

УДК 57.088.1

©Коллектив авторов

ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА РАЗМЕРА ПРОТЕОМОВ КЛЕТОК И ПЛАЗМЫ КРОВИ ЧЕЛОВЕКА

**С.Н. Нарыжный^{1,2*}, В.Г. Згода¹, М.А. Майнскова¹, Н.Л. Ронжина²,
Н.В. Белякова², О.К. Легина², А.И. Арчаков¹**

¹Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича,
119121, Москва, ул. Погодинская, 10; эл. почта: snaryzhny@mail.ru

²Петербургский институт ядерной физики имени Б.П. Константинова,
Национальный исследовательский центр “Курчатовский институт”,
188300, Ленинградская область, г. Гатчина

Огромный диапазон концентраций различных протеоформ и недостаточная чувствительность современных методов их детектирования не позволяют пока напрямую определить размеры протеомов (количество всех протеоформ), особенно в плазме. В наших исследованиях была предпринята оценка размера протеомов разного происхождения (начиная с клеток разного типа и заканчивая плазмой человека). Это было сделано на основе использования двумерного электрофореза (2-DE) с последующим окрашиванием белков красителями разной чувствительности. Была определена формула зависимости количества детектируемых полипептидов (пятен) в двумерном геле от чувствительности красителя. Экстраполируя данную зависимость на область чувствительности, позволяющую детектировать самую малую единичную молекулу белка, было подсчитано, что одна клетка человека (HepG2) может содержать около 70000 видов белка, а вся плазма – 1,5 млн. Применение данного подхода к другим клеточным объектам показало правомочность такой экстраполяции. Так, для гепатоцита получилось 70000 протеоформ, микоплазмы (*Acholeplasma laidlawii*) – 1100, дрожжей (*Saccharomyces cerevisiae*) – 40000, *Escherichia coli* – 6200, *Pyrococcus furiosus* – 3400. Важность полученных результатов состоит в том, что на их основании можно судить об организации протеома человека и планировать дальнейшие ходы по изучению его функционирования.

Ключевые слова: протеом, протеоформы, двумерный электрофорез.

DOI: 10.18097/PBMC20156102279

ВВЕДЕНИЕ

После завершения секвенирования генома человека и получения информации обо всех его генах похожая задача встала и перед исследователями протеома человека [1-5]. Соответственно, в 2010 году стартовал проект “Протеом человека” [6]. Хотя размер генома человека (количество генов) и его экспрессия достаточно изучены, общего мнения о размере протеома человека (количестве протеоформ) до сих пор нет. Следуя формуле “один ген – один белок”, должно быть как минимум ~20300 белков (полипептидов) [7, 8]. Они представлены каноническими или мастерными аминокислотными последовательностями. Модифицированные полипептиды, происходящие из одного и того же гена, могут составлять до 100 форм [9, 10]. Среди этих форм имеются белки с аминокислотными заменами (полиморфизм единичных аминокислот) (SAP),

продукты альтернативного сплайсинга (AS) и посттрансляционных модификаций (PTM) [9]. Также могут быть и продукты объединения разных генов [11]. Для всех этих мастерных и модифицированных молекул был предложен термин “белковый вид” или “протеоформа” [12, 13]. Основная проблема идентификации всех этих форм заключается в чувствительности существующих протеомных технологий и огромном динамическом диапазоне (6-7 порядков) концентраций разных белковых молекул, входящих в состав протеома [2, 3, 14, 15]. Современные методы детектирования белка пока не позволяют получить всю картину протеома человека, поэтому ближайшая задача протеомики состоит как минимум в оценке его размера. Эта задача и стояла в наших исследованиях.

Хотя двумерный электрофорез (2-DE) и не является идеальным методом разделения и идентификации всех белков, он всё-таки позволяет подойти близко

* - адресат для переписки

к определению количества протеоформ [16-18]. Он позволяет детектировать SAP и PTM, что невозможно сделать в широких масштабах существующими методами масс-спектрометрии (МС) [17]. Теоретически, различные протеомные подходы в комбинации с полной геномной информацией могли бы быть использованы для идентификации каждого белка в 2-DE геле [2, 3, 15-21]. Подход, который был использован в наших исследованиях, основан на 2-DE, методе, благодаря которому, в своё время и появился термин “протеом” [1]. Термин “белковый вид” или “протеоформа” относится к самой малой единице протеома, то есть полипептиду, имеющему уникальную последовательность и несущему уникальный набор посттрансляционных модификаций. Свойства 2-DE как раз и позволяют учитывать влияние посттрансляционных модификаций и сплайсинга на появление разнообразных белковых видов. Теоретически, такой протеомный подход в совокупности с генетическими данными мог бы быть использован для идентификации каждого белкового типа и получения информации о динамике протеомов в клетке или плазме. На практике, однако, ситуация находится пока на стадии, когда протеомике всё еще необходимо реализовать полный потенциал 2-DE. Ранее уже были попытки оценить размер протеома плазмы, которые давали самый широкий диапазон значений – от нескольких тысяч до миллионов [2, 6]. Для белков отдельной клетки человека (HeLa) результаты последних исследований позволяют говорить о присутствии как минимум 10 тысяч белковых видов [22]. Принимая это во внимание, мы полагаем, что наша предыдущая оценка размера протеомов на основе 2-DE очень близка к реальности. Была получена функция зависимости количества детектированных белковых пятен от чувствительности или предела чувствительности (LOD; limit of detection). Экстраполяция графика этой функции в область минимального LOD дала возможность подсчёта различных протеоформ, присутствующих на уровне одного полипептида [5]. В данной работе мы приводим наши подсчёты по оценке размеров расширенного круга клеточных протеомов, а также адаптируем данный подход для плазмы крови человека.

