

Sylwia GÓRCZYŃSKA-KOSIORZ

Śląski Uniwersytet Medyczny, Katedra i Klinika Chorób Wewnętrznych Diabetologii i Nefrologii

Bożena MAŁYSIAK-MROZEK, Dariusz MROZEK

Politechnika Śląska, Instytut Informatyki

SCENARIUSZE UŻYCIA SYSTEMU MAS4PSI W DIAGNOSTYCE MEDYCZNEJ

Streszczenie. System MAS4PSi (Multi Agent System For Protein Similarity searching) pozwala na szybkie, skalowalne i niezawodne poszukiwanie podobieństwa strukturalnego białek. Poszukiwanie podobieństwa strukturalnego białek jest kluczowe w prowadzeniu badań nad różnymi procesami biologicznymi i innymi obszarami, które mają swoją podstawę w tych procesach biologicznych. Celem niniejszego artykułu jest przybliżenie i określenie możliwych scenariuszy wykorzystania zbudowanego systemu MAS4PSi w powszechnie rozumianej diagnostyce medycznej zarówno na etapie opracowywania eksperymentów pomocnych we wprowadzeniu nowych badań, jak i na etapie typowo diagnostycznym.

Słowa kluczowe: bioinformatyka, białka, struktura, podobieństwo strukturalne, systemy wieloagentowe, obliczenia równoległe

SCENARIOS OF USING MAS4PSI SYSTEM IN MEDICAL DIAGNOSTICS

Summary. MAS4PSi (Multi-Agent System For Protein Similarity searching) is a system that allows fast, scalable and reliable protein structure similarity searching. Protein structure similarity searching is crucial in conducting research on a variety of biological processes, and other areas that have their basis in these biological processes. The purpose of this article is to define and present possible scenarios of using the MAS4PSi system in a broadly understood medical diagnostics, both in the design of experiments supporting new research, as well as the typical diagnostic stage.

Keywords: structural bioinformatics, protein structure, similarity, parallel computing, multi-agent systems

1. Wprowadzenie

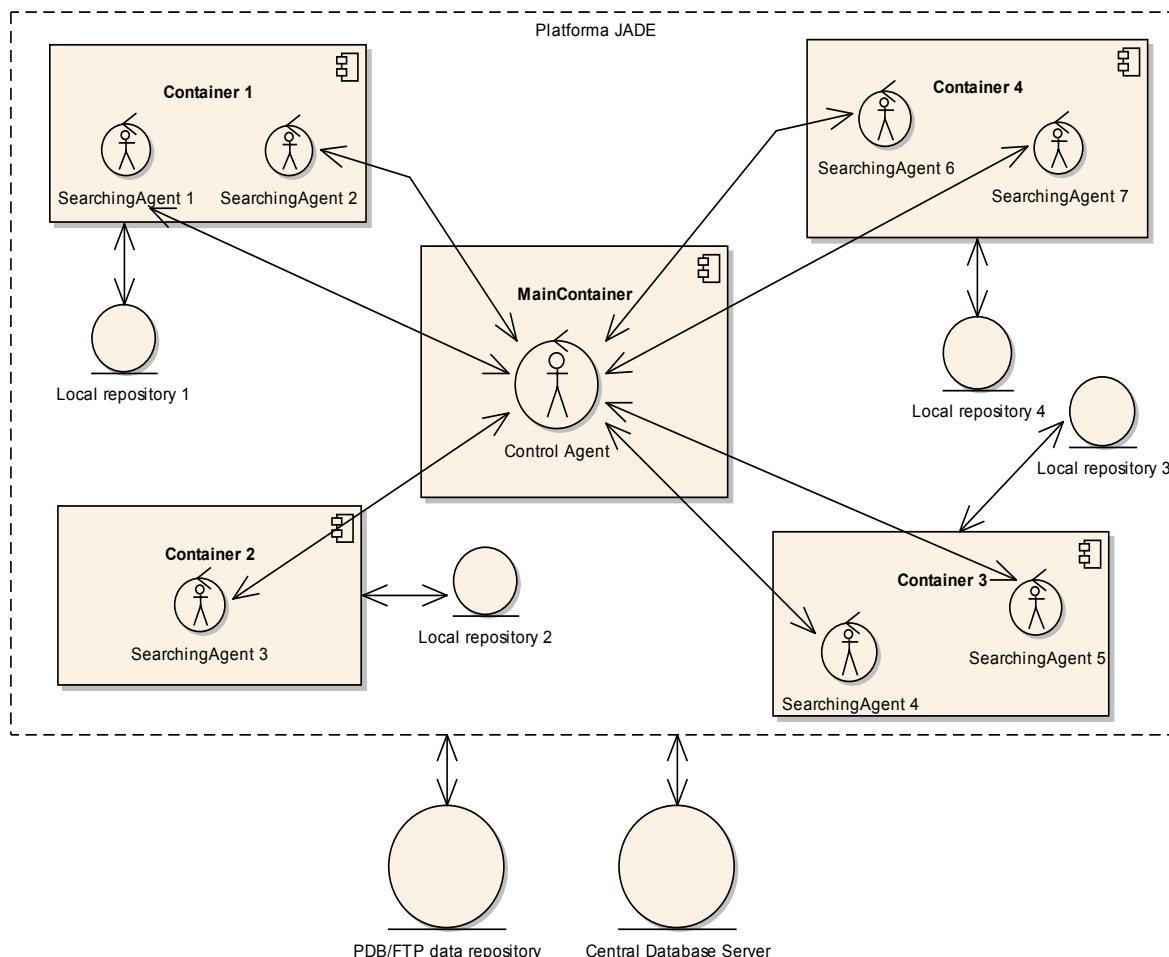
Proteomika to analiza proteomu, a jej celem jest poznanie i charakterystyka wszystkich białek powstających w wyniku ekspresji ludzkiego genomu [1]. Wiąże się ona zatem z badaniem zmian ekspresji białek w procesie rozwoju, rozumieniem złożoności systemu, jaki tworzą, białka oraz poznaniem ich roli w procesach patologicznych organizmów żywych. Bioinformatyka jako dziedzina i problemy badawcze, które się w niej pojawiają, stały się źródłem inspiracji dla twórców systemów komputerowych. Powstające oprogramowanie stanowi jedno z narzędzi w rozwiązywaniu problemów badawczych w tej dziedzinie. Oprogramowanie to jest często niezbędne do uzyskania nowych informacji [2]. Opracowanie algorytmów i programów służących do gromadzenia i analizy danych dotyczących struktur białkowych, porównywania ich do siebie, oceny ich podobieństwa i określenia różnic, jakie je dzielą, jest wykorzystywane w bioinformatyce proteomicznej i strukturalnej [3]. To nowe, dość wąskie obszary bioinformatyki, która korzysta z różnych programów dla oceny i analizy eksperymentów proteomicznych. Takie opracowania bioinformatyczne są często niezbędne do prawidłowej interpretacji otrzymanych wyników. Jednym z programów, które możemy wykorzystać do prowadzenia takich analiz, jest proponowany przez Autorów system MAS4PSi (od ang. *Multi Agent System For Protein Similarity searching*). Umożliwia on porównywanie do siebie różnych struktur białkowych, a dzięki temu – ocenę podobieństwa strukturalnego, co pośrednio przekłada się również na opis funkcjonalnego białka.

