

Dwuwymiarowa elektroforeza żelowa: od eksperymentu po profile ekspresji. Część druga – analiza obrazu

Two-dimensional gel electrophoresis: from experiment to protein expression profiles. Part two – image analysis

Agnieszka Suchwałko^{1,2}, Halina Podbielska¹

¹ Instytut Inżynierii Biomedycznej i Pomiarowej, Wydział Podstawowych Problemów Techniki, Politechnika Wroclawska, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, tel. +48 (71) 320 28 25, e-mail: agnieszka.suchwalko@pwr.wroc.pl

² MedicWave AB, Stalverksgatan 1, SE-302 45 Halmstad, Sweden

Streszczenie

Dwuwymiarowa (dwukierunkowa) elektroforeza żelowa (2-DE) jest metodą znaną od lat siedemdziesiątych poprzedniego stulecia. Zainteresowanie badaczy metodą 2-DE wynika z możliwości separacji nawet kilku tysięcy białek w jednym żelu, co pozwala na ich detekcję i identyfikację. Dzięki wysokiej rozdzielczości wyników separacji możliwe staje się ustalenie roli biologicznej poszczególnych białek czy odkrywanie wpływu czynników zewnętrznych na organizmy żywe na poziomie proteomu. Metoda 2-DE wciąż ewoluje. W ostatniej dekadzie nastąpił ogromny postęp w sposobie przeprowadzania eksperymentów z dużą powtarzalnością. Zmieniło się też całkowicie podejście do analizy obrazów żeli i wykorzystywane do tego celu algorytmy. Ze względu na tak szybki rozwój technik dwuwymiarowej elektroforezy żelowej, autorzy postanowili przedstawić obecny stan wiedzy w tym zakresie. Praca składa się z dwu części. W pierwszej opisano zmiany, jakie zaszły w technikach przeprowadzania eksperymentów elektroforetycznych. Druga część skupia się na analizie obrazów uzyskanych za pomocą metody 2-DE, z uwzględnieniem zmian w schemacie analizy.

Druga część pracy to opis dwóch schematów analizy: klasycznego i obecnie stosowanego. Szczegółowo opracowane zostały metody analizy pojedynczego obrazu 2-DE oraz analizy porównawczej serii obrazów. Przedstawione są algorytmy stosowane do normalizacji, segmentacji obrazu, detekcji plam, dopasowywania do siebie obrazów czy tworzenia map proteomowych jako najważniejsze w całej analizie. Opisano również generowanie profili ekspresji, metody identyfikacji protein oraz internetowe bazy danych 2-DE. Autorzy nie pominęli tak ważnego zagadnienia, jak sposoby porównywania systemów analizujących dane 2-DE. Ta część pracy podsumowuje rozwój, jaki nastąpił w ostatnim dziesięcioleciu w metodach stosowanych w analizie komputerowej obrazów 2-DE, którego najważniejszym krokiem było opracowanie nowego schematu analizy, opartego na tworzeniu map proteomowych.

Słowa kluczowe: elektroforeza żelowa, 2-DE, analiza porównawcza obrazów, schemat analizy 2-DE, mapa proteomowa, segmentacja obrazu, detekcja plam, identyfikacja protein

Abstract

Two-dimensional gel electrophoresis (2-DE) is a method commonly used since seventies of the previous century. It

owns its non weakening interest of researchers because of the possibility of separation even a few thousands of proteins in one gel, what allows for protein detection and identification. Thanks to high resolution of separation results, it is possible to determine a biological role of particular proteins or to discover influence of external factors on living organisms on the proteome level. 2-DE is still evolving. In the last decade, a great progress has been made in the way of performing experiments and their reproducibility. Approach to the analysis of gel images and algorithms used for the analysis have been entirely changed. For the reason of such instant changes of methods and techniques of the two-dimensional gel electrophoresis, the authors decided to describe state of the art. This work consists of two parts. The first part describes changes that have been made in performing of electrophoretic experiments. The second part concentrates on the analysis of obtained images, with emphasis the analysis workflow.

Second part is a description of two analysis workflows: the classical one and the currently used one. Analysis methods of the single 2-DE image and differential analysis of the image series are discussed in detail. Algorithms used for normalization, image segmentation, spot detection, image warping and creation of proteome maps are introduced as the most crucial in the whole analysis. Generation of expression profiles, protein identification methods, and the Internet databases of the 2-DE data are described. Authors have not missed as important issue as methods of comparison of systems for 2-DE data analysis. This part of the work summarises development that has been made for the last ten years in the methods used for the computer analysis of 2-DE images. The most important step of this development was elaboration of the workflow based on creation of proteome maps.

Key words: gel electrophoresis, 2-DE, differential image analysis, 2-DE analysis workflow, proteome map, image segmentation, spot detection, protein identification

Wstęp

Dwuwymiarowa (dwukierunkowa) elektroforeza żelowa (2-DE) jest metodą znaną od lat 70. ubiegłego stulecia [1]. Podstawową zasadą działania elektroforezy jest migracja naładowanych elektrycznie cząstek w polu elektrycznym. Jej zadaniem jest separacja cząsteczek badanej próbki, a efektem płytki pokryta żelem, zawierająca rozdzielone cząstki. Płytką żelowa może zostać zarejestrowana za pomocą odpowiedniego

skanera lub kamery jako obraz cyfrowy, możliwy do dalszej analizy za pomocą dedykowanych do tego celu programów komputerowych. Natomiast rozdzielone cząstki, mogą zostać wycięte z żelu (ręcznie lub z użyciem robota) i odpowiednio przygotowane, aby poddać je dalszej analizie (np. spektrometrii mas).

Najbardziej typowym zastosowaniem elektroforezy w proteomice jest separacja białek w zależności od pH oraz masy cząsteczkowej (M_r). Separacja taka jest możliwa, ponieważ proteiny i peptydy są substancjami amfoterycznymi i w związku z tym mogą zostać naładowane dodatnio albo ujemnie, w zależności od pH środowiska, w którym się znajdują. Dzięki temu mogą się przemieszczać w polu elektrycznym.

Ze względu na dużą liczbę trudnych danych, jakie niesie ze sobą każdy zarejestrowany cyfrowo obraz żelu pochodzący z eksperymentu elektroforetycznego, analiza ilościowa obrazów żeli uwzględniająca przetwarzanie obrazu, modelowanie zjawisk zachodzących w żelach czy estymację statystyczną zaburzonych danych stała się przedmiotem zainteresowań badaczy różnych dziedzin.

Na analizę danych pochodzących z eksperymentu 2-DE składa się wiele zagadnień. Surowe dane (niepoddane żadnemu przetwarzaniu) należy poddać obróbce wstępnej, która poprawi ich jakość (zmniejszy szumy itp.). Następnie należy wykonać rejestrację obrazu, czyli dopasować obrazy do teoretycznie uzyskanych danych na temat migracji cząstek i wykonać wszystkie możliwe czynności, które pozwolą na porównywanie obrazów między sobą. Na ten etap składa się usunięcie tła, normalizacja, dopasowanie przekształceniem afinicznym do wyników teoretycznych (jeśli są znane), dopasowanie obrazów między sobą (warping) i inne metody zależne od danych. Wykorzystanie zupełnie innych algorytmów jest niezbędne do stworzenia mapy proteomowej danego eksperymentu. Detekcja plam proteinowych to krok, na który składają się zwykle techniki segmentacji obrazu i modelowania znalezionych obiektów. Ostatni etap to wygenerowanie profili ekspresji. Te wszystkie kroki dzielą się na mniejsze: detekcja plam to również wyodrębnianie plam z kompleksów, a oprócz generowania profili ekspresji można zidentyfikować modyfikacje posttranslacyjne. Proces analizy danych 2-DE jest zatem bardzo skomplikowany i wymaga połączenia wiedzy z różnych dziedzin.

Analiza obrazów

Podejścia stosowane w analizie obrazów

Jednym z typowych zadań proteomiki jest poszukiwanie różnic w ekspresji protein pomiędzy próbką kontrolną (np. osoby zdrowe) a badaną (np. osoby chore) na serii obrazów żeli z eksperymentu 2-DE. Oznacza to poszukiwanie plam, które zniknęły, pojawiły się bądź zmieniły swój rozmiar lub ekspresję. Porównywanie obrazów żeli między sobą wyłącznie za pomocą ludzkiego wzroku wydaje się niemożliwe, jednak przez wiele lat badacze byli zmuszeni do analizowania wyników właśnie w taki sposób. Stosunkowo szybko powstały pierwsze programy analizujące obrazy żeli. Ich początkowe możliwości były bardzo niewielkie, głównie ze względu na ograniczone możliwości sprzętowe. Dzisiejsze oprogramowanie analizujące nie przypomina już pierwszych programów służących do tego celu.