МЕТОДИКА

Все используемые реагенты были получены от “Sigma-Aldrich” (США), если не указан другой производитель. Остальные реагенты были от следующих компаний: “Pierce” (США) (дитиотреитол, DTT), коктейли ингибиторов протеаз; “GE Healthcare” (США) (IPG DryStrip – гелевые полоски, IPG-буферы, DryStrip-покрывающая жидкость, Кумасси R350); “Promega” (США) (Trypsin Gold); “Bio-Rad” (США) (маркеры молекулярной массы белков для электрофореза); “Биолот” (Россия) (среды RPMI-1640 и DMEM для роста клеток, эмбриональная сыворотка телёнка); “Orange Scientific” (Бельгия) (культуральные флаконы Карреля).

Подготовка проб и двумерный электрофорез

Клетки *E. coli*, клетки печени человека (гепатоциты), HepG2, дрожжей (*Saccharomyces cerevisiae*), микоплазмы (*Acholeplasma laidlawii*) экстрагировали, как описано ранее [23-26]. Клетки растворяли в лизирующем буфере (LB) (7 М мочевины, 2 М тиомочевина, 4% CHAPS, 1% DTT, 2% IPG буфер (pH 3–10), коктейль протеазных ингибиторов). Разделение белков изоэлектрофокусированием (IEF) проводили, используя полоски DryStrip и следуя протоколу производителя. Образцы, находящиеся в лизирующем буфере LB, смешивали с регидрирующим раствором (7 М мочевины, 2 М тиомочевина, 2% CHAPS, 0,3% DTT, 0,5% IPG буфер, pH 3-10, 0,001% бромфенол голубой) в общем объёме 130 мкл (7-см полоска), или 200 мкл (11-см полоска), или 350 мкл (18-см полоска). Чтобы подготовить для разделения в первом направлении, IPG-полоски регидрировали ночь при 10°C, помещая их “вниз лицом” в регидратирующий раствор, содержащий белковые пробы, в Immobiline DryStrip Reswelling Tray. Фокусировку IEF проводили при 20°C, используя PROTEAN IEF cell (“Bio-Rad”), следуя протоколу производителя. Общее напряжение (Vh) было 16000, 30000, или 60000 для 7-см, 11-см, или 18-см IPG-полосок, соответственно. После IEF полоски вымачивали 10 мин в уравнивающем растворе (50 мМ трис-HCl (pH 6,8), 6 М мочевины, 2% SDS, 30% глицерин), содержащем 1% DTT, и повторяли инкубацию в уравнивающем растворе, содержащем вместо DTT 5% иодацетамид. Затем каждую полоску IPG помещали на верхний торец геля второго направления и запечатывали горячим раствором 0,5% агарозы, содержащим TGS (трис, глицин, SDS) электродный буфер (25 мМ трис (pH 8,3), 200 мМ глицин, 0,1% SDS). Разделение во втором направлении (SDS-PAGE) проводили в 12% полиакриламидном геле при комнатной температуре и при постоянной мощности (3 Вт/гель), используя электрофоретические ячейки MiniProtein 3 Dodeca, Criterion Dodeca, или Protean II XL Multi-cell (“Bio-Rad”) [25, 26].

Окрашивание

После 2-DE разделения гели фиксировали и промывали 20 мин на качалке в фиксирующем растворе (30% этанол, 10% уксусная кислота). Промывку повторяли как минимум один раз, и гели окрашивали 1,5 ч в красителе (0,1% Амидо Чёрный, или 0,1% Кумасси R250, или 0,1% Кумасси R350 в 30% этаноле, 10% уксусной кислоте) с последующей отмывкой в воде. Отмывку проводили как минимум 3 ч, многократно меняя воду до обесцвечивания фона. Серебрение проводили так, как описано ранее [5, 27, 28]. Радиоактивное мечение ¹²⁵I проводили, используя набор Pierce Iodination Tubes, как описано ранее [5, 29]. Флуоресцентное мечение белков в минимальном и насыщающем режиме проводили, используя краситель Cy5 (“BioDye”, Россия) по протоколу производителя [5]. Гели сканировали на денситометрах GS-800 (“Bio-Rad”) и Typhoon 9410

(“GE Healthcare”). Полученные изображения анализировали и пятна подсчитывали, используя программы ImageMaster 2D Platinum 7.0 (“GE Healthcare”) и PDQuest (“Bio-Rad”).

