

KRZYSZTOF PAWŁOWSKI

*Instytut Biologii Doświadczalnej im. M. Nenckiego PAN  
Pasteura 3, 02-093 Warszawa  
Wydział Rolnictwa i Biologii SGGW  
Nowoursynowska 166, 02-776 Warszawa  
E-mail: k.pawlowski@nencki.gov.pl*

## BIOINFORMATYKA W POSZUKIWANIU NOWYCH LEKÓW

### WSTĘP

Pod koniec XX w. w przemyśle farmaceutycznym ukształtował się paradygmat procesu tworzenia nowatorskich leków opartego na molekularnym mechanizmie choroby (ang. mechanism-based) (SAMS-DODD 2006). Wcześniejszy paradygmat oparty na fizjologii (ang. physiology-based) polegał na poszukiwaniu środka interwencji (leku), który bezpośrednio wpływałby korzystnie na widoczne, mierzalne symptomy choroby, u pacjenta lub w zwierzęcym modelu choroby. Poważnymi wadami tego paradygmatu było z jednej strony to, iż zależał krytycznie od wyboru prawidłowego modelu choroby, a z drugiej strony – ryzykowano, iż lek będzie zwalczał symptomy, a nie samą chorobę.

Nowy paradygmat mechanizmu działania opierał się na optymistycznym założeniu, iż dynamicznie rozwijające się techniki nowoczesnej biologii molekularnej i komórkowej pozwolą na dogłębne zrozumienie molekularnych mechanizmów chorób i w konsekwencji na zidentyfikowanie kluczowych „winowajców” postępu procesu chorobowego w osobach genów, których nefizjologiczne działanie ten proces uruchamia, wzmacnia bądź podtrzymuje. Dalej, w uproszczonej wersji, paradygmat ten skłaniał do poszukiwania genów związanych z chorobami, np. genów podatności na choroby, a następnie poszukiwania środków interwencji (leków) pozwalających na przywrócenie prawidłowego funkcjonowania produktów takich genów. W najprostszej wersji, poszukiwano pojedynczego genu (genu docelowego dla leku,

ang. drug target), takiego, że modulowanie jego działania pozwoliłoby na powstrzymanie lub odwrócenie procesu chorobowego. Jeśli więc na przykład za postępy choroby odpowiedzialne byłoby nadmierne przesyłanie sygnałów przez pewien receptor w błonie komórkowej, poszukiwano by leku blokującego ten receptor. W ramach tego paradygmatu kluczowy jest trafny wybór genu docelowego dla leku (ang. target identification). W sytuacji optymistycznej, precyzyjne uderzenie lekiem zaprojektowanym tak, aby działał na jeden „cel”, powinno pozwolić na wyleczenie choroby bez spowodowania niekorzystnych efektów ubocznych. Niestety, to nowoczesne podejście do tworzenia leków, wbrew powszechnym oczekiwaniom, nie zaowocowało lawiną nowych, bezpiecznych i skutecznych leków. W latach 1993–2007 łączne roczne wydatki przemysłu farmaceutycznego na badania i rozwój wzrosły około czterokrotnie (SHEKHAR 2008), osiągając poziom 60 mld US\$. Jednocześnie liczba nowych leków rocznie dopuszczonych do użycia pozostała niemal niezmienną. Biorąc pod uwagę, iż zwiększając wydatki, przemysł intensywnie wdrażał unikalne nowoczesne techniki badawcze, nie można było uniknąć rozczarowania wśród pacjentów, lekarzy, inwestorów i samych badaczy. Przykładem może być mukowiscydoza (ang. cystic fibrosis). Związek mutacji pewnego genu o nazwie *CFTR*, z tą chorobą odkryto w 1989 r. Mimo bardzo intensywnych badań, po niepowodzeniach terapii genowej (podawaniu „nieuszkodzonego”

białka CFTR), dopiero ostatnio lek na tę chorobę trafił do trzeciej fazy badań klinicznych (COUZIN-FRANKEL 2009). Należy pamiętać, że okres 10–15 lat to typowy czas między rozpoczęciem prac nad terapią a dopuszczeniem leku na rynek. Nawet jeśli projekty tworzenia nowych leków nie kończą się pomyślnie, jak ma to miejsce w większości przypadków i jak to było dotąd w przypadku mukowiscydozy, badania przynoszą głębsze zrozumienie procesu chorobowego. W przypadku genu *CFTR* okazało się, iż występuje w nim wiele mutacji, w różny sposób związanych z mechanizmem choroby, a także niezwiązanych z chorobą.