Początki analizowania obrazów żeli to programy, które nie były nawet w stanie wyświetlić obrazów, ze względu na brak wystarczających zasobów sprzętowych. Dostępne opcje pozwalały wyznaczyć charakterystyki ilościowe wyłącznie dobrze wyeksponowanych plam, a i to z dużym błędem. Sama technika elektroforezy również była daleka od dzisiejszej. Tło, artefakty czy szumy były nieodłącznymi składnikami analizy i nie było możliwości usunięcia ich w żaden sposób. Całość przetwarzania komputerowego odbywała się w trybie

tekstowym, a komunikacja z użytkownikiem miała miejsce wyłącznie za pomocą konsoli. Było to równoznaczne z ograniczoną dostępnością do oprogramowania – komputery nie były wówczas powszechnie osiągalne, a korzystanie z takiego programu było skomplikowane.

Od tych czasów dokonał się ogromny postęp. Twórcy oprogramowania nastawieni są na tworzenie systemów przyjaznych dla użytkownika, dających dużą swobodę wyboru opcji analizy czy wreszcie mających możliwość wygodnej i wysokiej jakości wizualizacji danych. Na proste w obsłudze oprogramowanie składają się zaawansowane algorytmy, których zadaniem jest w jak najkrótszym czasie uzyskanie z danych wszystkich możliwych informacji. Dużą wagę przykładają się do optymalizacji obliczeń. Chodzi tu zarówno o czas wykonywania operacji, zajętość pamięci przez proces analizy, jak i przeprowadzenie analizy z możliwie małym udziałem człowieka [2].

Zmieniony został zupełnie schemat analizy. Do niedawna najpierw wykonywana była detekcja plam na każdym z obrazów z eksperymentu, a następnie próbowano parować odpowiadające sobie plamy pomiędzy poszczególnymi obrazami i na tej podstawie porównywać ekspresję interesujących badacza protein [3, 4].

Klasyczny schemat analizy [3]:

1. obróbka wstępna obrazów: normalizacja, obcinanie i usunięcie tła, artefaktów i szumów;
2. detekcja plam: segmentacja obrazu, wyznaczanie granic plam, charakterystyka ilościowa znalezionych plam proteinowych;
3. parowanie plam na obrazach przez użytkownika, dopasowane pary są wykorzystywane jako podstawa do wyznaczenia przekształcenia dopasowującego pary obrazów do siebie;
4. automatyczne dopasowanie do siebie pozostałych znalezionych plam;
5. wygenerowanie profili ekspresji protein dla każdego z obrazów z eksperymentu;
6. wizualizacja danych i ich interpretacja;
7. tworzenie baz danych elektroforetycznych.

Takie podejście powodowało niekompletne profile ekspresji i problemy z dopasowaniem do siebie plam o słabej ekspresji, z różnych obrazów, z jednego eksperymentu. Dzisiejsze oprogramowanie najpierw próbuje dopasować do siebie wszystkie obrazy (przekształcenie dopasowujące, warping), wykonać fuzję obrazów do mapy proteomowej, później następuje detekcja plam na mapie proteomowej i powrót do miejsc, w których powinny znajdować się plamy na poszczególnych obrazach (odwrotne przekształcenie dopasowujące, odwrotny warping). Dzięki takiemu podejściu mamy pewność, że wszystkie plamy występujące w eksperymencie zostaną znalezione i scharakteryzowane na wszystkich żelach z eksperymentu. Schemat, wykonujący analizę w takiej kolejności, bywa nazywany schematem czwartej generacji i tak będzie określany w dalszej części pracy.

Analiza obrazów za pomocą schematu czwartej generacji

Analiza obrazów żeli opiera się zawsze na pewnym schemacie opisującym kolejne etapy od momentu wczytania danych cyfrowych obrazu do programu, aż do uzyskania kompletnych profili ekspresji, dla każdego z białek występujących w eksperymencie. Obecnie uznawany za najlepszy schemat analizy obrazów pochodzących z dwuwymiarowej elektroforezy żelowej składa się z następujących kroków:

1. obróbka wstępna obrazów (dekompozycja obrazu)
 - wyodrębnienie i usunięcie tła, artefaktów i szumów,
 - wygenerowanie przybliżonych obrazów o lepszej widoczności;

2. rejestracja obrazu
 - wyznaczenie transformacji afinicznych dopasowujących do teoretycznie wyznaczonych pI (punkt izoelektryczny) i M_r (masa cząsteczkowa),
 - wyznaczenie transformacji dopasowujących i odwrotnych transformacji dopasowujących do siebie obrazy (warping i warping odwrotny),
 - normalizacja poziomu natężenia pomiędzy obrazami;
3. fuzja obrazów do mapy proteomowej, wyznaczenie standardowych pozycji plam;
4. detekcja i modelowanie plam na mapie proteomowej;
5. detekcja i modelowanie plam na poszczególnych obrazach oraz ich analiza ilościowa zakończona wygenerowaniem profilu ekspresji.

Porównanie klasycznego schematu z powyżej zaprezentowanym schematem w kontekście zadania poszukiwania biomarkerów proteinowych można znaleźć w pracy [5].

Analiza pojedynczego obrazu

Zwykle do analizy wykorzystuje się całe serie obrazów powstałych w wyniku przeprowadzonego eksperymentu 2-DE [Suchwałko 2010]. Analiza serii obrazów żeli opiera się na niemal identycznych krokach powtarzanych dla każdego obrazu. Wyjątkiem są algorytmy wykorzystujące wiedzę zawartą we wszystkich obrazach z danego eksperymentu, ale tych jest wciąż stosunkowo niewiele. Istnieją również sytuacje, gdy obiektem zainteresowania jest analiza pojedynczego obrazu (singletonu). Efektem poddania pojedynczego obrazu obróbce jest usunięcie powstałych podczas eksperymentu błędów, poprawna segmentacja, detekcja i modelowanie plam, wreszcie możliwie precyzyjne ustalenie pI i M_r protein oraz poprawne scharakteryzowanie ich ekspresji. Typowe etapy analizy pojedynczego obrazu to:

1. obróbka wstępna: usunięcie smug lub szumów, korekcja tła i normalizacja natężenia;
2. wyznaczenie przekształcenia afinicznego dopasowującego do teoretycznie otrzymanych pI i M_r ;
3. segmentacja i modelowanie plam;
4. analiza ilościowa każdej z plam.

Obróbka wstępna

Ważnym etapem analizy pojedynczego obrazu jest korekcja przestrzenna. Bierze ona pod uwagę między innymi wycieki żelu, będące wynikiem zmian w globalnym polu elektrycznym podczas przeprowadzania eksperymentu elektroforezy. Zastosowanie korekcji znacznie upraszcza końcowy pomiar pI i M_r . Wyznaczenie przekształcenia, które pozwoli dopasować odkształcony obraz żelu do rzeczywistego położenia, wymaga rozwiązania dość skomplikowanych równań [6, 7].

Obrazy żeli wymagają specjalnego traktowania ze względu na podatność na przyleganie kurzu do powierzchni żelu. Problem ten ujawnia się po zapisie żelu do postaci cyfrowej. Jeszcze przed przystąpieniem do samej analizy okazuje się, że obraz jest zniekształcony. Zaburzenia zostały opisane jako szum o rozkładzie normalnym. Do ich usunięcia stosuje się filtry: lokalny gaussowski, rozpraszający, wielomianowy czy też lokalny medianowy [8]. Każdy z nich charakteryzuje się nieco innymi właściwościami, ale wszystkie wystarczająco dobrze radzą sobie z usuwaniem zaburzeń o rozkładzie normalnym. Obraz po przefiltrowaniu nie jest dokładnie taki, jakim była płytka żelowa przed zapisem. Do uzyskania obrazów możliwie wiernie odwzorowujących żel można zastosować modele probabilistyczne, jednak jest to proces czasochłonny ze względu na swoją złożoność obliczeniową.

Do wygładzenia obrazu stosuje się inne techniki przetwarzania obrazu, takie jak np. rozciąganie histogramu (transformacja liniowa, umożliwiającą wykorzystanie całego dostępnego

zakresu poziomów szarości). Należy jednak zawsze pamiętać, że wszelkie przekształcenia wykonywane na obrazie (z założenia nie dotyczy to filtrów) muszą być odwracalne i nie mogą w znaczący sposób zmienić entropii (miara informacji niesionej przez obraz) obrazu. Popularnymi technikami, których powinno się unikać, są między innymi: wyrównanie histogramu (rozciąganie histogramu jest przekształceniem, mającym całkiem odmienny charakter), poprawianie kontrastu czy korekcja gamma. Te transformacje powodują, że różne wartości natężenia mogą zostać przekształcone na jedną wartość i tracona jest bezpowrotnie informacja [9]. Takie przekształcenia można stosować, jeśli zmiany dotyczą kopii obrazów, na których nie ma być prowadzona dalsza analiza, zwłaszcza charakteryzacja plam.