Принцип метода оценки количества протеоформ

Подход, использованный в наших исследованиях, основан в первую очередь на 2-DE с учётом нескольких допущений. Первое допущение – каждое пятно на двумерном геле представляет определенный белковый вид (протеоформу). Строго говоря, это не совсем правильно [30, 31], так как некоторые пятна содержат несколько протеоформ (особенно после 2-DE низкого разрешения). Но, по-видимому, это не сильно влияет на наши вычисления, так как мы использовали 2-DE гели разного размера (разрешения) и это не сильно повлияло на конечные числа. Второе – используя краситель с повышенной чувствительностью (меньшим LOD) мы сможем детектировать больше протеоформ (N), чем с красителем с большим LOD при использовании одинакового количества белка [3]. Третье – имеется определённая зависимость количества детектированных в одинаковых количествах белка протеоформ ($Z=N/Q$) от чувствительности красителя (LOD). Соответственно, был предложен экспериментальный подход детектирования белковых пятен путём экстраполяции в область максимальной чувствительности или минимального LOD (рис. 1). Минимальный LOD (или максимальная чувствительность) означает возможность детектировать наименьшее количество белка в 2-DE. В нашем случае это одна копия самой малой полипептидной молекулы, которая имеет массу ~ 6 кДа (или 10^{-20} г). Эта масса взята из базы знаний UniProt, она также основана на пределах разделения в 2-DE [32].

В принципе, только с использованием такого красителя возможно детектирование всех протеоформ (даже самых малых в виде пятен) на 2-DE геле, где разделены белки одной клетки. Поэтому, чтобы определить количество таких протеоформ в единичной клетке (N_{\max}), необходимо найти функцию зависимости нормализованных на единицу белка (Z) количества пятен от LOD. Экстраполяция значений этой функции в область максимальной чувствительности (10^{-20} г) позволяет получить Z_{\max} . Умножение Z_{\max} на количество белка в единичной клетке (Q_{\min}) позволяет получить максимальное количество протеоформ в клетке (N_{\max}).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Экспериментальный подход по определению количества протеоформ с помощью экстраполяции в область максимальной чувствительности был использован в наших исследованиях [3, 5]. Максимальная чувствительность означает возможность детектировать единичные копии полипептидных молекул в образце ($PCN=1$), то есть минимальное количество белка. Данный подход основан на наблюдении, что используя краситель с большей чувствительностью, мы сможем детектировать больше протеоформ (N), чем с красителем меньшей чувствительности. Чтобы определить число протеоформ в единичной клетке, мы использовали функциональную зависимость нормализованного на единицу количества белка количества протеоформ (Z) от чувствительности или предела чувствительности красителя. Экстраполяция данной функции в область максимальной чувствительности ($PCN=1$) позволяет получить Z_{\max} . А в итоге, умножение Z_{\max} на количество белка в единичной клетке (Q) даёт максимальное количество