MAS4PSi jest zbudowanym na podstawie środowiska wieloagentowego [4-6] systemem umożliwiającym prowadzenie procesu poszukiwania podobieństwa strukturalnego białek (rys. 1). System tworzą:

- zestaw agentów wyszukiwania (ang. *Searching Agents*) prowadzących proces poszukiwania podobieństwa białek z użyciem jednego z zaimplementowanych algorytmów, np. CE [7], FATCAT [8] i in.;
- agent lub zbiór agentów zarządzających (ang. *Control Agent*) odpowiedzialnych za komunikację i rozdział zadań pomiędzy agentów wyszukiwania;
- platforma wieloagentowa JADE [9], pod kontrolą której działają wszystkie agenty systemu;
- centralny serwer bazy danych (ang. *Central Database Server*) przechowujący podstawowe informacje o strukturze białek, informacje opisowe, a także informacje o prowadzonych lub przeprowadzonych już poszukiwaniach wraz z informacjami o wynikach tych poszukiwań;
- wiele lokalnych repozytoriów danych (ang. *Local repository*) przechowujących podzbiory struktur białkowych replikowanych z bazy Protein Data Bank [10], wykorzystywanych

podczas prowadzenia procesu poszukiwania podobieństwa; zbiory te pozwalają uniknąć problemu wąskiego gardła podczas dostępu do centralnego serwera;

- zdalna baza danych Protein Data Bank (PDB), dostępna za pośrednictwem protokołu FTP, używana w przypadku braku wymaganych białek w lokalnych bazach danych (ang. *PDB/FTP data repository*).



Rys. 1. Architektura i główne elementy funkcjonalne systemu MAS4PSi

Fig. 1. Architecture and main functional elements of the MAS4PSi system

System MAS4PSi umożliwia współbieżną realizację procesu poszukiwania podobieństwa strukturalnego białek przez rozproszenie i zrównoleglenie obliczeń na wielu agentach wyszukiwania. Dzięki temu czas prowadzenia procesu poszukiwania podobieństwa skraca się odwrotnie proporcjonalnie do liczby agentów biorących udział w procesie. Szczegóły działania systemu zostały przedstawione w poprzednich pracach Autorów, m.in. w [11-13]. W niniejszym artykule przedstawiono natomiast wybrane scenariusze zastosowania zbudowanego przez Autorów systemu MAS4PSi w szeroko pojętej diagnostyce medycznej.

2. Scenariusz 1: Ocena wpływu zmiany polimorficznej lub mutacyjnej wykrytej w DNA na strukturę i funkcjonalność białka

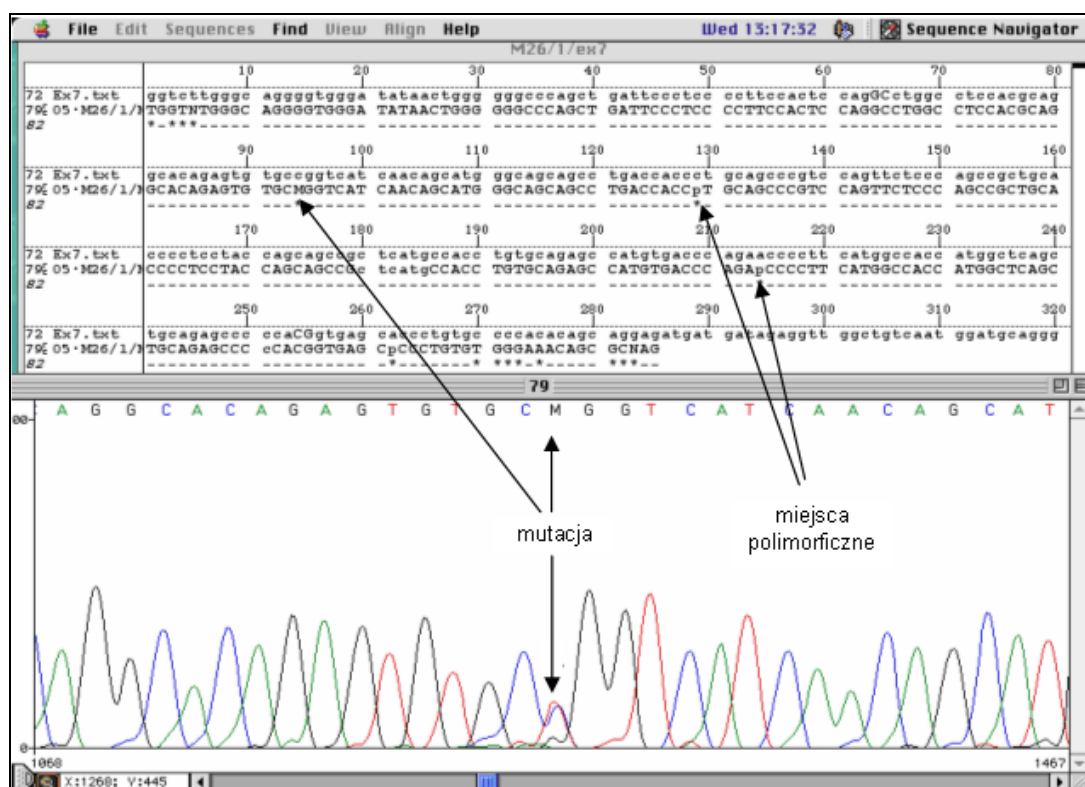
Zmiany w sekwencji DNA mogą być spontaniczne lub spowodowane działaniem czynników chemicznych bądź fizycznych. Jeżeli występują one z częstotliwością poniżej 1% w badanej populacji, określa się je jako zmiany mutacyjne – mutacje genomowe. Skutkiem takich zmian w sekwencji nukleotydowej może być zmiana funkcji białka – utrata lub rzadziej nabycie nowej funkcji. Niestety, rozpoznając zmianę w strukturze DNA, rzadko możemy jednoznacznie odpowiedzieć na pytanie, jak wpływa ona na zmianę funkcji czy struktury przestrzennej białka. Jeśli zmiana w DNA związana jest w konsekwencji z zamianą na aminokwas kończący translację, czyli stop kodon, to możemy jednoznacznie określić, że dojdzie do zmiany funkcji białka z powodu skrócenia jego łańcucha. Każda inna zmiana w DNA może mieć wpływ na strukturę i funkcję białka, lecz określenie rodzaju zmiany nukleotydowej oparte jest głównie na ocenie fenotypu probanta (mutant) lub ocenie korelacji występowania schorzenia u osób z mutacją w badanej rodzinie.