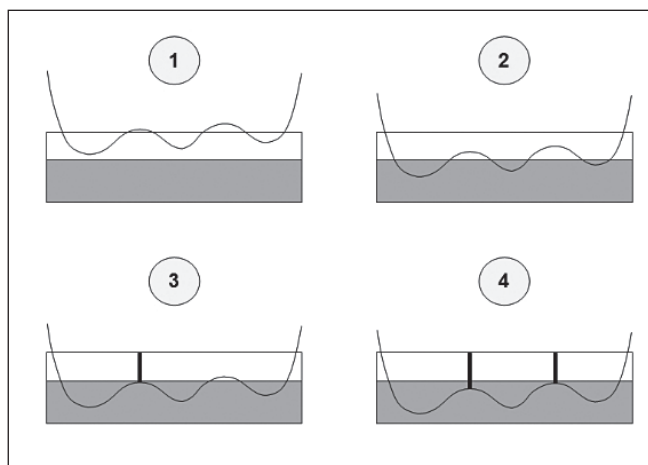
Przed analizowaniem informacji niesionej przez obraz żelu należy usunąć jego tło, jako nieistotny dla analizy element obrazu. Idealne jest białe tło, czyli sytuacja, gdy każdy punkt wchodzący w skład tła ma maksymalną wartość natężenia. Tło powstaje na skutek przyłączenia barwnika do substancji niebędących proteinami, których nie udało się usunąć (np. składników żelu czy płytki, na której umieszczony jest żel). Ponieważ zawsze tło jest widoczne, zatem niezbędne jest zastosowanie złożonych algorytmów, aby poprawnie i bez utraty istotnych danych usunąć tło. Należy zwrócić uwagę na fakt, że tło znajduje się również w miejscach, w których są plamy. Prosty sposób jest wykorzystanie algorytmu opartego na znalezieniu minimów lokalnych tła [10]. Znaną metodą jest również wyznaczenie wielkości natężenia tła dla każdej z plam przez obliczenie wartości, która reprezentuje 10% wszystkich natężeń znajdujących się na granicy plamy [8]. Inny sposób to dopasowanie wielomianu do znanych natężeń tła i estymacja wartości, które znajdują się pod plamami [11]. Istnieje również algorytm, który wyznacza wartości tła, „tocząc” kulę pod powierzchnią trójwymiarowych płaszczyzn reprezentujących żele. Kula musi być odpowiednio duża, żeby nie zagłębić się zbyt mocno w plamach [12]. Uproszczona wersja algorytmu, stosująca obiekt cylindryczny w jednym z wymiarów, jest przydatna do usuwania smug w tym wymiarze [13]. Taki cylinder zagłębia się w smugi, ale nie jest w stanie zagłębić się w plamy, dzięki czemu można stworzyć obraz smug, a następnie odjąć go od obrazu plam.

Zostały zidentyfikowane cztery przyczyny powstawania zaburzeń: struktura żelu; charakterystyka detergentu, który zapewnia transport, warunki panujące w buforze oraz natura pola elektrycznego [11].

Detekcja plam proteinowych – segmentacja obrazu oraz modelowanie plam

Detekcja plam ma na celu wyodrębnienie wszystkich plam proteinowych znajdujących się na obrazie. Na końcu tego etapu znane są charakterystyki znalezionych protein w postaci współrzędnych środków plam, natężeń i właściwości geometrycznych. Nie jest to proste zadanie, ponieważ proteiny bardzo różnią się między sobą. Bywają one trudne do odróżnienia od artefaktów i szumów, często są przysłonięte smugami lub nachodzą na siebie. Mocno nasycone klastry protein nie mają widocznych granic pomiędzy białkami [14, 15]. Rozwiązanie tego problemu wiąże się po raz kolejny z zastosowaniem zaawansowanych algorytmów obróbki obrazu. Typowy schemat detekcji plam składa się z następujących kroków [11]:

1. znalezienie jak największej liczby punktów (często środków) należących do różnych plam;
2. segmentacja obrazu na części, z których każda zawiera jedną z plam, bądź klastr proteinowy (na podstawie punktów wyznaczonych w poprzednim kroku);
3. modelowanie każdego z segmentów w taki sposób, aby wyodrębnić z niego dane do dalszej analizy, bądź kolejne podregiony, które zawierają plamy.



Rys. 1 Schemat działania algorytmu działów wodnych (WST). Działanie algorytmu WST (watershed segmentation) można sobie wyobrazić jako stopniowe zanurzanie w wodzie płaszczyzny reprezentującej żel, którą nakłuto w minimach lokalnych (punkty znajdujące się najgłębiej – najciemniejsze na obrazie). Najpierw woda przedostaje się przez nakłucia i stopniowo wypełnia całą płaszczyznę. W miejscach gdzie łączy się woda, która dostała się przez dwa różne minima lokalne, stawiana jest tama (dział wodny)

Początkowo detekcja plam była realizowana przez filtr LoG [16-18]. Próbowano również stosować pierścienie, które wskazywały środek elips po wstępnym podziale obrazu na plamy i tło z użyciem wartości granicznej [2]. Zupełnie inne podejście to podział obrazu na segmenty przy założeniu, że plamy należą do wypukłych obszarów obrazu, które wyznaczane były za pomocą maski $n \times n$ [3]. Stosowano również operacje morfologiczne, na przykład serie filtrów o zróżnicowanych rozmiarach masek do zidentyfikowania plam o różnych wielkościach [4]. Jednak żadna z wymienionych metod nie cieszy się takim powodzeniem jak segmentacja obrazu metodą działów wodnych (watershed segmentation – WST) [5].

Terminologia dotycząca zagadnienia segmentacji metodą działów wodnych została zapożyczona z geografii. Dział wodny należy rozumieć jako granicę pomiędzy regionami terenu (zlewniami), w których zbiera się woda. Woda przekraczająca dział wodny zaczyna przelewać się do sąsiedniej zlewni. Idea algorytmu bazuje na analogii żelu do rzeźby terenu. Zagłębieniami są plamy, im mniejsza wartość natężenia danego punktu, tym znajduje się on głębiej poniżej powierzchni. Powierzchnią jest wartość bieli, czyli 2^{16} dla obrazów 16-bitowych. Metoda wyznaczająca działy wodne etykietuje punkty, w taki sposób, że każda zlewnia ma niepowtarzalną, w obrębie danego obrazu, etykietę. Dodatkowo wprowadzana jest etykieta W, która oznacza działy wodne. Dzięki temu każdy punkt obrazu należy albo do zlewni, albo do działu wodnego. Klasycznym algorytmem jest ten zaproponowany przez Vincenta i Soille'a w pracy [19]. W pierwszym kroku znajdowane są lokalne minima, czyli najciemniejsze punkty obrazu. W następnej części algorytmu – obrazowo mówiąc – nakłuwamy minima lokalne, a całą powierzchnię stopniowo zanurzamy w wodzie. Miejsca, gdzie woda, która dostała się przez dwa różne minima, łączy się, tworząc jeden basen, oznaczane są jako działy wodne. Intuicyjny schemat algorytmu działów wodnych przedstawia rys. 1. W konsekwencji działania algorytmu działów wodnych otrzymujemy podział obrazu żelu na segmenty, z których każdy zawiera jakąś plamę lub klastr proteinowy. Metoda jest wrażliwa na szumy i artefakty, należy ją więc stosować po wcześniejszym usunięciu z obrazu zbędnych obiektów. Więcej informacji o różnych algorytmach działów wodnych (również o ich wersjach równoległych) można znaleźć w pracach [5, 7-11, 20].

Inne metody nie bazują na tak intuicyjnej idei i opierają się bardziej na koncepcjach matematycznych, takich jak na przykład markowskie pola losowe [12], oparte na Laplasjanie. Algorytm wykorzystuje całą dostępną informację z eksperymentu do modelowania szumu, dzięki czemu nawet plamy o słabej ekspresji zostają znalezione [13].

Najczęściej spotykane w oprogramowaniu algorytmy, których zadaniem jest określenie, które punkty obrazu tworzą plamę proteinową, opierają się na dopasowaniu dwuwymiarowego rozkładu normalnego oraz modelowaniu iteracyjnym [21]. Najlepsze wyniki widać się dawać hybryda segmentacji i modelowania.

Detekcja plam w kompleksach proteinowych

Białka w żelu często tworzą kompleksy (klastry, skupienia), czyli formacje składające się z co najmniej kilku plam proteinowych nachodzących na siebie bądź zlewających się. W takich rejonach informacja o natężeniu nie jest bardzo przydatna [22], gdyż jest ono bardzo zbliżone co do wartości na obszarze tworzącym kompleks i mocno kontrastuje z rejonem otaczającym kompleks. Algorytm zaproponowany w pracy [22] rozcina plamy, znajdując punkty siodłowe rozważanej powierzchni. Inna praca sprowadza problem rozdzielania kompleksów do pokrycia obszaru elipsami [16]. Ze względu na generowanie przez metodę wielu zbędnych i nieoptymalnych wyników ta sama praca [16] proponuje rozwiązanie bazujące na programowaniu liniowym, upraszczające i poprawiające końcowy wynik.

Analiza ilościowa plam

Po znalezieniu plam i rozdzielaniu ich kompleksów pozostało jeszcze zadanie charakteryzacji ilościowej wszystkich białek. Typowe wyznaczane charakterystyki to [18, 23]:

- powierzchnia plamy: liczba pikseli powierzchnia piksela;
- gęstość optyczna plamy: $\max_{(x,y) \in \text{spot}} I(x,y)$ – maksymalne natężenie plamy;
- całkowita gęstość optyczna plamy: $\sum_{(x,y) \in \text{spot}} I(x,y)$ – suma wszystkich natężeń.

