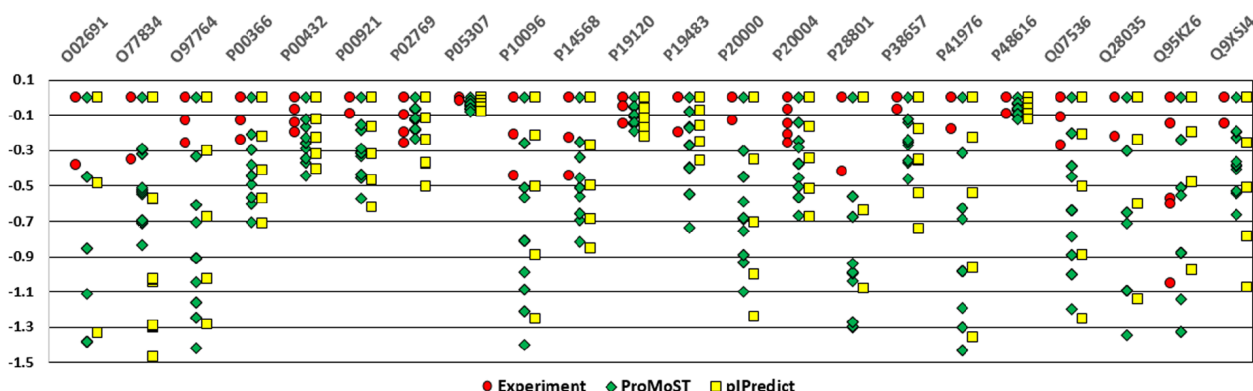


Рисунок 4. Пример сравнения предсказанных значений pI для виртуального набора PTM (фосфорилирование остатков треонина или серина и спонтанное дезаминирование остатков глутамина или аспарагина и их комбинация) с экспериментально определёнными протеоформами, полученными анализе белков тканей почки *Bos Taurus* [9]. Для ряда белков были обнаружены несколько вариантов положения идентифицированных белков (выделены красным). По оси ординат - изменение pI протеоформы относительно варианта с наиболее основным значением. Предположение об обнаруженном варианте можно сделать для белков O02691, O77834, P10096, P14568, P41976, Q28035, Q95KZ6.

ЛИТЕРАТУРА

1. Skvortsov, V.S., Alekseychuk, N.N., Khudyakov, D.V., & Romero, I.R. (2015) pIPredict: a computer tool for predicting isoelectric points of peptides and proteins. *Biomeditsinskaya khimiya*, 61(1), 83-91. DOI: 10.18097/PBMC20156101083
2. Po, H.N., & Senozan, N.M. (2001) The Henderson-Hasselbalch equation: its history and limitations. *Journal of Chemical Education*, 78(11), 1499. DOI: 10.1021/ed078p1499
3. Patrickios, C.S. (1995) Polypeptide amino acid composition and isoelectric point: 1. A closed-form approximation. *Journal of colloid and interface science*, 175(1), 256-260. DOI: 10.1006/jcis.1995.1454
4. Gauci, S., Van Breukelen, B., Lemeer, S.M., Krijgsveld, J., & Heck, A.J. (2008) A versatile peptide pI calculator for phosphorylated and N-terminal acetylated peptides experimentally tested using peptide isoelectric focusing. *Proteomics*, 8(23-24), 4898-4906. DOI: 10.1002/pmic.200800295
5. Perez-Riverol Y., Audain E., Millan A., Ramos Y., Sanchez A., Vizcaíno J.A., Wang R., Muller M., Machado Y.J., Betancourt L.H., Padrón G., Besada V. (2012) Isoelectric point optimization using peptide descriptors and support vector machines. *Journal of proteomics*, 75(7), 2269-2274. DOI: 10.1016/j.jprot.2012.01.029
6. Skvortsov, V.S., Alekseychuk, N.N., & Rybina, A.V. (2017) Correction of the electrophoretic shift in virtual 2D SDS-PAGE electrophoresis. *Biomeditsinskaya khimiya*, 63(3), 278-283. DOI: 10.18097/PBMC20176303278
7. Bjellqvist B., Hughes G.J., Pasquali Ch., Paquet N., Ravier F., Sanchez J.-Ch., Frutiger, S. & Hochstrasser D.F. (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, 14, 1023-1031. DOI: 10.1002/elps.11501401163
8. Halligan, B.D. (2009) ProMoST: a tool for calculating the pI and molecular mass of phosphorylated and modified proteins on two-dimensional gels. In *Phospho-Proteomics* (pp. 283-298). Humana Press. DOI: 10.1007/978-1-60327-834-8_21
9. Talamo, F., D'Ambrosio, C., Arena, S., Del Vecchio, P., Ledda, L., Zehender, G., Ferrara, L. & Scaloni, A. (2003) Proteins from bovine tissues and biological fluids: defining a reference electrophoresis map for liver, kidney, muscle, plasma and red blood cells. *Proteomics*, 3(4), 440-460. DOI: 10.1002/pmic.200390059

Поступила: 12. 04. 2018.
Принята к публикации: 22. 06. 2018.

pIPredict VERSION 2: NEW FEATURES AND PTM ANALYSIS

V.S. Skvortsov, N.N. Alekseychuk, Yu.V. Miroshnichenko, A.V. Rybina*

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

pIPredict was created as a tool for prediction of the isoelectric point of peptides and proteins. It can also generate virtual 2D electrophoresis maps. The method of pI prediction is based on the Henderson-Hasselbach equation. In a new version the ProMoST and our new scales of pKa values were added. The other added features included: correction of electrophoretic shift by analyzing amino acid composition of proteins and prediction of pI values for proteins with a new set of posttranslational and other chemical modifications. Prediction of pI for proteins with PTM can be used to predict position of modified proteoforms on the virtual 2D electrophoresis map or as the tool of identifying which particular proteoform was observed in the experiment. The program also includes several widely used pKa scales, that can partially calculate values for proteins with some post-translational modifications. pIPredict is created as JAVA application and is freely available at <http://www.ibmc.msk.ru/LPCIT/pIPredict>.

Key words: protein; peptide; isoelectric point; posttranslational modifications; property prediction

ACKNOWLEDGEMENTS

This work was performed within the framework of the Program for Basic Research of State Academies of Sciences for 2013-2020.