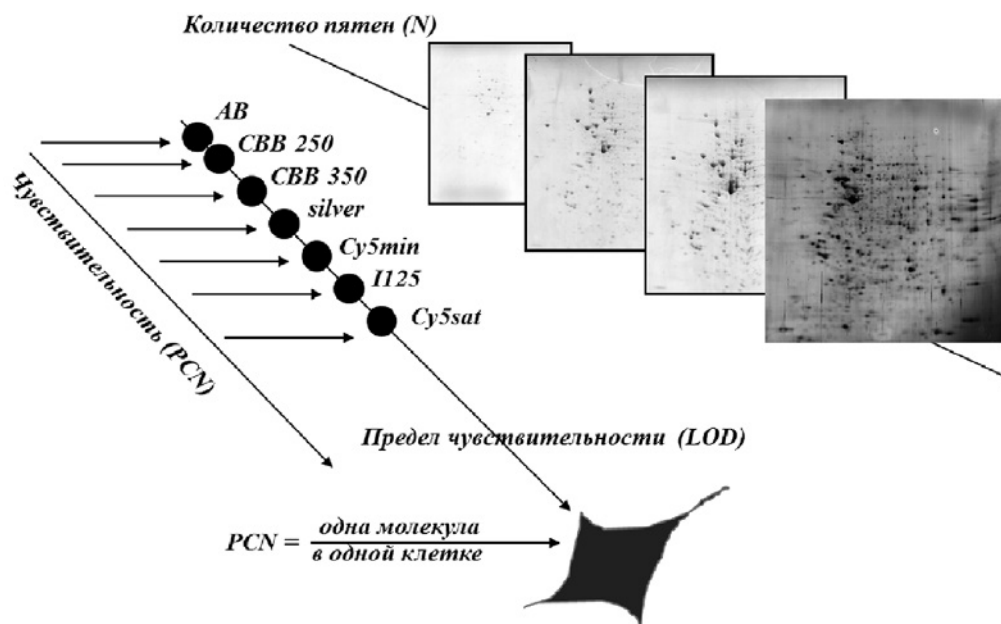


Рисунок 1. Схема экспериментального подхода по оценке количества протеоформ окраской двумерных гелей красителями разной чувствительности.

КОЛИЧЕСТВО ПРОТЕОФОРМ В ПРОТЕОМАХ КЛЕТОК И ПЛАЗМЫ КРОВИ

протеоформ (N_{\max}) в клетке. Чтобы получить такую функцию, нам необходимо сначала подсчитать предел чувствительности для всех красителей, используемых в эксперименте. Для этого несколько образцов с понижающимся содержанием бычьего сывороточного альбумина (БСА) прогоняли одномерным электрофорезом SDS-PAGE. Несколько гелей окрашивали различными красителями, и оценивали минимальное количество БСА, которое можно было детектировать [5]. Среди всех красителей Амидо Чёрный имеет наименьшую чувствительность, ($LOD = 5 \times 10^{-8}$ г) на полосу ($PCN = 4 \times 10^{12}$), ^{125}I и насыщенный $Cy5$ – наивысшую чувствительность, LOD составили 2×10^{-11} г и 10^{-11} г ($PCN = 1,6 \times 10^9$ и $0,8 \times 10^9$) на полосу, соответственно (таблица). Ранее были протестированы белковые экстракты бактерий, *E. coli*, *P. furiosus* [5]; в данной работе мы расширили наши исследования с использованием образцов микоплазмы. По сравнению с дрожжами, клетками человека и особенно плазмой эти объекты имеют протеомы гораздо меньшего размера. Чтобы вычислить нормализованное на единицу количества белка (Z) количество белковых пятен, необходимо вначале выявить зависимость количества этих пятен (N) от количества нанесенного белка. Для этого было окрашено различными красителями несколько наборов двумерных гелей (2-DE), содержащих разное количество белка (как минимум 3 геля для каждого количества). Для каждого красителя был построен график зависимости количества пятен (N) от количества нанесенного белка (Q). Полученные графики хорошо описываются линейной функцией с коэффициентом корреляции $>0,9$, угол наклона которой и даёт нормализованное число пятен $Z=N/Q$ (число пятен нормализовано на микрограмм белка). Далее, чтобы оценить количество протеоформ в одной клетке (N_{\max}), была получена зависимость между нормализованным

количеством протеоформ (Z) и чувствительностью детектирования (LOD) [5]. Экстраполяция данной функции в область максимальной чувствительности (10^{-20} г) позволяет определить Z_{\max} . В итоге умножение Z_{\max} на количество белка в одной клетке (Q_{\min}) даёт количество протеоформ в клетке (N_{\max}). Для бактерий такие вычисления дают очень похожие величины ($\sim 4 \times 10^{10}$ – для *E. coli*, и $2,8 \times 10^{10}$ – для *P. furiosus*) [5]. Используя формулу $Z=N/Q$ (1), мы можем рассчитать количество протеоформ в клетке, $N_{\max} = Z_{\max} \times Q_{\min}$ (2). В этом случае Q_{\min} обозначает количество белка в одной клетке, которое равно $\sim 1,55 \times 10^{-7}$ мкг для *E. coli* и $\sim 1,2 \times 10^{-7}$ мкг для *P. furiosus* [33]. Так, $Z=N/Q$ (1), $N_{\max} = Z_{\max} \times Q_{\min}$ (2). Для *E. coli* количество протеоформ $\sim 4 \times 10^{10} \times 1,55 \times 10^{-7} = 6200$. Для *P. furiosus* $\sim 2,8 \times 10^{10} \times 1,2 \times 10^{-7} = 3400$ [5].

Ранее, используя ZOOM® IEF как стратегию фракционирования, Richardson и соавт. сумели детектировать с помощью 2-DE 5525 белковых пятен, представляющих протеом *E. coli* [33]. Они утверждают, что это число отражает почти все типы полипептидов в одной клетке *E. coli*, так как техника позволяет детектировать единичные белковые копии [33]. По предыдущим оценкам Tonella и соавт. [34], *E. coli* может содержать 6022 протеоформ. Эти числа очень близко соотносятся с нашими вычислениями (6200), подтверждая правомочность используемого нами подхода.