Inne charakterystyki można wyznaczyć na podstawie parametrów modeli opisujących plamy proteinowe [17].

Ekspresja białek jest zwykle opisywana za pomocą gęstości optycznej plam oraz całkowitej gęstości optycznej plam, które są unormowane dla całego obrazu. Całkowita gęstość optyczna lepiej opisuje ekspresję białek [24].

Analiza porównawcza obrazów żeli

Analiza porównawcza obrazów żeli pochodzących z eksperymentów 2-DE to zastosowanie dostępnych technik, w celu wydobycia informacji, które pozwolą na zidentyfikowanie protein, różniących się ekspresją w zależności od badanego czynnika. Analiza porównawcza związana jest również z pozyskiwaniem biomarkerów [5]. Takie podejście pozwala na zastosowanie technik nieopierających się na wcześniej postawionych hipotezach i może wykorzystywać algorytmy oparte na statystyce [25, 26] czy pozyskiwaniu danych (data mining). Zastosowanie aparatu statystycznego wymaga przygotowania danych w taki sposób, aby spełniały założenia postawione konkretnym metodom. Uzyskanie informacji z danych jest możliwe wyłącznie przez szczegółową analizę natężeń, normalizację, odpowiednią transformację danych, wyznaczenie wariancji oraz wykorzystanie odpowiednich metod statystycznych. Wszystkie te operacje wykonywane są w odpowiedniej kolejności, w poszczególnych etapach analizy.

Eksploracyjna analiza znormalizowanych natężeń plam wykazała silną zależność pomiędzy wariancją a średnią między proteinami [27]. Wariancja danych pochodzących z żeli ma

wiele źródeł: biologiczne rozbieżności w poziomie protein pomiędzy próbkami, błędy podczas ekstrakcji białek, barwienie żeli, proces elektroforezy oraz błędy w analizie ilościowej. W pracy [27] zaproponowano metody, które sprawdziły się w modelowaniu podobnych zaburzeń dla mikromacierzy. Efektem było usunięcie korelacji przez zastosowanie przesuniętej transformacji logarytmicznej przy jednoczesnym zachowaniu heterogeniczności wariancji dla różnych protein.

Ważnym zagadnieniem jest przygotowanie eksperymentu. Żeby w pełni skorzystać z narzędzi matematycznych, statystycznych czy analizy danych, należy odpowiednio przygotować eksperyment. Najlepszym rozwiązaniem jest przeprowadzenie eksperymentu po konsultacji z osobą, która miałaby dokonać analizy uzyskanych danych, jeśli ma być ona inna niż standardowa (oferowana przez oprogramowanie do analizy obrazów żeli). W taki sposób łatwiej jest ustalić, ile próbek należy przebadać, żeby wyniki były wiarygodne ze statystycznego punktu widzenia. Ma to znaczenie na przykład dla wyznaczenia biologicznej wariancji badanych próbek [28]. Naturalnie im większa liczba danych, tym więcej można znaleźć i zidentyfikować protein, które będą mogły być wzięte pod uwagę w analizie. Białko, które ujawniło się raz bądź dwa razy, nie może być brane pod uwagę, jeśli eksperyment składał się z 20 żeli. Informacje na temat przygotowywania eksperymentu pod kontem późniejszej analizy są opisane w pracach [25, 29].

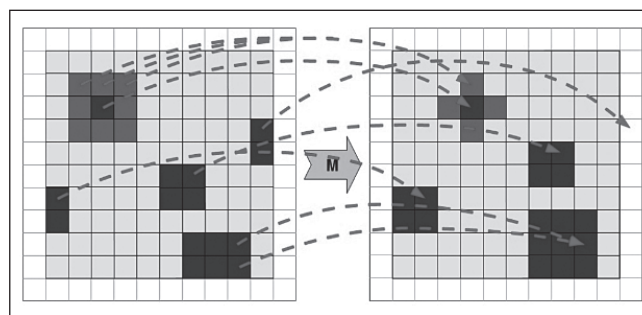
Obróbka wstępna

Ten etap analizy obejmuje operacje podobne do obróbki wstępnej dla pojedynczego obrazu 2-DE. Zawiera się w nim usunięcie szumów, tła, artefaktów czy rozciągnięcie histogramu. Należy też wykonać, dla każdego obrazu z eksperymentu, dopasowanie analogiczne do opisanego w części o analizie singletonu obrazu 2-DE. Chodzi o możliwie dobre przybliżenie teoretycznie wyznaczonych pI i M_r . Jest to ważne przy późniejszym wyznaczeniu pozycji standardowych protein. Dopasowanie takie określone jest przekształceniem afinicznym (korygującym przesunięcia i obroty), które w żaden sposób nie zmienia informacji niesionej przez obraz i jest odwracalne.

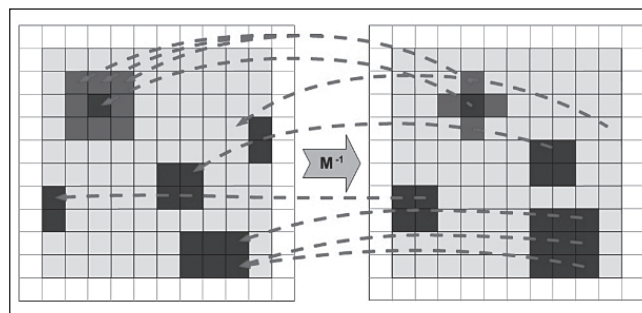
Rejestracja obrazu

Rejestracją obrazu w analizie porównawczej obrazów pochodzących z dwuwymiarowej elektroforezy żelowej nazywa się wszystkie operacje, które należy wykonać, aby móc dokonać analizy porównawczej dwóch obrazów. Do takich zadań należą między innymi normalizacja natężenia czy dopasowanie geometryczne obrazów. Za niepożądane można uznać zaburzenia natężenia, wynikające z procesu zapisu obrazu żelu do postaci cyfrowej, czy zniekształcenia geometryczne, będące następstwem samego eksperymentu. Usuwając niechciane modyfikacje, równocześnie należy zachować faktyczne różnice między obrazami niosące informację o zmianie ekspresji białek, ich pH czy masy cząsteczkowej.

Problem definiuje krotka (I_s, I_r, M, sim) , gdzie I_s to obraz źródłowy (oryginalny), I_r to obraz referencyjny, M oznacza odwzorowanie obrazu oryginalnego w obraz referencyjny, a sim jest funkcją opisującą podobieństwo między obrazami [30]. Wartość sim jest maksymalna, gdy dopasowanie jest najlepsze (optymalne). Oznacza to, że najlepsze dopasowanie znajduje się poprzez maksymalizację funkcji podobieństwa obrazów. Żeby to osiągnąć, szuka się transformacji w postaci funkcji mapującej, która odwzorowuje każdy piksel należący do przestrzeni obrazu źródłowego w jakiś piksel należący do przestrzeni obrazu referencyjnego. Jednak wykonując to zadanie wprost, okaże się, że część pikseli na obrazie wynikowym I (obraz powstały z I_s po zastosowaniu przekształcenia)



Rys. 2 Przykład działania przekształcenia M . Przekształcenie działa z obrazu źródłowego (I_s , po lewej stronie) w obraz referencyjny (I_r , po prawej stronie). Oznacza to, że obraz wynikowy (I) będzie przekształceniem obrazu I_s , w taki sposób, aby odpowiadające sobie charakterystyczne punkty znajdowały się w tych samych miejscach. Na przykładzie widać, że przekształcenie M może odwzorować kilka punktów w jeden, a inne odwzorować poza obszarem obrazu



Rys. 3 Przykład działania przekształcenia M^{-1} . Przekształcenie działa z obrazu referencyjnego (I_r , po prawej stronie) w obraz źródłowy (I_s , po lewej stronie). Grot strzałek pokazują, skąd brane są wartości natężeń dla punktów znajdujących się na początku strzałek. Jeśli grot strzałki znalazłby się poza obrazem, przypisana zostałaby maksymalna dostępna wartość natężenia bądź średnia natężeń najbliższych znajdujących się punktów

nie będzie wypełniona, ponieważ transformacja nie jest 1:1 i kilka punktów może zostać odwzorowanych w jeden, a inne w żaden. Co więcej, może zdarzyć się, że punkty po transformacji znajdują się poza obrazem. Schemat działania przekształcenia M pokazuje rys. 2.

Rozwiązaniem problemu jest wykorzystanie odwzorowania odwrotnego do M , oznaczanego w dalszej części pracy przez M^{-1} . Wówczas analizuje się kolejno punkty obrazu wynikowego I i na podstawie M^{-1} ustala się punkt z I_s , który go zastąpi. W ten sposób cały obraz wynikowy jest wypełniony, co nie oznacza, że wykorzystane zostały wszystkie punkty z I_s . Schemat działania przekształcenia M^{-1} pokazuje rys. 3. Przekształcenie, które pozwala dopasować jeden obraz do drugiego, jest zagadnieniem pochodzącym z grafiki komputerowej i nazywa się z angielskiego warpingiem bądź przekształceniem warpującym. Więcej o transformacjach dopasowujących obrazy można znaleźć w pracy [31]. Algorytmy warpingu są obecnie na bardzo wysokim poziomie i pozwalają na bardzo dobre dopasowanie obrazów. Wiadomo, że algorytmy stosujące jedną funkcję do dopasowania obrazów są za słabe, ze względu na nierówne odkształcenia w różnych częściach obrazu. Znacznie lepiej sprawdzają się metody hierarchiczne, takie jak HGT (*hierarchical grid transform*) opisany w pracy [32]. W większości systemów niezbędne jest wskazanie przez użytkownika co najmniej kilku punktów, na podstawie których znaleziona zostanie transformacja dla całego obrazu. Powstają jednak pierwsze programy oferujące automatyczne dopasowanie par obrazów.