Tematem niniejszego artykułu jest bioinformatyka w procesie tworzenia nowych leków. Ponieważ metody obliczeniowe biologii, jak pokazemy szczegółowo poniżej, są nieodłącznym elementem każdego etapu tego procesu, przedstawiając wkład bioinformatyki, będziemy jednocześnie omawiać słabości współczesnego paradygmatu opartego

na mechanizmie i odnosić się do wyzwań, jakie stają przed nauką poszukującą nowych terapii.

W niniejszym, krótkim przeglądzie zastosowań bioinformatyki w przemyśle farmaceutycznym, termin ten będzie z jednej strony stosowany w sposób zawężony w porównaniu z przyjętym w przemyśle, ponieważ nie będzie tu mowy np. o biomedycznych bazach danych i innych podobnych narzędziach informatycznych. Z drugiej strony, termin „bioinformatyka” będzie tu używany w sposób szerszy niż to stosuje się w przemyśle, gdyż będzie mowa o zastosowaniach z pogranicza biologii obliczeniowej i chemii obliczeniowej.

Ponieważ w procesie tworzenia nowych leków ma zastosowanie każda niemal poddziedzina bioinformatyki, nie będziemy tu przedstawiać szczegółowo różnych metod bioinformatycznych, raczej postaramy się koncentrować się na ich roli w procesie poszukiwania nowych terapii.

## ETAPY PROCESU TWORZENIA NOWYCH LEKÓW

Proces tworzenia nowatorskich leków (Tabela 1) na chorobę, dla której dostępne terapie są niezadowalające, z reguły zaczyna się od (1) zrozumienia molekularnych i komórkowych mechanizmów procesu chorobowego, związanych z wybranym przez ekspertów medycznych efektem biologicznym, odpowiedzialnym za określony aspekt choroby. To zrozumienie ułatwia (2) poszukiwanie i potwierdzenie właściwego punktu interwencji terapeutycznej (ang. drug target), zwykle białka, którego funkcję biologiczną ma modulować projektowany lek. Kolejnym eta-

pem, jest (3) opracowanie odpowiedniego środka interwencji (tzw. kandydata na lek, ang. candidate drug; zwykle jest to mała cząsteczka chemiczna, ang. small-molecule drug, albo też np. przeciwciało terapeutyczne bądź inne białko). Dalsze kroki to (4) potwierdzenie skuteczności i doskonalenie kandydata na lek w układach modelowych i w modelach zwierzęcych. Zanim potencjalny lek zostanie dopuszczony do pierwszych badań klinicznych, niezbędna jest (5) ocena jego bezpieczeństwa.

## ZROZUMIENIE MECHANIZMÓW CHOROBY

Zrozumienie molekularnych mechanizmów choroby jest bardzo długotrwałym procesem, często dziełem pokoleń badaczy. Zwykle na taki proces składa się wiele wycinkowych badań i obserwacji z zakresu medycyny, biologii molekularnej i komórkowej, biochemii, genetyki i wielu innych dziedzin. Badania takie są najczęściej prowadzone w wielu ośrodkach akademickich i przemysłowych, a ich wyniki – rozproszone w wielu publikacjach, których liczba dla jednej choroby nieraz idzie w tysiące. Niezbędnymi na-

rzędziami bioinformatycznymi, pozwalającymi wykorzystywać ten gmach wiedzy o mechanizmach chorobowych do projektowania leków, są literaturowe bazy danych, np. Medline, oraz narzędzia umożliwiające przeszukiwanie i analizowanie literatury biomedycznej, łącznie z informacjami o obiektach biologicznych, których ona dotyczy, np. OMIM. Obiektami takimi mogą być geny, białka, metabolity oraz obiekty abstrakcyjne, takie jak anotacje funkcjonalne czy ścieżki sygnalizacyjne i metaboliczne. Narzędzia do analizy sieci

Tabela 1. Typowe zastosowania bioinformatyki w procesie tworzenia nowych leków.