Dzięki oprogramowaniu takiemu jak MAS4PSi możliwe staje się choćby teoretyczne określenie wpływu zmiany sekwencji nukleotydowej na strukturę przestrzenną białka, a możliwość porównania otrzymanej struktury białkowej ze znanymi białkami o dokładnie znanych funkcjach pozwoli na hipotetyczną ocenę funkcji nabytej lub utraconej w wyniku mutacji.

2.1. Przykład 1

Przebieg scenariusza:

1. Identyfikacja nieznannej i nieopisanej do tej pory mutacji w genie HNF1-alfa u pacjenta z podejrzeniem cukrzycy MODY3 na podstawie sekwencyjnej analizy DNA (rys. 2). Zmiana mutacyjna dotyczy zmiany aminokwasu na inny aminokwas, co nie zmienia ramki odczytu, lecz może wpływać na zmianę struktury białka.
2. Określenie korelacji oznaczonej mutacji w rodzinie probanta – mutację określono jako związaną z występowaniem cukrzycy w badanej rodzinie, nie oceniono natomiast, jak zmienia się funkcja kodowanego białka.
3. Białko HNF1-alfa jest czynnikiem transkrypcyjnym i zmiana jego struktury bezpośrednio może się przekładać na zmianę jego funkcji. Poznanie przestrzennej struktury mutantu na podstawie postranslacyjnej struktury pierwszorzędowej pozwoli na hipotetyczną ocenę utraty lub zmiany funkcji czynnika HNF1-alfa, a tym samym – na próbę określenia zmian odpowiadających za fenotyp pacjenta.



Rys. 2. Zestawienie sekwencji eksonu 7 genu HNF1-alfa pacjenta i sekwencji referencyjnej w programie Sequence Navigator. Zamiana nukleotydu w pozycji 1340 C/T powoduje zmianę aminokwasu proliny na leucynę w kodonie 447 podczas procesu translacji. Miejsce wystąpienia mutacji, mające wygląd dwóch nachodzących się przebiegów, zostało oznaczone strzałką i literą (M), w sekwencji literami (p) oznaczono miejsca polimorficzne

Fig. 2. Alignment of patient's and reference sequences of exon 7 for gene HNF1-alpha in the Sequence Navigator program. Replacement of nucleotide C/T at position 1340 causes the change of the amino acid Proline for Leucine at codon 447 in the translation process. Location of the mutation, visible as two overlapping lines, has been marked with an arrow and the letter (M) in the sequence, letters (p) indicate polymorphic places

2.2. Przykład 2

Przebieg scenariusza:

1. Na podstawie analizy sekwencyjnej określono zmianę w DNA na pograniczu intron/egzon, która powodowała wypadnięcie jednego egzonu. Następnym etapem analizy było sekwencjonowanie cDNA obejmującego dwa sąsiadujące egzony. Stwierdzono wypadanie jednego z nich. Dla określenia konsekwencji fenotypowych zmian w sekwencji nukleotydowej analizowano u wszystkich członków rodziny.
2. Określono brak jednoznacznej korelacji fenotypu z genotypem, co przemawia raczej za alternatywnym składaniem niż mutacją powodującą wypadanie jednego z egzonów w mRNA. Powoduje to powstanie izoformy danego białka.
3. Analiza z wykorzystaniem oprogramowania MAS4PSi pozwala na ocenę różnic strukturalnych pomiędzy dwoma izoformami jednego białka.

2.3. Przykład 3

Przebieg scenariusza:

1. Metodami biologii molekularnej, takimi jak PCR-RLFP czy sekwencjonowanie, określono obecność zmian w genie MTHFR (reduktazy metylenotetra-hydrofolianowej). We wcześniejszych badaniach określono różną aktywność enzymu MTHFR w grupach badanych z różnymi schorzeniami, np. miażdżycą, zakrzepicą, nefropatią cukrzycową, nawracającymi poronieniami. Oceniono, że zmniejszona aktywność enzymu była związana z obecnością polimorfizmu C677T i/lub A1298C.
2. Fenotypowo u osób mających jedną lub obie zmiany wykazuje się zwiększoną predyspozycję do różnych schorzeń.
3. Ocena struktury 3D białka, która będzie zawierała jedną lub obie zmiany aminokwasowe związane z występowaniem opisywanych miejsc polimorficznych w genie MTHFR, pozwala na analizę miejsc aktywnych białka i określenie przyczyny spadku aktywności enzymu i umożliwi próbę określenia podłoża predyspozycji genetycznej.