Оценка количества протеоформ в клетках человека

Так же, как и в случае с микроорганизмами, 2-DE гели с разным количеством белка клеток печени или линии HepG2 были окрашены различными красителями. Были построены зависимости количества детектированных пятен (N) от количества нанесенного белка [5]. Из трёх самых чувствительных методов, $Cy5_{sat}$, $Cy5_{min}$, ^{125}I ,

Таблица. Предел чувствительности (LOD) при детектировании белков различными красителями, соответствующие значения Z (N/Q) и оценка количества протеоформ (N_{\max}).

Детектирование	LOD^*		$Z^* = N/Q$ (мкг $^{-1}$)						
	г	PCN	HepG2*	<i>E. coli</i> *	<i>P. furiosus</i> *	Микоплазма	Дрожжи	Печень	Плазма
Амидо Чёрный	5×10^{-8}	5×10^{12}	$0,8 \pm 0,1$	$0,6 \pm 0,07$	$0,6 \pm 0,07$	$0,4 \pm 0,05$	$0,3 \pm 0,03$	$1,3 \pm 0,2$	$1,0 \pm 0,1$
Кумасси R250	$2,5 \times 10^{-8}$	$2,5 \times 10^{12}$	$1,6 \pm 0,2$	N/A	N/A	N/A	N/A	N/A	$2 \pm 0,3$
Кумасси R350	10^{-8}	1012	$2,1 \pm 0,3$	$2,1 \pm 0,3$	$2,2 \pm 0,3$	$0,8 \pm 0,1$	$0,7 \pm 0,1$	$2,1 \pm 0,3$	$1,8 \pm 0,3$
Серебро	10^{-9}	1011	10 ± 2	21 ± 3	18 ± 3	16 ± 3	17 ± 3	10 ± 2	13 ± 3
$Cy5_{min}$	$0,3 \times 10^{-10}$	3×10^9	97 ± 12	302 ± 36	275 ± 30	50 ± 6	65 ± 7	100 ± 12	20 ± 4
^{125}I	2×10^{-11}	2×10^9	124 ± 14	N/A	N/A	N/A	N/A	N/A	45 ± 8
$Cy5_{sat}$	10^{-11}	109	154 ± 18	N/A	N/A	N/A	N/A	N/A	41 ± 8
LgZ_{\max}	-	-	8	10,55	10,45	11	10,5	8	6,17
Q_{\max} (мкг)	-	-	$0,7 \times 10^{-8}$	$1,2 \times 10^{-7}$	$1,55 \times 10^{-7}$	4×10^{-6}	7×10^{-4}	7×10^{-4}	1
N_{\max}	-	-	70000	6200	3400	1600	40000	70000	1500000

Примечание: N - количество пятен; Q - количество белка в микрограммах (мкг); LgZ_{\max} - экстраполяция к пределу чувствительности в 1 PCN или 10^{-20} г; LOD (limit of detection) - предел чувствительности в граммах (г) или в PCN (protein copy number), количестве детектируемых полипептидов с массой в 6 кДа; N_{\max} - оценка размера протеомов (количество протеоформ), полученная после умножения Z_{\max} на массу протеома (Q_{\max}) в микрограммах (мкг); N/A - нет данных. * - некоторые данные адаптированы из [5].

величины Z составили 154, 97, и 124, соответственно (таблица). Значения Z для печени и HepG2 оказались весьма схожими как для серебра (10), так и для самых слабых методов окрашивания – Кумасси R350 – 2,1, Кумасси R250 – 1,6 и Амидо Чёрный – 0,8). Соответственно, зависимость Z от LOD для клеток печени и HepG2 практически совпадает. Линейная зависимость для HepG2 клеток имеет коэффициент корреляции $R = -0,99$, а для клеток печени – $-0,85$. Экстраполяция к минимальному LOD (PCN = 1) даёт величину Z_{\max} равной 108. Следуя формуле $Z=N/Q$, мы находим максимальное количество протеоформ в клетке $N_{\max} = Z_{\max} \times Q_{\min}$. Здесь Q_{\min} обозначает количество белка в одной клетке [35, 36]. Для гепатоцита или клетки HepG2 Q_{\min} равно примерно 7×10^{-4} мкг [36]. Следовательно, $N_{\max} = Z_{\max} \times Q_{\min} = 108 \times 7 \times 10^{-4} = 7 \times 10^4$ протеоформ. То есть, в одной клетке HepG2 или гепатоците содержится как минимум 70×10^3 полипептидных форм. Эта оценка согласуется с экспериментальными данными Nagaraj и соавт. по оценке протеома клеток HeLa [22]. В данной работе, где был проведён МС/МС анализ без учёта PTM и альтернативных сплайс (AS) вариантов, авторы идентифицировали 10255 белков. В нашем случае свойства 2-DE позволяют учитывать PTM [37] и AS [38] при оценке размера протеома. Хотя максимальное число пятен в клетках HepG2 составило всего 3245 [5], экстраполяция в область детектирования единичных молекул доводит эту величину до 70000, указывая на то, что PTM и AS увеличивают разнообразие протеоформ как минимум в 7 раз.