Etap procesu tworzenia nowych leków	Zadania bioinformatyki
analiza molekularnych mechanizmów procesu chorobowego	<ul style="list-style-type: none"> <li>• analizy danych mikromacierzowych (ekspresji genów) w chorobie, w kontekście wiedzy biologicznej</li> <li>• analizy funkcjonalnego znaczenia polimorfizmów związanych z chorobą</li> <li>• badanie zależności między danymi klinicznymi a obrazem molekularnym choroby</li> </ul>
poszukiwanie właściwego punktu interwencji terapeutycznej (białko docelowe, drug target)	<ul style="list-style-type: none"> <li>• analiza zależności w grupach genów związanych z chorobą – analiza ścieżek</li> <li>• symulacje ścieżek sygnalizacyjnych</li> <li>• poszukiwanie i katalogowanie białek podatnych na uderzenie lekiem (druggable)</li> <li>• przewidywanie struktur potencjalnych białek docelowych</li> </ul>
opracowanie odpowiedniego środka interwencji (kandydata na lek)	<ul style="list-style-type: none"> <li>• budowa modelu struktury białka docelowego</li> <li>• dokowanie potencjalnych kandydatów do struktury białka docelowego</li> </ul>
potwierdzanie skuteczności i doskonalenie kandydata na lek	<ul style="list-style-type: none"> <li>• analizy danych mikromacierzowych – badanie molekularnych mechanizmów odpowiedzi na lek (farmakogenomika, pharmacogenomics)</li> </ul>
ocena bezpieczeństwa kandydata na lek	<ul style="list-style-type: none"> <li>• analizy danych mikromacierzowych – badanie efektów podania leku na stan komórki (toksykogenomika, toxicogenomics)</li> </ul>

zależności między obiektami biomedycznymi (ang. Biomedical Entity Relationship Systems, BERS) (EKINS i współaut. 2007) pozwalają na przykład wyszukiwać, w literaturze i powiązanych z nią bazach danych, informacje o zależnościach między procesami chorobowymi a grupami genów i białek.

Typowymi eksperymentami, z których można czerpać pomysły na nowe scenariusze terapeutyczne oparte na mechanizmie, są badania w populacjach powiązań genetycznych (ang. genetic association studies) z chorobami oraz badania mikromacierzowe, porównujące ekspresję genów w tkankach chorych i zdrowych. Z pierwszego rodzaju badań, uzyskuje się informacje o korelacjach między występowaniem pewnych mutacji w jednym lub kilku genach a podatnością na chorobę. Z badań drugiego rodzaju, otrzymuje się dane o zmianach w poziomie ekspresji genów w chorobie (mogą to być duże grupy genów) oraz, często, informacje o skorelowaniu poziomu ekspresji genów z różnymi danymi klinicznymi opisującymi nasilenie procesu chorobowego.

Zarówno badania genetyczne, jak i mikromacierzowe w kluczowy sposób zależą od zaawansowanych metod statystycznych, pozwalających wykryć statystycznie istotny sygnał biologiczny, związany z chorobą, separując go od innych sygnałów biologicznych, np. związanych ze zmiennością osobniczą, niejednorodnym składem tkankowym badanych próbek, wpływem środowiska na badanych dawców. Jednak nawet takie istotne wyniki mogą mieć znikomą wartość praktyczną dla procesu projektowania nowych leków, jeśli nie pozwalają na tworzenie użytecznych hipotez terapeutycznych. Ułatwiają to narzędzia typu BERS, pozwalające rozpatrywać związane z chorobą geny w kontekście dostępnej wiedzy biologicznej oraz ich relacji z innymi obiektami biomedycznymi.