Istnieją sytuacje, w których nie da się dopasować żadnego przekształcenia, na przykład jeśli proteina, na skutek badane-



go czynnika, zmieniła swoją masę na tyle mocno, że przesuwała się na obrazie w inne miejsce, „mijając” po drodze inne białka. Jeśli nie da się ustalić transformacji, należy zastosować inną technikę, na przykład DIGE (*difference gel analysis*). Różnorodne algorytmy warpingu, począwszy od klasycznych liniowych czy wielomianowych, przez hierarchiczne, aż po równoległe, są opisane między innymi w pracach [19, 32-34]. Porównania różnych metod dopasowujących obrazy znajdują się w pracach [32, 35].

W przypadku analizy serii obrazów ważna jest ich normalizacja. Natężenie plam na dwóch obrazach może się znacząco różnić, chociaż faktyczne natężenie jest niemal równe. Normalizacja pozwala usunąć różnice między obrazami, które powstały na skutek różnej objętości próbek ładowanych do IEF, niezgodności w transferze protein pomiędzy pierwszym a drugim wymiarem eksperymentu (znikanie protein między wymiarami) czy też zróżnicowanego czasu ekspozycji żelu podczas zapisu obrazu. Normalizacją nazywany jest w tym przypadku proces, w wyniku którego wartości natężenia opisujące poszczególne plamy proteinowe będą mogły być porównywane pomiędzy różnymi obrazami z eksperymentu. Normalizację, wykonuje się zwykle w oparciu o założenie o jednakowej ilości białek na każdym z obrazów. Na tej podstawie wyznacza się proporcjonalną wartość natężenia dla każdej plamy w obrębie obrazu. Inne podejście mówi, że natężenie obrazu często jest zdeterminowane przez kilka protein o bardzo silnej ekspresji, w związku z czym należy usunąć z procesu normalizacji plamy mające największe wartości natężenia (usunięcie obserwacji odstających).

Istnieją też podejścia, sugerujące, że rozkłady natężeń między obrazami są takie same. Na tej podstawie próbuje się dopasowywać do nich na przykład rozkład normalny. Jeszcze inna hipoteza mówi o tym, że tylko część plam należących do obrazu powinna być brana pod uwagę podczas normalizacji [21]. Metody początkowo opracowane dla mikromacierzy z powodzeniem znalazły zastosowanie również na tym etapie analizy [22].

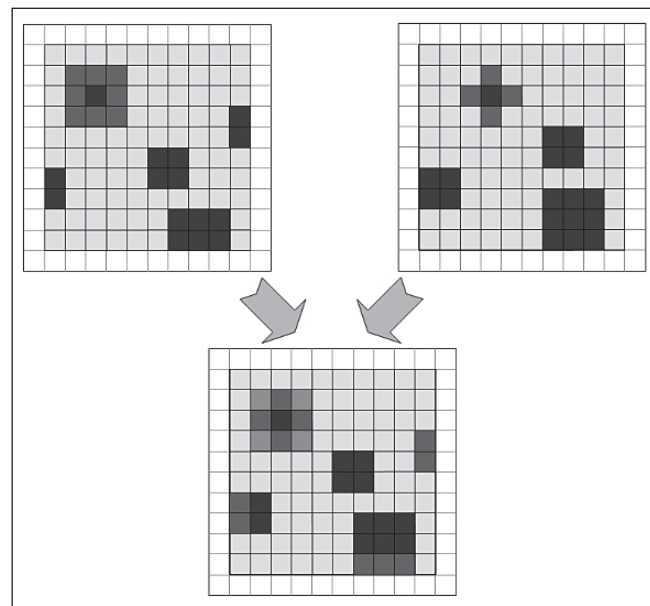
Inaczej przedstawia się problem normalizacji dla techniki DIGE. W samym eksperymencie została uwzględniona próbka z tak zwanym standardem, zawierająca mieszaninę wszystkich próbek z eksperymentu w jednakowych ilościach. Normalizacja polega na podzieleniu natężeń plamy przez odpowiadające mu natężenie plamy z obrazu standardowego.

Jak było wspomniane w pierwszej części pracy dotyczącej eksperymentu [36], próbka może zostać podzielona na kilka części, każda o węższym gradiencie pH niż całość próbki (np. 3-7 pH oraz 6-11 pH). Po zapisie obrazu żeli do postaci cyfrowej mamy do czynienia z kilkoma obrazami dla jednej próbki, które wymagają scalenia. Ma to miejsce właśnie na etapie rejestracji obrazu [37]. Łączenie obrazów odbywa się zwykle z udziałem człowieka, którego zadaniem jest wskazanie odpowiadających sobie plam proteinowych na zachodzących na siebie pod względem pH krańcach obrazów. Na tej podstawie program scala obrazy, tworząc jeden obraz dla jednej próbki.

Zagadnienia dotyczące rejestracji obrazu, rejestracji obrazu biomedycznego czy wreszcie miar podobieństwa między obrazami są opisane szczegółowo w pracach [38-47] oraz w książkach [48, 49].

Mapa proteomowa

Poprawne wyznaczenie przekształcenia dopasowującego obrazy (dla każdej pary obrazów jest to inne przekształcenie) jest kluczowym zadaniem tej części analizy. Dzięki znalezieniu optymalnych dopasowań można w prosty sposób utworzyć mapę proteomową. Mapa proteomowa jest fuzją wszystkich obrazów, jakie zostały otrzymane w wyniku przeprowadzenia eksperymentu 2-DE [50]. Powstaje ona przez prze-



Rys. 4 Uproszczony przykład tworzenia mapy proteomowej. Na podstawie informacji o położeniu, wielkości i natężeniu punktów na dwóch obrazach jest tworzona mapa. Mapa odzwierciedla relatywne położenia punktów ich natężenia oraz wielkość

kształcenie wartości natężenia poszczególnych pikseli ze wszystkich obrazów przez pewną funkcję, w odpowiadające im piksele na mapie. Jako wymagania dla określenia standardowej pozycji plam, podane zostały cztery warunki [50]:

- standardowa pozycja proteiny powinna znajdować się możliwie blisko jej relatywnych położen na wszystkich żelach z eksperymentu, po uwzględnieniu przesunięć i obrotów powstałych podczas zapisu obrazu;
- przestrzenne relacje pomiędzy plamami powinny zostać zachowane na tyle dobrze, na ile to możliwe;
- powinno być możliwe znalezienie odpowiadających sobie plam na zestawie żeli w oparciu o ich pozycje standardowe;
- pozycja standardowa powinna pozwalać na predykcję protein jeszcze niezidentyfikowanych, na podstawie ich pozycji.

Standardowe pozycje plam generowane są w dwóch krokach. Pierwszy krok to wyznaczenie standardowego położenia kilku ustalonych plam (tych, które posłużyły do wyznaczenia przekształceń dopasowujących), przez uśrednienie współrzędnych, dla których wcześniej została zastosowana transformacja afiniczna, redukująca przesunięcia, skalowania i obroty, różniące teoretycznie wyznaczone pI i M_r , od faktycznie uzyskanych podczas eksperymentu. Drugi krok to wygenerowanie przybliżonych pozycji pozostałych plam na podstawie punktów należących do zbioru znanych już plam. Funkcją zaproponowaną w pracy [50] jest średnia ważona, która ciemnym pikselom przypisuje dużą wagę, a jasnym małą. Efektem zastosowania takiego przekształcenia jest dobra widoczność plam, które nie występują na wszystkich obrazach z eksperymentu, oraz tych, które są słabo widoczne, ale obecne na większości obrazów. Wygenerowana w taki sposób mapa, na pierwszy rzut oka, wygląda jak jeden z obrazów z eksperymentu. Zasadnicza różnica polega na tym, że na mapie znajdują się wszystkie plamy, które ujawniły się w eksperymencie. Można łatwo zaobserwować plamy proteinowe, które wystąpiły tylko na części obrazów lub mają bardzo słabą ekspresję i przez to na części obrazów zlewały się z sąsiadującymi, silniejszymi proteinami. Bardzo uproszczony przykład tworzenia mapy proteomowej pokazuje rys. 4.



Detekcja i dopasowywanie plam proteinowych

Detekcja plam proteinowych odbywa się najpierw na mapie proteomowej, a następnie na każdym z obrazów z osobna. Należy pamiętać, że obrazy po transformacji dopasowującej zmieniły się i analiza ilościowa nie jest na nich możliwa. Analizę wykonuje się na pierwotnych obrazach, na podstawie danych uzyskanych z mapy proteomowej.