3. Scenariusz 2: Poszukiwanie genów kandydujących dla schorzeń genetycznych na podstawie określenia zmian strukturalnych białka powodowanych przez SNP w badanym genie

SNP, czyli pojedyncze zmiany nukleotydu, występują w DNA średnio co 100-300 par zasad oraz odpowiadają za około 90% zmienności międzyosobniczej. Częstość ich występowania w danej populacji jest różna i różne jest również znaczenie biologiczne tych zmian. Poniżej przedstawiono przykłady wykorzystania systemu MS4PSi w analizie konsekwencji występowania SNP w łańcuchach DNA.

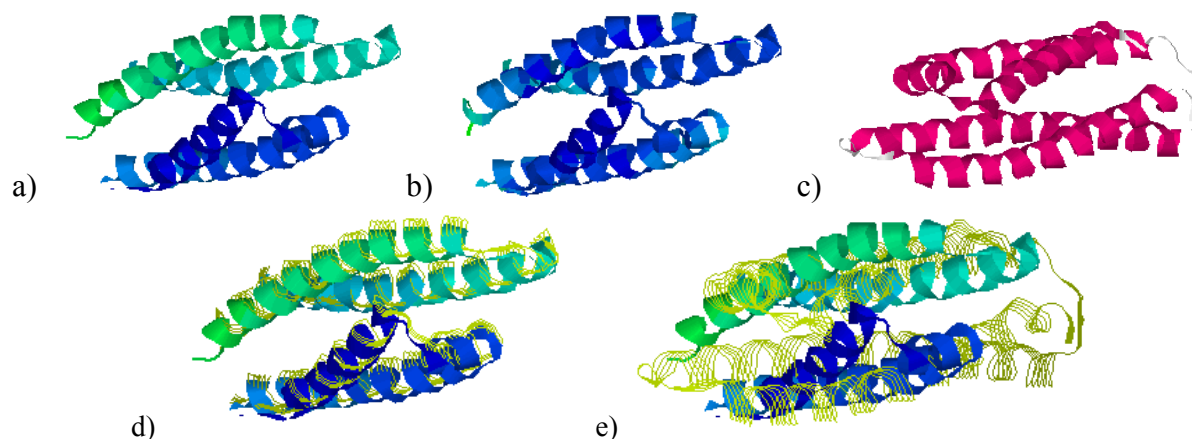
3.1. Przykład 1

Apolipoproteina E (ApoE) odgrywa istotną rolę w metabolizmie cholesterolu i triglicerydów. Jest jedną z pięciu rodzajów apolipoprotein występujących we krwi, wytwarzaną głównie w wątrobie i mózgu. Jej funkcje to transport lipidów z miejsca ich powstawania lub wchłaniania do tkanek, gdzie są magazynowane, oraz transport lipidów, a w szczególności cholesterolu, z narządów obwodowych do wątroby w celu ich wydalenia. ApoE reguluje również aktywność enzymów biorących udział w procesie metabolizmu lipidów i lipoprotein. Gen kodujący ApoE znajduje się na chromosomie 19, zawiera 3,7 kbp i składa się z 4 egzonów. W bazie NCBI opisano w tym genie 88 różnych SNP, każdy ma swój numer rs oraz przypisaną do niego lokalizację w genie ApoE.

W literaturze najczęściej omawiane są allele E2, E3 oraz E4. Najbardziej popularny jest allele E3, który występuje u większości populacji i nie koreluje w istotny sposób z chorobami sercowo-naczyniowymi. Allel E4 był opisywany jako związany ze zwiększonym ryzykiem rozwoju miażdżycy oraz ze zwiększonym ryzykiem choroby Alzheimera o późnym początku (rozвивa się ona u osób po 65 roku życia). Ponadto możemy obserwować efekt addytywny – obecność jednego allelu E4 niesie umiarkowanie zwiększone ryzyko choroby, natomiast posiadanie dwóch alleli E4/E4 związane jest już z dużo większym ryzykiem wystąpienia choroby Alzheimera. Mechanizm choroby nie jest jednak prosty i sama obecność allelu E4 nie wpływa bezpośrednio na zachorowanie, ma natomiast znaczenie dla przedklinicznego leczenia oraz przyspieszenia wystąpienia objawów choroby. Dla określenia obecności alleli E2, E3 oraz E4 opisano genetyczną klasyfikację, która została oparta na ocenie dwóch SNP, rs 429358 i rs 7412, w następującej konfiguracji:

- ApoE2 allele = rs429358 (T) + rs7412 (T)
- ApoE3 allele = rs429358 (T) + rs7412 (C)
- ApoE4 allele = rs429358 (C) + rs7412 (C)

Dostępne w bazach PDB (Protein Data Bank, <http://www.pdb.org>) struktury białkowe, wyszukane za pomocą systemów takich jak MAS4PSi, pozwalają na porównanie poszczególnych form białkowych ApoE między sobą (rys. 2) oraz analizę ewentualnych różnic lub odkształceń konformacyjnych wynikających z różnic genetycznych. Na rys. 3 przedstawiono konformacje ApoE reprezentowane przez struktury drugorzędowe.