Dodatkową trudnością w odważnym formułowaniu nowych hipotez terapeutycznych jest zrozumiała obawa, zwłaszcza laboratoriów przemysłowych, przed wykorzystywaniem informacji o mało znanych genach do dalszych badań (PAWŁOWSKI 2008). Uważa się, że takie geny są bardzo ryzykownymi

objektami badań, ponieważ szanse szybkiego opracowania terapii modulującej ich działanie są znikome. Obok genów „popularnych”, będących tematem setek i tysięcy prac, jest wiele genów, o których nie wiadomo niemal nic (HOFFMANN i VALENCIA 2003). Wielu badaczy, również akademickich, ma niestety tendencję do wstępnego filtrowania wyników wielkoskalowych eksperymentów, jak np. eksperymenty mikromacierzowe, i rozpatrywania wyłącznie danych dla genów dobrze zbadanych. Typowym przykładem może być praca dotycząca wpływu dymu tytoniowego na nabłonek oskrzelikowy, której autorzy, uzyskawszy dane na temat zmian ekspresji około dwudziestu tysięcy ludzkich genów, przeanalizowali w swojej publikacji zaledwie czterdzieści cztery geny, dobrze wcześniej przebadane i pasujące do przyjętego przez nich scenariusza (HACKETT i współaut. 2003).

Podejście polegające na poszukiwaniu pojedynczego genu „odpowiedzialnego” za chorobę jest tyle optymistyczne, co naiwne. Może to zilustrować tytuł pewnego kome-

ntarza na temat odkrycia genu *ADAM33* jako istotnie powiązanego z astmą: „*ADAM 33: just another asthma gene...?*”. Z tytułu tego przebija lekka irytacja, gdyż nie był to pierwszy gen okrzyknięty jako „ten jedyny” odpowiedzialny za chorobę. W miarę pojawiania się coraz to nowych doniesień genetycznych o „nowym genie choroby X” oraz doniesień genomicznych (mikromacierzowych) o nowych „genach z nadekspresją w chorobie Y” wzrastała świadomość, że za choroby odpowiedzialne jest zwykle działanie grup genów, tworzących ścieżki sygnalizacyjne, regulacyjne bądź metaboliczne (ang. pathways).

Budowa wiedzy o molekularnym obrazie choroby, to dopiero początek drogi do nowych terapii. Z różnych eksperymentów uzyskuje się dane o związku między różnicami w aktywności genów bądź różnymi wariantami genów a procesem chorobowym. Jak tę wiedzę o korelacji własności obiektów mikroskopowych (geny) i makroskopowych (osobniki) przetłumaczyć na scenariusze terapeutyczne.

#### WYBÓR WŁAŚCIWEGO PUNKTU INTERWENCJI TERAPEUTYCZNEJ – CZYLI TRUDNA SZTUKA ZNAJDOWANIA CELU

W poszukiwaniu genów docelowych dla leków (ang. target discovery) istotne jest skatalogowanie wszystkich potencjalnych genów docelowych. Takie katalogowanie staje się możliwe dzięki zsekwencjonowaniu genomu człowieka oraz genomów wielu patogenów. Praktyka projektowania leków uczy, iż skuteczne leki zwykle modulują działanie białek z kilku zaledwie klas. Są to enzymy (zwłaszcza kinazy i proteazy), receptory błony komórkowej (zwłaszcza receptory sprzężone z białkami G, GPCR), kanały jonowe oraz jądrowe receptory hormonów. Zadaniem bioinformatyki, oprócz katalogowania dobrze znanych rodzin białkowych zawierających białka docelowe znanych leków, jest poszukiwanie nowych potencjalnych białek docelowych dla leków, na przykład przez szukanie białek o bardzo odległym podobieństwie do znanych białek docelowych. Przykładem może być wykrycie w rodzinie białek, uważanych wcześniej za kanały jonowe, domeny strukturalnej pełniącej prawdopodobnie funkcję enzymatyczną (proteazy). Odkrycia tego dokonano metodą przewidywania struktury białka przez wykazanie bardzo odległych podobieństw sekwencji, na granicy wykrywalności (PAWŁOWSKI i współaut. 2006). Sto-

sując podobne metody można ocenić, że w genomie człowieka dostępne dla leków jest ok. 20% genów (PLEWCZYŃSKI i RYCHLEWSKI 2009).