Pierwsze podejście to dzielenie mapy na rozłączne segmenty za pomocą jednego z algorytmów działów wodnych. Następnie w każdym z segmentów wyszukiwana jest plama. Zgodnie z założeniem algorytmu działów wodnych wszystkie segmenty zawierają minima lokalne. Od minimów lokalnych rozpoczyna się poszukiwanie plamy. Ostatecznie każdy piksel obrazu albo należy do jakiejś plamy, albo należy do tła obrazu. Algorytmy detekcji plam mogą uwzględniać różne charakterystyki obrazu: natężenie, nachylenie czy klasyfikacja punktu na podstawie jego otoczenia [21]. Programy zwykle pozwalają na modyfikację wygenerowanych granic plam, pozwalając w ten sposób na łączenie, dzielenie czy zmianę kształtu znalezionych obiektów. Przy tych operacjach warto zachować dużą ostrożność, gdyż indywidualne podejście może mieć wpływ na późniejszą powtarzalność wyników czy profile ekspresji.

Innym podejściem do detekcji plam proteinowych jest próba modelowania plam z wykorzystaniem dwuwymiarowego rozkładu normalnego. Jest to sposób na wyznaczenie granic plam białkowych. Takie podejście ogranicza jednak możliwości modelowania kształtu plam. Opracowane zostały również algorytmy hybrydowe, które zaczynają od wyznaczenia segmentów, a następnie na podstawie rozkładu normalnego wyznaczają granice plam [21].

Po znalezieniu i zweryfikowaniu przez człowieka granic plam na mapie program za pomocą wcześniej wyznaczonego przekształcenia M (żeby dopasować obrazy użyte zostało przekształcenie M^{-1}) transformuje znalezione plamy na pierwotne obrazy. Następnie sprawdza, czy i w tych miejscach występują plamy. Jeśli występują, to automatycznie są one kojarzone z plamami z mapy i odpowiadającymi sobie plamami z pozostałych obrazów. Ta część analizy, to dopasowywanie do siebie plam (*spot matching*). Naturalnie plamy różnią się kształtem i rozmiarem między obrazami, dlatego niezbędny jest krok, który pozwoli na wyznaczenie granic danej plamy na każdym z obrazów z osobna. Plamy zazwyczaj występują na wszystkich obrazach z eksperymentu, ale różnią się pod względem ekspresji i – co bardzo istotne – położeniem na obrazie. Sporadycznie zdarza się, że pewna proteina występuje tylko na części obrazów. Dzięki fuzji obrazów do mapy proteomowej kroki: detekcji plam na poszczególnych obrazach oraz dopasowywania do siebie plam stały się znacznie prostsze, zajmują mniej czasu i są bardziej wiarygodne. Najważniejszym efektem tworzenia map proteomowych jest tworzenie kompletnych profili ekspresji, bez obserwacji brakujących. Plamy, które nie zostały znalezione na części obrazów, zostaną scharakteryzowane, a ich natężenia będą bliskie zeru [51].

Analiza ilościowa oraz profile ekspresji

Jako wynik etapu detekcji plam otrzymuje się informacje o każdej ze znalezionych plam białkowych. Informacje te nazywane są profilami ekspresji protein. Znane są: znormalizowane natężenie, pozycja środka, zarys, tło i inne charakterystyki. Nie można jednak porównać tych danych bezpośrednio z zawartością protein w zbadanej próbce. Na różnice pomiędzy wynikami analizy a pierwotną próbką znaczny wpływ mają: utrata części próbki podczas wprowadzania do IEF (obrazowanie izoelektryczne, pierwszy wymiar 2-DE), efektywność transferu między wymiarami, utrata protein podczas wybarwienia, efektywność wybarwienia, krzywa wybarwienia

protein względem czasu, krzywa wybarwienia względem koncentracji protein oraz blaknięcie barwników [21]. Wszystkie informacje dotyczące natężeń analizowanych danych, należy traktować jako ilościowe, a nie jakościowe. Oznacza to, że można porównywać ekspresję pewnej proteiny pomiędzy poszczególnymi obrazami z eksperymentu, na przykład ustalając jej natężenie w jednej z próbek kontrolnych i stwierdzając, że w warunkach eksperymentu zmieniła swoją ekspresję dwukrotnie w porównaniu z próbką kontrolną. Nie dotyczy to pI oraz M_r , które mogą być dość dokładnie estymowane w oparciu o położenie plamy w żelu i znane zależności wpływu barwnika.

Na podstawie wygenerowanych profili ekspresji można ustalić, które proteiny zareagowały na czynnik będący obiektem badania oraz jak intensywna była reakcja. To pozwala stawiać hipotezy na temat funkcji biologicznej białek. Często stosuje się technikę 2-DE jako narzędzie do poszukiwania hipotez, a następnie inne metody, które pozwalają na weryfikację znalezionych zależności. Dobrym przykładem jest tu szukanie biomarkerów, czyli protein, które są nierozłącznie związane z fenotypem. Do poszukiwań stosuje się szeroką gamę metod – od standardowych testów statystycznych po uczenie maszynowe [21].

Charakter profili ekspresji jest zbliżony do charakteru danych mikromacierzowych. Analiza profili ekspresji wykorzystuje (z pewnymi modyfikacjami) metody z powodzeniem stosowane w analizie danych mikromacierzowych. Profile można analizować w pewnym zakresie wewnątrz programów do analizy obrazów, jednak nie zapewniają one zbyt szerokiej gamy metod. Znacznie większe możliwości ma analiza za pomocą innych uniwersalnych narzędzi, specjalizujących się w metodach statystycznej analizy danych oraz pozyskiwania wiedzy z danych. Takie uniwersalne narzędzia zwykle nie oferują prostej dla użytkownika, wizualnej reprezentacji znalezionych informacji. Taką funkcjonalność mają wyłącznie systemy do analizy obrazów 2-DE. Bardziej szczegółowy opis programów i oferowanych przez nie możliwości znajduje się w pracy [21].

Sposoby porównywania systemów do analizy obrazów żeli

Wszystkie dostępne systemy, oprócz samej analizy obrazów, oferują podstawowe funkcje, takie jak: import i eksport danych w standardowych formatach graficznych (tiff, gif, jpeg, png itp.), opisywanie i poprawianie kształtu plam, porównywanie obrazów i samych plam, przeszukiwanie listy plam czy wreszcie połączenie z internetowymi bazami danych [30].

Nie istnieją prace porównujące wszystkie dostępne systemy do analizy obrazów 2-DE, ponieważ jest ich dużo i są stosunkowo drogie. Zwykle porównanie dotyczy dwóch, sporadycznie kilku programów [26, 52-56]. Obecnie na rynku znajduje się spora grupa programów do analizy obrazów żeli. Wśród nich wciąż są takie, które wywodzą się od pierwszych, akademickich systemów. Obecnie dostępne na rynku programy to: SameSpots Progenesis (Nonlinear Dynamics), ImageMaster 2D (AP Biotech), DeCyder™ (tylko DIGE; GE Healthcare), Delta2D (DECODON), REDFIN (LUDESI), PDQuest (BioRad), GELLAB II (Scanalytics, Fairfax), Melanie 7.0 (GeneBio). Istnieje również niewielka liczba darmowych pakietów do analizy obrazów żeli, jak na przykład digeR, program napisany w języku R, z graficznym interfejsem użytkownika [57].

Porównanie działania oprogramowania do analizy obrazów żeli opiera się na specjalnie do tego przygotowanych testach [26, 58]. Testy badają nie tylko precyzję i niezawodność, ale też powtarzalność wyników analizy między poszczególnymi programami. Można znaleźć informacje po-

równawcze na temat poszczególnych etapów analizy, takich jak detekcja plam, wrażliwość na nachodzenie na siebie protein, różnice w analizie ilościowej plam czy też odróżnianie białek od artefaktów [59]. Badania takie przeprowadza się zarówno na obrazach symulowanych [59, 60], o znanych właściwościach, takich jak liczba plam lub ich wielkości, oraz na obrazach rzeczywistych. Do testowania tworzone są również odpowiednie modele i systemy uczące się na prawdziwych obrazach [60].

Bazy danych 2-DE

Po uzyskaniu listy znalezionych i scharakteryzowanych cząsteczek możliwe jest porównanie ich z wynikami zawartymi w bazach danych opublikowanych w internecie. Serwer, który zawiera najwięcej danych pochodzących z dwuwymiarowej elektroforezy żelowej to ExPASy (*Expert Protein Analysis System*). Znajduje się on pod adresem <http://www.expasy.org/>. Znaleźć tam można dane pochodzące z różnych eksperymentów, zawierające informacje o funkcji protein (o ile jest ona znana), szczegółach przygotowania eksperymentu oraz jego wynikach. Wszystkie zestawy opatrzone są informacjami o autorze, referencjami do prac, które na ich podstawie powstały, opisem metod identyfikacji, danymi o zakresie pI i innymi szczegółami pozwalającymi na porównanie z posiadanymi danymi. Takie dane pomagają na przykład w weryfikacji prowadzonych eksperymentów, mogą też posłużyć jako dane testowe dla nowych metod analizy. Waga istnienia takich baz danych jest nie do przecenienia. W jednym miejscu można znaleźć informacje o tym, jakie proteiny zostały już zidentyfikowane, jakie prowadzono eksperymenty i jakie funkcje poszczególnych białek są znane. Pozwala to stawiać nowe hipotezy i weryfikować już istniejące. Można postarać się o umieszczenie wyników własnych eksperymentów i analiz na serwerach ExPASy.