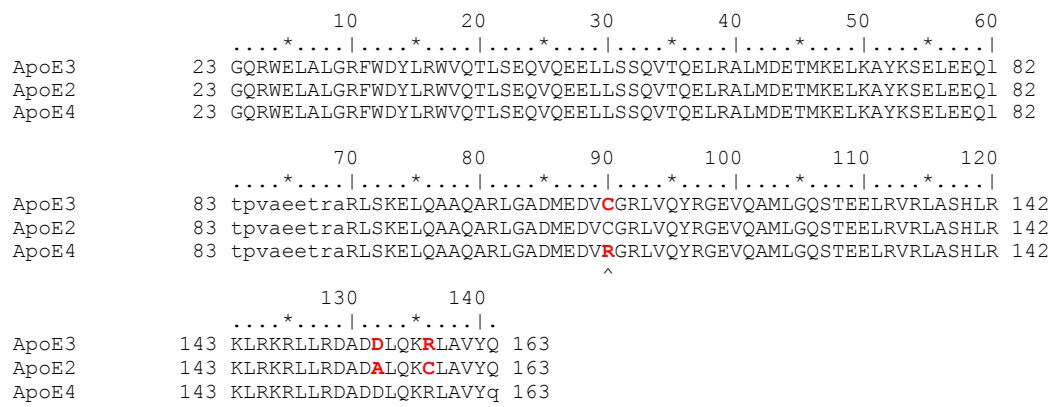


Rys. 3. Struktury przestrzenne (fragmenty): a) ApoE3, b) ApoE2, c) ApoE4, oraz porównanie struktur przez nałożenie na siebie: d) ApoE3 i ApoE2, e) ApoE3 i ApoE4 (druga struktura została oznaczona półprzezroczystą wstążką)

Fig. 3. Parts of spatial structures of: a) ApoE3, b) ApoE2, c) ApoE4, and comparison of structures by their superposition: d) ApoE3 and ApoE2, e) ApoE3 and ApoE4 (second structure is marked by using semitransparent ribbon)

Na rys. 4 przedstawiono porównanie sekwencji aminokwasów apolipoprotein E3 oraz E2 i E4 z zaznaczeniem powstałych mutacji. Pozwala to na określenie wpływu SNP na zmianę

struktury ApoE oraz porównanie tych danych z badaniami empirycznymi, takimi jak np. dobra odpowiedź na leczenie farmakologiczne choroby Alzheimera u pacjentów mających przynajmniej jeden allel E4.



Rys. 4. Porównanie na poziomie sekwencji z zaznaczonymi mutacjami

Fig. 4. Comparison on the level of primary structures with marked mutations

3.2. Przykład 2

Sposób działania:

1. Wśród danych literaturowych podawane są geny kandydujące odpowiadające za predyspozycje do wielu schorzeń, takich jak: otyłość, nadciśnienie tętnicze, cukrzyca, alergia itd., często w schorzeniach dziedziczonych wielogenowo. Brakuje jednak jednoznacznych informacji o jednym SNP korelującym z badanym schorzeniem. Jeśli projektujemy badanie (doświadczenie naukowe) mające dostarczyć informacji, czy dany gen jest związany z chorobą, konieczny jest wybór odpowiedniego SNP.
2. Korzystając z ogólnie dostępnych baz danych (np. NCBI), oceniamy wstępnie SNP pod względem: częstości występowania w badanej populacji, miejsca położenia w genie, wcześniej opisywanych korelacji z fenotypem. Niemniej jednak wybór optymalnego SNP do szerokiego badania w populacji z kilkudziesięciu, a czasem z kilkuset możliwych jest bardzo trudny. Nie bez znaczenia jest również koszt wykonania badania w wybranej populacji oraz czas, w jakim może być ono wykonane.
3. Program MAS4PSi pozwala na wstępną ocenę wpływu SNP na strukturę białka, a tym samym umożliwi celowany wybór tych SNP, które „najlepiej rokuja” do dalszych badań molekularnych. Pozwoli to na obniżenie kosztów eksperymentu oraz czasu pracy, w jakim badanie populacyjne mogłoby być wykonane.

4. Scenariusz 3: Ocena struktury 3D białka i określenie związku z występowaniem schorzenia

Liczba białek proteomu jest znacznie większa niż liczba kodujących je genów. Dzięki Projektowi Poznania Ludzkiego Genomu (HGP) ustalono, że ludzki genom zawiera w przybliżeniu 30-50 tysięcy genów kodujących białka. Liczba białek ludzkiego proteomu poznanych do tej pory to około 500 tysięcy. Oznacza to, że jeden gen może kodować około 10 różnych wariantów białek. Za występowanie takiej dysproporcji odpowiadają procesy molekularnego składania pre-mRNA, tzw. alternatywny splicing, oraz modyfikacje postranslacyjne, takie jak fosforylacja czy glikozylacja, które wpływają na ostateczną strukturę białka.

Poniżej przedstawiono dwa przykłady możliwości wykorzystania systemu MAS4PSi do analizy związków struktury białka z występowaniem schorzenia.

4.1. Przykład 1

Przebieg scenariusza:

1. Pacjent choruje na schorzenie genetyczne, np. mukowiscydozę, która jest związana ze zmianami w genie CFTR. Obraz tej choroby jest bardzo niejednorodny, a zmiany w sekwencji nukleotydowej w genie CFTR charakteryzują się różną manifestacją kliniczną, od pełnoobjawowej mukowiscydozy do zaburzeń ze strony układu rozrodczego, takich jak niedrożność nasieniowodów.
2. W celu wykonania analizy od pacjenta izolujemy białko będące produktem genu CFTR i przy użyciu dostępnych technik pozyskujemy jego strukturę 3D.
3. Przy użyciu oprogramowania MAS4PSi poszukujemy białek podobnych strukturalnie i porównujemy je z prawidłowym produktem genu CFTR.
4. Na podstawie przeprowadzonych analiz spróbujemy ocenić związek struktury 3D badanego przez nas białka z występowaniem choroby oraz jej obrazem u pacjenta. Możemy również spróbować określić wpływ zmutowanego białka na zachowanie prawidłowych funkcji białka CFTR. Oczywiście będą to teoretyczne rozważania, jeżeli jednak uda się nam zgromadzić większą grupę chorych, którzy pozwolą na przeprowadzenie takiego badania, zgromadzimy bazę danych, dzięki której każdy następny pacjent będzie mógł być pełniej diagnozowany.