Оценка количества протеоформ в плазме

В отличие от ситуации с единичной клеткой, протеом которой имеет максимальную массу 7×10^{-10} г (гепатоциты), протеом плазмы гораздо больше, и, что особенно существенно, не имеет явно выраженных размеров. “Среднестатистический” индивидум имеет ~2,5 литра или 250 г белка в плазме с концентрацией ~100 мг/мл [39, 40]. Процесс истощения 6 мажорных белков, включая сывороточный альбумин, иммуноглобулины IgG/IgA, трансферрин, антитрипсин и гаптоглобин приводит к удалению примерно 90% белка. Соответственно, белки истощенной плазмы представляют 10% или 10 мг в каждом мл. Даже предполагая, что плазма в человеческом теле является гомогенным раствором, мы все равно должны определить предел этой гомогенности. Другими словами, необходимо определить минимальный объём плазмы, который отражает свойства всей плазмы организма. Принимая в расчет динамический диапазон концентраций различных плазменных белков в 7 порядков, для наших расчётов мы приняли объём в 0,1 мкл. Соответственно, Q_{\min} плазмы будет равным не более 1 мкг. Параметры LOD и Z были рассчитаны по такой же схеме, что и в случае клеток HepG2. Некоторые числа были рассчитаны на основе предыдущих публикаций [3, 5]. Дополнительные данные были получены с использованием окрашивания белков Кумасси R350, серебром или

мечением ^{125}I (таблица). Зависимость Z от LOD была отражена в логарифмических шкалах (рис. 2). На оси абсцисс указан предел чувствительности (PCN) для каждого красителя, а на оси ординат – нормализованное количество пятен (Z). Полученная зависимость очень хорошо коррелирует с линейной функцией с коэффициентом корреляции $R = -0,925$. Экстраполируя функцию в область детектирования одного полипептида, мы получаем величину \hat{Z} (Z_{\max}) равную 1500000. Из уравнения $Z=N/Q$ мы можем оценить максимальное количество протеоформ с использованием уравнения $N_{\max} = Z_{\max} \times Q_{\min}$. $N_{\max} = 1500000/\text{мкг} \times 1 \text{ мкг} = 1500000$. Следовательно, протеом плазмы человека может содержать не менее $1,5 \times 10^6$ протеоформ. Так как в процессе приготовления образца иммуноглобулины были удалены вместе с другими мажорными белками, то данная оценка не учитывает их разнообразия в плазме, которое может дополнительно добавить ещё столько же протеоформ [40]. Данная оценка коррелирует с нашими предыдущими вычислениями [3, 5] и комбинаторными предсказаниями других авторов [19, 20].

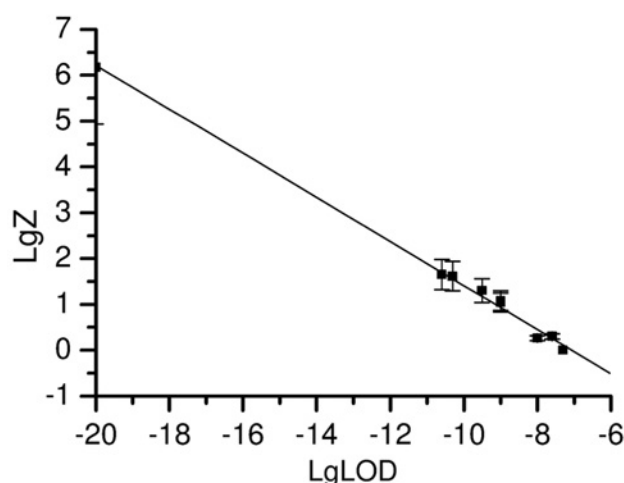


Рисунок 2. Оценка размера протеома плазмы. Зависимость Z (N/Q) от предела чувствительности (LOD) с использованием логарифмической шкалы. Положение точек соответствует линейному уравнению: $LgZ = -3,4 - 0,48LgLOD$, $R = -0,96$, $SD = 0,1987$, $p < 1,51 \times 10^{-4}$.