Identyfikacja protein

Identyfikacja protein na podstawie mas peptydów zaobserwowanych w eksperymencie spektrometrem mas (*peptide mass fingerprinting* – PMF) to podstawowa metoda identyfikacji stosowana po eksperymentach techniką 2-DE. Ze względu na swoją dokładność często stosuje się też chromatografię cieczową, połączoną ze spektrometrią mas (LC-MS/MS). PMF jest jednak zazwyczaj pierwszym narzędziem stosowanym do identyfikacji, głównie dlatego, że metoda jest szybka i przyjazna dla użytkownika. Dostarcza ona charakterystycznych dla każdej proteiny danych, które są następnie porównywane z teoretycznymi oraz z danymi zawartymi w bazach danych.

Metody identyfikacji nie stanowią dobrej podstawy do pracy z próbkami pochodzącymi z organizmów, których genomy jeszcze nie zostały w pełni opisane. Problemem jest też identyfikacja białek, które uległy modyfikacjom posttranslacyjnym, gdyż nie dają się one porównać z niezmienionymi danymi teoretycznymi [61]. Gdy PMF nie wystarcza, można ustalić sekwencje aminokwasów za pomocą MALDI-MS (*matrix-assisted laser desorption/ionization mass spectrometry*) lub MS/MS (wielokrotna spektrometria mas). Największą wadą MS/MS jest czas niezbędny do przeprowadzenia całego procesu identyfikacji. Błędy w identyfikacji są obiektem zainteresowania badaczy [62], a metody identyfikujące są wciąż rozwijane i coraz bardziej dokładne.

Podsumowanie

Proteomika od lat rozwija się bardzo intensywnie [63]. Analiza proteomu jako ostatniego poziomu ekspresji genów rozpoczęła się właśnie dzięki metodzie 2-DE [64]. Mimo po-

wstania wielu nowych technik i podejść dwuwymiarowa elektroforeza żelowa jest jednak wciąż jedną z niezastąpionych technik proteomiki i jako taka wymaga ogromnych nakładów pracy, aby otrzymywane wyniki były jeszcze bardziej precyzyjne, a analiza przebiegała prościej i szybciej. Ostatnia dekada pozwoliła na wprowadzenie wielu udogodnień dla badaczy posługujących się tą metodą.

Zaliczają się do nich: opracowanie nowego schematu analizy danych 2-DE, analiza danych otrzymanych metodą DIGE, zaawansowane algorytmy znacznie upraszczające porównanie danych pochodzących z różnych próbek, nowoczesne techniki wizualizacji itd. Dzięki tworzeniu map proteomowych analiza jest bardziej precyzyjna i wiarygodna, a zastosowanie zaawansowanych algorytmów pozwoliło znacząco skrócić czas jej trwania. Wciąż brakuje pełnej automatyzacji zarówno eksperymentu, jak i działania oprogramowania. Sytuacją w pełni satysfakcjonującą będzie posiadanie systemów analizujących, których pracę człowiek będzie musiał jedynie sprawdzać, ale nie będzie musiał w niej uczestniczyć. ■

Literatura

1. T. Rabilloud: *Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains*, Proteomics, vol. 2, 2002, s. 3-10.
2. J. Fiévet, C. Dillmann, G. Lagniel, M. Davanture, L. Negroni, J. Labarre, D. de Vienne: *Assessing factors for reliable quantitative proteomics based on two-dimensional gel electrophoresis*, Proteomics, vol. 4, 2004, s. 1939-1949.
3. K. Pleissner, H. Oswald, S. Wegener: *Proteomics: from protein sequence to function*, Garland Science, 2001.
4. K.P. Pleißner, F. Hoffmann, K. Kriegel, C. Wenk, S. Wegner, A. Sahlström, H. Oswald, H. Alt, E. Fleck: *New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis gel databases*, Electrophoresis, vol. 20, 1999, s. 755-765.
5. J. E. Bandow, J.D. Baker, M. Berth, C. Painter, O.J. Sepulveda, K.A. Clark, I. Kilty, R.A. VanBogelen: *Improved image analysis workflow for 2-D gels enables large-scale 2-D gel-based proteomics studies—COPD biomarker discovery study*, Proteomics, vol. 8, 2008, s. 3030-3041.
6. J.S. Gustafsson, A. Blomberg, M. Rudemo: *Warping two-dimensional electrophoresis gel images to correct for geometric distortions of the spot pattern*, Electrophoresis, vol. 23, 2002, s. 1731-1744.
7. J.S. Gustafsson, R. Ceasar, C.A. Glasbey, A. Blomberg, M. Rudemo: *Statistical exploration of variation in quantitative two-dimensional gel electrophoresis data*, Proteomics, vol. 4, 2004, s. 3791-3799.
8. R.D. Appel, J. Vargas, P.M. Palagi, D. Walther, D.F. Hochstrasser: *Melanie II: a third generation software package for analysis of two-dimensional electrophoresis images-II Algorithms*, Electrophoresis, vol. 18, 1997, s. 2735-2748.
9. R. Rangayyan: *Biomedical image analysis*, CRC, 2005.
10. J.J. Tyson R.H. Haralick: *Computer analysis of two-dimensional gels by a general image processing system*, Electro-phoresis, vol. 7(3), 1986, s. 107-113.
11. A.W. Dowsey, M.J. Dunn, G.Z. Yang: *The role of bioinformatics in two-dimensional gel electrophoresis*, Proteomics, vol. 3, 2003, s. 1567-1596.
12. S.R. Sternberg: *Grayscale morphology*, Comput. Vision Graph. Image Process, vol. 35(3), 1986, s. 333-355.
13. M.M. Skolnick: *Application of morphological transformations to the analysis of two-dimensional electrophoretic gels of biological materials*, vol. 35(3), 1986, s. 306-332.
14. M.C. Pietrogrande, N. Marchetti, F. Dondi, P.G. Righetti: *Spot overlapping in two-dimensional polyacrylamide gel electrophoresis maps: relevance to proteomics*, Electrophoresis, vol. 24, 2003, s. 217-224.
15. M.C. Pietrogrande, N. Marchetti, F. Dondi, P.G. Righetti: *Spot overlapping in two-dimensional polyacrylamide gel electrophoresis separations: a statistical study of complex protein maps*, Electrophoresis, vol. 23, 2002, s. 283-291.
16. R.D. Appel, P.M. Palagi, D. Walther, J.R. Vargas, J.C. Sanchez, F. Ravier, C. Pasquali, D.F. Hochstrasser: *Melanie II—a third-generation software package for analysis of two-dimensional electrophoresis images: I. Features and user interface*, Electrophoresis, vol. 18, 1997, s. 2724-2734.
17. N.L. Anderson, J. Taylor, A.E. Scandora, B.P. Coulter, N.G. Anderson: *The TYCHO system for computer analysis of two-dimensio-*