4.2. Przykład 2

Przebieg scenariusza:

1. U badanego pacjenta została znaleziona zmiana w DNA, która dotychczas nie była nigdzie wcześniej opisana w żadnej bazie danych, ani jako miejsce polimorficzne, ani jako

zmiana mutacyjna. Analizując sekwencję aminokwasową, możemy określić, jak zmiana nukleotydowa w DNA wpłynie na sekwencję aminokwasową, ale nie zawsze jesteśmy w stanie powiedzieć, czy dojdzie do zmiany w strukturze białka i jak ta zmiana wpłynie na modelowanie przestrzenne białka.

2. Oszacowanie wpływu zmian sekwencyjnych na zamianę modelowania przestrzennego białka przez znalezienie podobnych struktur za pomocą systemu MAS4PSi może być pomocne w określaniu dostępności centrów aktywnych oraz biodostępności substratu dla białka enzymatycznego.

Przykład takiej analizy przedstawiono w publikacji Autorów [14] dla białka RAB5A, gdzie wykorzystano dane o profilach energetycznych z bazy EDB [15] oraz algorytm EAST [16-18].

5. Scenariusz 4: Zastosowanie programu MAS4PSi w proteomice klinicznej

Do wykrywania markerów biologicznych stosowane są techniki pozwalające na porównywanie złożonych mieszanin białkowych i ilościową ocenę ich poszczególnych składników. Metody te umożliwiają określenie zmian ekspresji białka, modyfikacji potranslacyjnych, interakcji między białkami i ich rozmieszczeniem w komórce. W proteomice klinicznej, dla której podstawowym celem jest poszukiwanie różnic w profilach białkowych osób chorych i zdrowych, stosowane są głównie elektroforeza dwuwymiarowa i spektrometria masowa. Obie metody umożliwiają identyfikację pojedynczych białek – pojedynczych markerów diagnostycznych. Analiza statystyczna wielu zmiennych charakteryzujących profile białkowe pozwala na identyfikację grup białek, dla których zmiany ekspresji mają znaczenie diagnostyczne lub prognostyczne. Analiza porównawcza struktur 3D przy użyciu programu MAS4PSi może być jednym z elementów umożliwiających pełniejszą diagnozę lub czynnikiem weryfikującym ją. Może również wspomagać poszukiwania zmienionych białek, które występują w stanach patologicznych, przez porównywanie ich struktur z fizjologicznie występującymi formami. Przydatność diagnostyczna identyfikowanych białek powinna być również weryfikowana innymi metodami, takimi jak testy immunodiagnostyczne lub/i Western blotting.

Przykładem takich zastosowań praktycznych, które w przyszłości mogą wspomóc dotychczasową diagnostykę laboratoryjną przy wykorzystaniu odpowiednio skonstruowanych narzędzi analizy bioinformatycznej, są poszukiwania różnic profili białkowych, opisujących stany chorobowe, takie jak:

1. Proteomika w badaniach choroby Alzheimera i schizofrenii. Badania dotyczą chorób neurodegradacyjnych. W chorobie Alzheimera badania proteomiczne koncentrują się na po-

szukiwaniach nowych potencjalnych markerów choroby, które pozwoliłyby również na przewidywanie szybkości rozwoju choroby. W badaniach nad schizofrenią określono kilka potencjalnych markerów, które mogłyby być wykrywane i oceniane na podstawie proteomu wątroby i krwinek czerwonych, a nie tylko płynu mózgowo-rdzeniowego.

2. Proteomika w prenatalnym rozpoznawaniu zespołu Downa. Ponieważ tradycyjne prowadzenie prenatalnej diagnostyki tego schorzenia obarczone jest nadal dość znacznym odsetkiem wyników fałszywie dodatnich, badanie proteomiczne skierowane są na odszukanie nowych biomarkerów wykrywanych w surowicy matki, które pozwolą na szybką i celowaną diagnostykę.
3. Proteomika w badaniach chorób reumatycznych. Zastosowanie proteomiki do diagnostyki tych schorzeń umożliwi ocenę pełnego profilu nieprawidłowości immunologicznych, co pozwoliłoby na pełne różnicowanie tych schorzeń oraz wczesne ich wykrycie.
4. Białkowe markery chorób nowotworowych. Badania proteomu skupiają się głównie na wykryciu wczesnych markerów nowotworzeni, a także ocenie zmian w ekspresji białek podczas progresji choroby.
5. Określenie zmian w proteomie przy narażeniu na dym papierosowy, benzen, metale ciężkie – ocena proteomu osób narażonych na kontakt ze szkodliwymi substancjami, zwłaszcza przy przewlekłym narażeniu.
6. Proteomika a fizjologia i patofizjologia nerek – zastosowanie proteomiki w badaniach moczu, który jest materiałem łatwym do pobrania i powszechnie dostępnym; poszukiwanie markerów związanych z zaburzeniami funkcji nerek oraz ocena ekspresji białek korelujących z procesami starzenia.

W każdym z wymienionych przypadków analiza strukturalna białek mogłaby być dobrym uzupełnieniem istniejących metod.

6. Scenariusz 5: Użycie MAS4Pi w projektowaniu doświadczeń i eksperymentów naukowych mających zastosowanie w medycynie i farmacji

Przy znajomości sekwencji aminokwasowej lub struktury białkowej i przy użyciu odpowiedniego oprogramowania, takiego jak MAS4PSi, możliwe jest określenie podobieństwa białek. Dotyczy to zarówno stopnia homologiczności, podobieństwa funkcji, jak i zastosowania porównywanych, badanych białek jako potencjalnych antygenów, np. szczepionkowych. Oczywiście wszystkie te analizy nie zastąpią w pełni empirycznych doświadczeń i eksperymentów, ale mogą być bardzo pomocne w projektowaniu doświadczeń, co nie jest bez wpływu na koszty eksperymentu. Decydującą rolę w doborze odpowiedniego białka,

mającego w przyszłości być np. antygenem wykorzystanym do produkcji szczepionek, będzie miał wynik eksperymentu naukowego, jednakże wartość analiz informatycznych może znacznie skrócić czas realizacji eksperymentu oraz jego koszty.