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Так как ситуация с размером протеомом человека ещё не до конца решена, необходимо искать различные пути её решения, в противном случае конечная цель проекта “Протеом человека” получить полный каталог протеоформ останется недостижимой. Недостаточная чувствительность методов детектирования белковых молекул не позволяет пока получить полную картину человеческого протеома. Но используя имеющиеся технологии, можно, как минимум, оценить его размер. С этой целью мы предложили подход, в котором используется зависимость детектирования количества протеоформ

от чувствительности метода. Ограниченные в наших экспериментах пределом чувствительности в 10^8 белковых молекул, мы экстраполировали экспериментальные данные в область чувствительности единичных молекул. Данная экстраполяция позволила нам оценить размеры протеомов разного происхождения, включая человеческие клетки, а также плазму человека. Необходимо подчеркнуть, что единичная клетка и плазма являются очень разными объектами. В сущности, клетка – это микро, а плазма – самый большой макрообъект человеческого тела, активно взаимодействующий со всеми органами человека. Если клетка, как протеом, является функциональной единицей с ограничениями в пространстве (объём клетки $\sim 7500 \text{ см}^3$), плазма – это набор разных протеомов с общим объёмом до 2,5 л. Такой большой объём ведет к очень сильному разбавлению минорных белков и широкому диапазону их концентраций – более чем 10 порядков концентраций в молярном выражении [19, 41]. Таким образом, плазма включает в себя практически всё разнообразие протеоформ, присутствующих в клетках разного типа человека, и низкокопийные белки составляют большую часть её протеома. Большинство этих белков не являются ни структурными, ни функциональными компонентами плазмы, а попадают в неё из различных клеток тела. И наоборот, единичная клетка имеет меньшее количество протеоформ, адаптированных для выполнения специфических клеточных функций.

Работа выполнена в рамках Программы фундаментальных научных исследований государственных академий наук на 2013–2020 годы. Авторы признательны чл.-корр. РАН В.М. Говоруну и В.Т. Пешехонову за предоставление, соответственно, клеток микоплазмы и клеток дрожжей.

ЛИТЕРАТУРА

- Wilkins M.R., Appel R.D., Van Eyk J.E., Chung M.C.M., Görg A., Hecker M., Huber L.A., Langen H., Link A.J., Paik Y.-K. et al. (2006) *Proteomics*, **6**, 4-8.
- Archakov A.I., Ivanov Y.D., Lisitsa A.V., Zgoda V.G. (2007) *Proteomics*, **7**, 4-9.
- Archakov A., Ivanov Y., Lisitsa A., Zgoda V. (2009) *Proteomics*, **9**, 1326-1343.
- Archakov A., Zgoda V., Kopylov A., Naryzhny S., Chernobrovkin A., Ponomarenko E., Lisitsa A. (2012) *Expert Rev. Proteomics*, **6**, 667-676.
- Naryzhny S., Lisitsa A., Zgoda V., Ponomarenko E., Archakov A. (2014) *Electrophoresis*, **35**, 895-900.
- Omenn G.S. (2004) *Proteomics*, **4**, 1235-1240.
- Beadle G.W., Tatum E.L. (1941) *Proc. Natl. Acad. Sci. USA*, **27**, 499-506.
- International Human Genome Sequencing Consortium (2004) *Nature*, **431**, 931-945.
- Lisitsa A., Moshkovskii S., Chernobrovkin A., Ponomarenko E., Archakov A. (2014) *Expert Rev. Proteomics*, **1**, 121-129.
- Li M., Durbin K.R., Sweet S.M., Tipton J.D., Zheng Y., Kelleher N.L. (2013) *Proteomics*, **13**, 2585-2596.
- Casado-Vela J.C., Lacal J.C., Elortza F. (2013) *Proteomics*, **13**, 5-11.
- Smith L.M., Kelleher N.L. (2013) *Nat. Methods*, **10**, 186-187.
- Jungblut P.R., Holzhütter H.G., Apweiler R., Schlüter H. (2008) *Chem. Cent. J.*, **2**, 16.
- Garbis S., Lubec G., Fountoulakis M. (2005) *J. Chromatogr. A*, **1077**, 1-18.
- Corthals G.L., Wasinger V.C., Hochstrasser D.F., Sanchez J.C. (2000) *Electrophoresis*, **21**, 1104-1115.
- O'Farrell P.H. (1975) *J. Biol. Chem.*, **250**, 4007-4021.
- Klose J. (1999) *Methods Molecular Biology*, **112**, 147-172.
- Gygi S.P., Corthals G.L., Zhang Y., Rochon Y., Aebersold R. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 9390-9395.
- Adkins J.N., Varnum S.M., Auberry K.J., Moore R.J., Angell N.H., Smith R.D., Springer D.L., Pounds J.G. (2002) *Mol. Cell. Proteomics*, **1**, 947-955.
- Nielsen M.L., Savitski M.M., Zubarev R.A. (2006) *Mol. Cell. Proteomics*, **5**, 2384-2391.
- Omenn G.S., States D.J., Adamski M., Blackwell T.W., Menon R., Hermjakob H., Apweiler R., Haab B.B., Simpson R.J., Edde J.S. et al. (2005) *Proteomics*, **5**, 3226-3245.
- Nagaraj N., Wisniewski J.R., Geiger T., Cox J., Kircher M., Kelso J., Pääbo S., Mann M. (2011) *Mol. Systems Biology*, **7**, 548.
- Naryzhny S.N. (2009) *Anal. Biochem.*, **392**, 90-95.
- Naryzhny S.N. (1996) *Anal. Biochem.*, **238**, 50-53.
- Naryzhny S.N., Lee H. (2001) *Electrophoresis*, **22**, 1764-1775.
- Naryzhny S.N., Lee H. (2003) *Proteomics*, **3**, 930-936.
- Shevchenko A., Tomas H., Havlis J., Olsen J.V., Mann M. (2006) *Nature protocols*, **1**, 2856-2860.
- Winkler C., Denker K., Wortelkamp S., Sickmann A. (2007) *Electrophoresis*, **28**, 2095-2099.
- Chizzonite R., Truitt T., Podlaski F.J., Wolitzky A.G., Quinn P.M., Nunes P., Stern A.S., Gately M.K. (1991) *J. Immunol.*, **147**, 1548-1556.
- Thiede B., Koehler C.J., Strozynski M., Treumann A., Stein R., Zimny-Arndt U., Schmid M., Jungblut P.R. (2013) *Mol. Cell. Proteomics*, **2**, 529-538.
- Deracinois B., Flahaut C., Duban-Deweere S., Karamanos Y. (2013) *Proteomes*, **1**, 180-218.
- UniProt Consortium (2008) *Nucleic Acids Res.*, **36**, 190-195.
- Richardson M.R., Liu S., Ringham H.N., Chan V., Witzmann F.A. (2008) *Electrophoresis*, **12**, 2637-2644.
- Tonella L., Walsh B.J., Sanchez J.C., Ou K., Wilkins M.R., Tyler M., Frutiger S., Gooley A.A., Pescaru I., Appel R.D. et al. (1998) *Electrophoresis*, **11**, 1960-1971.
- Alberts B., Bray D., Lewis J., Raff M., Roberts K., Watson J. (1994) *Mol. Biol. Cell* (3rd edn.), NYC: Garland Publishing.
- Lodish H., Berk A., Zipursky S.L., Matsudaira P., Baltimore D., Darnell J. (2000) *Molecular Cell Biology* (4th edn.), NYC: W.H. Freeman and Company.
- Lim H., Yates III, J.R. (2001) *Encyclopedia of Life Sciences*. www.els.net.
- Janke C., Holzer M., Klose J., Arendt T. (1996) *FEBS Lett.*, **379**, 222-226.
- Milo R., Jorgensen P., Moran U., Weber G., Springer M. (2010) *Nucleic Acids Res.*, **38**, 750-753.
- Anderson N.L., Anderson N.G. (2002) *Mol. Cell Proteomics*, **11**, 845-867.
- Hortin G.L., Sviridov D. (2010) *J. Proteomics*, **73**, 629-636.