- nal gel electrophoresis patterns, *Clinical chemistry*, vol. 27, 1981, s. 1807-1820.
18. A.D. Olson M.J. Miller: *Elsie 4: quantitative computer analysis of sets of two-dimensional gel electrophoretograms*, *Analytical biochemistry*, vol. 169, 1988, s. 49-70.
 19. L. Vincent, P. Soille: *Watersheds in digital spaces: an efficient algorithm based on immersion simulations*, *IEEE transactions on pattern analysis and machine intelligence*, vol. 13, 1991, s. 583-598.
 20. N.H. Packer, M.S. Ball, P.L. Devine: *Glycoprotein detection of 2-D separated proteins*, *Methods in molecular biology* (Clifton, N.J.), vol. 112, 1999, s. 341-352.
 21. M. Berth, F.M. Moser, M. Kolbe, J. Bernhardt: *The state of the art in the analysis of two-dimensional gel electrophoresis images*, *Applied microbiology and biotechnology*, vol. 76, 2007, s. 1223-1243.
 22. D.P. Kreil, N.A. Karp, K.S. Lilley: *DNA microarray normalization methods can remove bias from differential protein expression analysis of 2D difference gel electrophoresis results*, *Bioinformatics* (Oxford, England), vol. 20, 2004, s. 2026-2034.
 23. H.Y. Tan, T.W. Ng, O.W. Liew: *Effects of light spectrum in flatbed scanner densitometry of stained polyacrylamide gels*, *BioTechniques*, vol. 42, 2007, s. 474-478.
 24. K. Takahashi, M. Nakazawa, Y. Watanabe, A. Konagaya: *Fully-automated spot recognition and matching algorithms for 2-D gel electrophoretogram of genomic DNA*, *Genome informatics, Workshop on Genome Informatics*, vol. 9, 1998, s. 161-172.
 25. S.M.N. Hunt, M.R. Thomas, L.T. Sebastian, S.K. Pedersen, R.L. Harcourt, A.J. Sloane, M.R. Wilkins: *Optimal replication and the importance of experimental design for gel-based quantitative proteomics*, *Journal of proteome research*, vol. 4(3), 2005, s. 809-819.
 26. N.A. Karp, D.P. Kreil, K.S. Lilley: *Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis*, *Proteomics*, vol. 4, 2004, s. 1421-1432.
 27. K. Löster C. Kannicht: *2D-electrophoresis. Detection of glycosylation and influence on spot pattern*, *Methods in molecular biology* (Clifton, N.J.), vol. 194, 2002, s. 301-316.
 28. S. Luhn, M. Berth, M. Hecker, J. Bernhardt: *Using standard positions and image fusion to create proteome maps from collections of two-dimensional gel electrophoresis images*, *Proteomics*, vol. 3, 2003, s. 1117-1127.
 29. G. Van den Bergh, L. Arckens: *Protein profiling based on two-dimensional difference gel electrophoresis*, *Methods in molecular biology* (Clifton, N.J.), vol. 439, 2008, s. 211-224.
 30. M.D. Miller, R.A. Acey, L.Y.T. Lee, A.J. Edwards: *Digital imaging considerations for gel electrophoresis analysis systems*, *Electrophoresis*, vol. 22, 2001, s. 791-800.
 31. J. Prehm, P. Jungblut, J. Klose: *Analysis of two-dimensional electrophoretic protein patterns using a video camera and a computer. II. Adaptation of automatic spot detection to visual evaluation*, *Electrophoresis*, vol. 8(12), 1987, s. 562-572.
 32. K. Conradsen, J. Pedersen: *Analysis of two-dimensional electrophoretic gels*, *Biometrics*, vol. 48, 1992, s. 1273.
 33. A. Paradelo, J.P. Albar: *Advances in the analysis of protein phosphorylation*, *Journal of proteome research*, vol. 7, 2008, s. 1809-1818.
 34. J. Roerdink, A. Meijster: *The watershed transform: definitions, algorithms and parallelization strategies*, *Mathematical Morphology*, vol. 41, 2001, s. 187-228.
 35. J. Smółka: *Watershed based region growing algorithm*, *Annales Informatica UMCS, Lublin*, vol. 3, 2005, s. 169.
 36. A. Suchwałko, H. Podbielska: *Dwuwymiarowa elektroforeza żelowa: od eksperymentu po profile ekspresji. Część pierwsza – eksperyment*, *Inżynieria Biomedyczna – Acta Bio-Optica et Informatica Medica*, vol. 16, 2010, s. 285-292.
 37. J.A. Westbrook, J.X. Yan, R. Wait, S.Y. Welson, M.J. Dunn: *Zooming-in on the proteome: very narrow-range immobilised pH gradients reveal more protein species and isoforms*, *Electrophoresis*, vol. 22, 2001, s. 2865-2871.
 38. A. Moga, B. Cramariuc, M. Gabbouj: *An efficient watershed segmentation algorithm suitable for parallel implementation*, *IEEE Comput. Soc. Press*.
 39. W.F. Patton: *Detection technologies in proteome analysis*, *Journal of Chromatography. B, Analytical technologies in the biomedical and life sciences*, vol. 771, 2002, s. 3-31.
 40. T. Rabilloud: *Detecting proteins separated by 2-D gel electrophoresis*, *Analytical Chemistry*, vol. 72, 2000, s. 48A-55A.
 41. W. Weiss, F. Weiland, A. Görg: *Protein detection and quantitation technologies for gel-based proteome analysis*, *Methods in molecular biology* (Clifton, N.J.), vol. 564, 2009, s. 59-82.
 42. J. De Bock, P. De Smet, W. Philips: *A fast sequential rainfalling watershed segmentation algorithm*, *Advanced Concepts for Intelligent Vision Systems*, Springer, 2005, s. 476-482.
 43. A. Meijster, J. Roerdink: *A proposal for the implementation of a parallel watershed algorithm*, *Computer Analysis of Images and Patterns*, Springer, s. 790-795.
 44. S.Z. Li: *Markov random field modelling of fMRI data using a mean field EM-algorithm4*, *Methods in Computer Vision and Pattern*, 1999, s. 317-330.
 45. M.S. Baker: *Automatic registration and segmentation algorithm for multiple electrophoresis images*, *SPIE*, 2000.
 46. A. Efrat, F. Hoffmann, K. Kriegel, C. Schultz, C. Wenk: *Geometric algorithms for the analysis of 2D-electrophoresis gels*, *Journal of Computational Biology*, vol. 9(2), 2002, s. 299-315.
 47. P.F. Lemkin, J.E. Myrick, K.M. Upton: *Splitting merged spots in two-dimensional polyacrylamide gel electrophoresis gel images*, *Applied and theoretical electrophoresis: the official journal of the International Electrophoresis Society*, vol. 3, 1993, s. 163-172.
 48. K. Miura: *Imaging technologies for the detection of multiple stains in proteomics*, *Proteomics*, vol. 3, 2003, s. 1097-1108.
 49. P.F. Lemkin, L.E. Lipkin: *GELLAB: A computer system for 2D gel electrophoresis analysis. II. Pairing spots*, *Computers and biomedical research, an international journal*, vol. 14, 1981, s. 355-380.
 50. A.T. Rosengren, J.M. Salmi, T. Aittokallio, J. Westerholm, R. Lahesmaa, T.A. Nyman, O.S. Nevalainen: *Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels*, *Proteomics*, vol. 3, 2003, s. 1936-1946.
 51. M. Eravci, S. Fuxius, O. Broedel, S. Weist, S. Eravci, U. Mansmann, H. Schluter, J. Tiemann, A. Baumgartner: *Improved comparative proteome analysis based on two-dimensional gel electrophoresis*, *Proteomics*, vol. 7, 2007, s. 513-523.
 52. J.I. Garrels: *The QUEST system for quantitative analysis of two-dimensional gels*, *The Journal of Biological chemistry*, vol. 264, 1989, s. 5269-5282.
 53. M.J. Dutt, K.H. Lee: *The scaled volume as an image analysis variable for detecting changes in protein expression levels by silver stain*, *Electrophoresis*, vol. 22, 2001, s. 1627-1632.
 54. C. Glasbey, K.V. Mardia: *A review of image-warping methods*, *Journal of Applied Statistics*, vol. 25, 1998, s. 155-171.
 55. J. Salmi, T. Aittokallio, J. Westerholm, M. Griese, A. Rosengren, T.A. Nyman, R. Lahesmaa, O. Nevalainen: *Hierarchical grid transformation for image warping in the analysis of two-dimensional electrophoresis gels*, *Proteomics*, vol. 2, 2002, s. 1504-1515.
 56. T. Aittokallio, J. Salmi, T.A. Nyman, O.S. Nevalainen: *Geometrical distortions in two-dimensional gels: applicable correction methods*, *Journal of Chromatography. B, Analytical technologies in the biomedical and life sciences*, vol. 815, 2005, s. 25-37.
 57. K. Kaczmarek, B. Walczak, S. De Jong, B. Vandeginste: *Comparison of image-transformation methods used in matching 2D gel electrophoresis images*, *Acta Chromatographica*, no. 13, 2003, s. 7-21.
 58. A. Görg, W. Postel, A. Domscheit, S. Günther: *Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (Hordeum vulgare): method, reproducibility and genetic aspects*, *Electrophoresis*, vol. 9, 1988, s. 681-692.
 59. T. Voss, P. Haberl: *Observations on the reproducibility and matching efficiency of two-dimensional electrophoresis gels: consequences for comprehensive data analysis*, *Electrophoresis*, vol. 21, 2000, s. 3345-3350.
 60. G. Horgan, A. Creasey, B. Fenton: *Superimposing two-dimensional gels to study genetic variation in malaria parasites*, *Electrophoresis*, vol. 13(1), 1992, s. 871-875.
 61. P.R. Graves, T.A.J. Haystead: *Molecular biologist's guide to proteomics*, *Microbiology and molecular biology reviews: MMBR*, vol. 66, 2002, s. 39-63.
 62. R.A. Bradshaw, A.L. Burlingame, S. Carr, R. Aebersold: *Protein identification: the good, the bad, and the ugly*, *Molecular & cellular proteomics: MCP*, vol. 4, 2005, s. 1221-1222.
 63. M.R. Wilkins, R.D. Appel, J.E. Van Eyk, M.C.M. Chung, A. Görg, M. Hecker, L.A. Huber, H. Langen, A.J. Link, Y.-K. Paik, S.D. Patterson, S.R. Pennington, T. Rabilloud, R. J. Simpson, W. Weiss, M.J. Dunn: *Guidelines for the next 10 years of proteomics*, *Proteomics*, vol. 6, 2006, s. 4-8.
 64. P.H. O'Farrell: *High resolution two-dimensional electrophoresis of proteins*, *The Journal of Biological Chemistry*, vol. 250, 1975, s. 4007-4021.

otrzymano / received: 03.05.2010 r.
zaakceptowano / accepted: 14.07.2010 r.