7. Scenariusz 6: Opracowanie nowych substancji leczniczych, czyli projektowanie leków na podstawie porównywania struktur białkowych

Od kilkunastu lat znane są powszechnie osobnicze różnice w reakcji na różne leki. Ocenia się, że mają na nie wpływ takie czynniki, jak:

- wpływ polimorfizmu genetycznego enzymów z grupy CYP na farmakokinetykę i farmakodynamikę leku,
- genetyczne warianty białkowych transporterów leków,
- wiązanie leków z różnymi wariantami genetycznymi białek osocza.

Większość tych czynników związana jest z osobniczo zmiennymi strukturami białkowymi, które występują zarówno na etapie wchłaniania, transportu leku, jak i jego biotransformacji. Wieloetapowość tego procesu i złożoność oddziaływań biologicznych mogą częściowo być wyeliminowane przez wcześniejsze badanie oddziaływań białko-białko na podstawie oceny i porównania struktur białkowych przy użyciu programu MAS4PSi.

Porównanie struktur białek biorących udział w metabolizmie leku u różnych pacjentów pozwoli na wybranie optymalnego, najbardziej uniwersalnego projektu leku. Użycie programu MAS4PSi umożliwi przeprowadzenie wielu teoretycznych testów na różnych etapach metabolizmu leku, które pozwolą na dobór najlepszej struktury leku – optymalnej dla szerokiej populacji. Taka strategia wstępnego projektowania „uniwersalnego leku” pozwoli na zmniejszenie kosztów eksperymentów doświadczalnych oraz pozwoli na przewidywanie i wcześniejsze rozwiązywanie problemów związanych z biotransformacją oraz potencjalnymi działaniami ubocznymi leku.

Drugim nurtem w zakresie farmakologii, w którym możemy wykorzystać oprogramowanie MAS4PSi, jest poszukiwanie nowych, lepszych odpowiedników znanych już farmaceutyków, które często również są białkami lub bezpośrednio wchodzą z nimi w interakcję. Oprogramowanie MAS4PSi powala na wyszukanie podobnych do znanego już leku struktur białkowych. Będziemy również mogli ocenić, jak niewielka nawet zmiana struktury ocenianego leku wpłynie na interakcje z poszczególnymi białkami biorącymi udział w jego metabolizmie. Taka ocena w bezpośredni sposób przekłada się na określenie biodostępności czy szlaku metabolizmu leku w organizmie.

8. Scenariusz 7: Identyfikacja wyizolowanych białek na podstawie strukturalnego przyporządkowania białka

Oprogramowanie MAS4PSi wydaje się cennym narzędziem pozwalającym na określenie i ocenę nieznanych dotąd białek zarówno pod względem struktury, jak i funkcjonalności. Dzięki porównywaniu rozległej bazy białkowej o znanych funkcjach i ustalonej strukturze możliwe jest ocenianie nowych, nieznanych struktur oraz przyporządkowanie ich do odpowiednich grup białek.

8.1. Przykład 1

U pacjentów z wybraną chorobą nowotworową z surowicy krwi wyizolowano białko o nieznaną strukturę. Potrzebna jest nam informacja: czy znamy już białko o takiej strukturze? A jeżeli nie, to do jakiej grupy białek możemy je zakwalifikować? Jaka może być jego funkcja w rozwoju, przebiegu choroby?

Dzięki oprogramowaniu MAS4PSi możliwe jest porównanie struktur białkowych do siebie oraz odnalezienie struktury badanego białka bądź najbardziej do niej podobnej. Taki algorytm postępowania pozwoli na określenie i rozpoznanie białka wyizolowanego od chorych i być może opracowanie nowego markera dla badanej choroby nowotworowej. Prawdopodobnie możliwe będą również ustalenie funkcji ocenianego białka w rozwoju choroby lub odpowiedzi na leczenie oraz ocena wpływu białka na określenie skuteczności terapii przeciwnowotworowej.

8.2. Przykład 2

Każdy zajmujący się wyczynowo sportem zawodnik przed zawodami musi być poddany kontroli antydopingowej. Badane są standardowe substancje dopingowe, ale zdarza się, że kontrola ta komercyjnymi testami nie wykryje wszystkich niedozwolonych substancji. Jeżeli u danego sportowca zostałyby wykryte nieznaną substancję o charakterze białka, podejrzana o działanie dopingowe, to przy użyciu programu MAS4PSi można przeprowadzić analizę porównawczą struktur białkowych dostępnych w bazie danych. Pozwoli to na uzyskanie odpowiedzi na następujące pytania:

- Z jaką substancją mamy do czynienia?
- Jaka jest jej struktura lub do jakiej grupy białkowej należy?
- Czy może mieć działanie dopingowe?

Użycie oprogramowania MAS4PSi umożliwia szybką i skuteczną analizę oraz może się przyczynić do poszerzenia dotychczasowej wiedzy.

9. Podsumowanie

System MAS4PSI można szeroko stosować w analizie złożonych procesów zachodzących w komórkach organizmów żywych, a w konsekwencji również i w diagnostyce medycznej oraz farmacji ukierunkowanej na tworzenie nowych biocząstek. Niektóre z opisanych w niniejszym artykule scenariuszy wciąż pozostają w obszarze potencjalnych zastosowań systemu MAS4PSI. Jest to niestety konsekwencją braku odpowiedniej aparatury diagnostycznej przez środowisko medyczne. Wydaje się jednak bardzo prawdopodobne, że w niedalekiej przyszłości metody diagnostyczne będą oparte nie tylko na analizie sekwencyjnej materiału genetycznego, lecz także na dokładnej analizie strukturalnej cząstek, jakimi są białka. Już dziś systemy porównujące struktury białkowe są stosowane w codziennej praktyce przez koncerny farmaceutyczne podczas projektowania nowych leków. Rozszerzenie wachlarza dalszych zastosowań jest tylko kwestią czasu.