Поступила: 16. 02. 2015.

EXPERIMENTAL ESTIMATION OF PROTEOME SIZE FOR CELLS AND HUMAN PLASMA

*S.N. Naryzhny^{1,2}, V.G. Zgoda¹, M.A. Maynskova¹, N.L. Ronzhina¹,
N.V. Belyakova², O.K. Legina², A.I. Archakov¹*

¹Institute of Biomedical Chemistry,

10 Pogodinskaya str., Moscow, 119121 Russia; e-mail: snaryzhny@mail.ru

²Konstantinov Petersburg Nuclear Physics Institute, Gatchina, Leningrad District, Russia

Huge range of concentrations of different protein and insufficient sensitivity of methods for detection of proteins at a single molecule level does not yet allow obtaining the whole image of human proteome. In our investigations, we tried to evaluate the size of different proteomes (cells and plasma). The approach used is based on detection of protein spots in 2-DE after staining by protein dyes with different sensitivities. The function representing the dependence of the number of protein spots on sensitivity of protein dyes was generated. Next, by extrapolation of this function curve to theoretical point of the maximum sensitivity (detection of a single smallest polypeptide) it was calculated that a single human cell (HepG2) may contain minimum 70000 proteoforms, and plasma – 1.5 mln. Utilization of this approach to other, smaller proteomes showed the competency of this extrapolation. For instance, the size of mycoplasma (*Acholeplasma laidlawii*) was estimated in 1100 proteoforms, yeast (*Saccharomyces cerevisiae*) - 40000, *E. coli* – 6200, *P. furiosus* – 3400. In hepatocytes, the amount of proteoforms was the same as in HepG2 – 70000. Significance of obtained data is in possibilities to estimating the proteome organization and planning next steps in its study.

Key words: proteome, proteoforms, 2DE.