BIBLIOGRAFIA

1. Gu J., Bourne P. E.: *Structural Bioinformatics (Methods of Biochemical Analysis)*. Wiley-Blackwell, 2009.
2. Eidhammer I., Jonassen I., Taylor W. R.: *Protein Bioinformatics: An Algorithmic Approach to Sequence and Structure Analysis*. Wiley, 2004.
3. Gibrat J. F., Madej T., Bryant S. H.: Surprising similarities in structure comparison. *Curr Opin Struct Biol.* 6(3), 1996, s. 377÷385.
4. Padgham L., Winikoff M.: *Developing Intelligent Agent Systems: A Practical Guide*. Halsted Press, New York, USA 2004.
5. Wooldridge M.: *An Introduction to Multiagent Systems*, 1st edition. John Wiley & Sons, 2002.
6. Mrozek D., Małysiak B., Augustyn W.: Agent-supported Protein Structure Similarity Searching. *Lecture Notes in Artificial Intelligence*, Springer-Verlag GmbH, Vol. 5044, 2008, s. 49÷61.
7. Shindyalov I. N., Bourne P. E.: Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering*, Vol. 11(9), 1998, s. 739÷747.
8. Ye Y., Godzik A.: Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics*, Vol. 19(2), 2003, s. 246÷255.
9. Bellifemine F. et al.: JADE, A White Paper, 2003, <http://jade.tilab.com/papers/2003/WhitePaperJADEEXP.pdf>

10. Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H. et al.: The Protein Data Bank. *Nucleic Acids Res.*, No. 28, 2000, s. 235÷242.
11. Momot A., Małysiak-Mrozek B., Kozielski S., Mrozek D. et al.: Improving Performance of Protein Structure Similarity Searching by Distributing Computations in Hierarchical Multi-Agent System. 2nd International Conference on Computational Collective Intelligence – Technologies and Applications (ICCCI-2010), Kaohsiung, Taiwan, *Lecture Notes in Artificial Intelligence*, Springer-Verlag, LNAI 6421, 2010, s. 320÷329.
12. Małysiak-Mrozek B., Momot A., Mrozek D., Hera Ł. et al.: Scalable system for protein structure similarity searching. *Proc. of the Third International Conference on Computational Collective Intelligence: Technologies and Applications (ICCCI'11)*, Vol. 6923, Part II, Springer-Verlag, Berlin, Heidelberg, s. 271÷280.
13. Małysiak-Mrozek B., Momot A., Mrozek D., Momot M.: Architektura hierarchicznego systemu wieloagentowego dla procesu poszukiwania podobieństwa białek. *Studia Informatica*, Vol. 33, No. 2A (105), Gliwice 2012, s. 83÷97.
14. Mrozek D., Małysiak-Mrozek B., Kozielski S.: Energy Properties of Protein Structures in the Analysis of the Human RAB5A Cellular Activity. *Advances in Intelligent and Soft Computing*, Vol. 59, Springer Verlag GmbH, 2009, s. 121÷131.
15. Mrozek D., Małysiak-Mrozek B., Kozielski S., Świerniak A.: The Energy Distribution Data Bank: Collecting Energy Features of Protein Molecular Structures. *Ninth IEEE International Conference on Bioinformatics and BioEngineering*, IEEE, 2009, s. 301÷306.
16. Mrozek D., Małysiak-Mrozek B.: An Improved Method for Protein Similarity Searching by Alignment of Fuzzy Energy Signatures. *International Journal of Computational Intelligence Systems*, Atlantis Press, 2011, s. 75÷88.
17. Mrozek D., Małysiak-Mrozek B., Kozielski S.: Alignment of protein structure energy patterns represented as sequences of Fuzzy Numbers. *Annual Meeting of the North American Fuzzy Information Processing Society*, 2009. NAFIPS 2009. Cincinnati, USA, IEEE, 2009, s. 1÷6.
18. Mrozek D., Małysiak B., Kozielski S.: An optimal alignment of proteins energy characteristics with crisp and fuzzy similarity awards. *Fuzzy Systems Conference, FUZZ-IEEE 2007*. IEEE International, 2007, s. 1÷6.

Wpłynęło do Redakcji 16 stycznia 2013 r.

Abstract

Proteomics is an analysis of the proteome, whose goal is to study and characterize all the proteins produced by the expression of human genome. It thus involves the study of changes in protein expressions, understanding the complexity of the systems, which proteins are, and understanding roles of proteins in pathological processes in living organisms. Bioinformatics as a field of research, and all problems that occur in it, has become a source of inspiration for the developers of computer systems. The resulting software is one of the tools to solve research problems in this area. The software is often necessary to obtain new, useful information. Development of algorithms and programs for collecting and analyzing data on protein structures, comparing them to each other in order to assess their similarities and differences, is used in proteomics and structural bioinformatics. These are new and quite narrow areas of bioinformatics, which makes use of a variety of programs for the assessment and analysis of proteomics experiments. Such bioinformatics software is often necessary for the correct interpretation of results of these experiments. One of the programs that we can use to conduct such an analysis is the MAS4PSi system (Multi Agent System For Protein Similarity searching) shown in the paper. It allows comparison of different protein structures to each other, and thus evaluation of their structural similarity, which also allows the functional description of proteins. MAS4PSi is a system that allows fast, scalable and reliable protein structure similarity searching. In this article we define and present possible scenarios of using the MAS4PSi system in a broadly understood medical diagnostics, both in the design of experiments supporting new research, as well as the typical diagnostic stage.

Adresy

Sylwia GÓRCZYŃSKA-KOSIORZ: Śląski Uniwersytet Medyczny; Katedra i Klinika Chorób Wewnętrznych Diabetologii i Nefrologii, ul. 3-go Maja 13-15, 41-800 Zabrze, Polska, sykos@wp.pl.

Bożena MAŁYSIAK-MROZEK: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-100 Gliwice, Polska, bozena.malysiak-mrozek@polsl.pl.

Dariusz MROZEK: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-100 Gliwice, Polska, dariusz.mrozek@polsl.pl.