Pomocnicze zadania bioinformatyki przy określaniu genów docelowych dla leków, to określanie relacji między genem będącym potencjalnym „celem” leku a jego homologami u człowieka oraz w organizmach modelowych (HOLBROOK i SANSEAU 2007). Wiedza na ten temat pozwala, z jednej strony, odpowiednio zaplanować projektowanie środków interwencji terapeutycznej (zwykle lek ma uderzać specyficznie w wybrane białko docelowe, ale nie w jego bliskie homologi). Z drugiej strony, wiedza o relacjach ortologii między człowiekiem a np. myszą, pozwala właściwie planować i interpretować doświadczenia na zwierzętach. Jeśli np. u myszy, ludzkie białko docelowe ma więcej niż jeden bliski homolog, to interpretacja doświadczeń na myszach ze znokautowanym jednym z tych genów jest problematyczna.

Pierwszym przybliżeniem listy możliwych białek docelowych dla leków w danej chorobie jest zatem wyliczenie wszystkich genów, w które można „uderzyć” lekiem (ang. drug-gable), i które przy tym podlegają ekspresji

w chorej tkance lub w organizmie patogenu (np. bakterii chorobotwórczej). Taka lista przedstawia sobą jednak znikomą wartość, o ile białka te nie będą poparte hipotezami dotyczącymi ich zaangażowania w proces chorobowy bądź w istotne procesy życiowe patogenu. Hipoteza ta, z jednej strony, uprawdopodobnia dane białko jako „cel terapeutyczny”, a z drugiej strony, umożliwia opracowanie metody sprawdzenia zasadności wykorzystania danego białka jako punktu docelowego dla leku. Hipoteza dotycząca zaangażowania białka w proces chorobowy może być zbudowana na przykład na podstawie zmian ekspresji genu w chorych tkankach, w porównaniu ze zdrowymi, albo zmian ekspresji genu w modelowych układach hodowli tkankowych.

Istotną trudnością w budowaniu hipotez terapeutycznych jest fakt, że zmiany w zachowaniu genów „w chorobie” mogą dotyczyć czterech grup genów: A) genów, których „złośliwe” działanie „napędza” proces chorobowy, B) genów, które są zaangażowane w zwalczanie choroby, C) genów, dla których korelacje ich właściwości z chorobą są efektem ubocznym choroby lub D) te korelacje są przypadkowe.

Ponieważ, jak mówiliśmy wcześniej, uważa się, iż za choroby odpowiedzialne jest zwykle działanie grup genów („ścieżek”), „celem” leku powinna być raczej ścieżka, niż gen. A ponieważ ze względów praktycznych łatwiej „uderzyć” w pojedynczy gen, jako genów docelowych należy używać niekoniecznie tych genów, które najwyraźniej „ujawniają się” w eksperymentach, ale tych, których modulowanie działania może najskuteczniej modulować działanie całej ścieżki. Zatem, analizując doświadczenia, których celem jest poznanie mechanizmów choroby i wybór punktu interwencji, poszukuje się zwykle nie genów, których właściwości korelują się z chorobą, ale ścieżek (powiązanych funkcjonalnie grup genów), które ulegają zmianom w chorobie. Na przykład, zamiast poszukiwać genów o nadekspresji w chorobie, poszukuje się ście-

żek, dla których istotna część genów podlega nadekspresji w chorobie. Istotne są tu uporządkowane anotacje funkcjonalne genów, czyli opis funkcji biologicznej i molekularnej genów według pewnego uporządkowanego słownika terminów biologicznych. Zwykle stosuje się tu słownik Ontologii Genów (<http://www.geneneontology.org>) (OSBORNE i współaut. 2007). Dalej, stosuje się tu narzędzia typu BERS, aby znaleźć grupy genów zmienione w chorobie – albo znane ścieżki, albo też nieopisane wcześniej grupy genów powiązane zależnościami funkcjonalnymi. Przykładem takiej analizy jest praca dotycząca ścieżek sygnalizacyjnych związanych z jednym z najgroźniejszych nowotworów, rakiem trzustki (JONES i współaut. 2008).

Jeśli znalezione ścieżki pasują do wiedzy o efektach biologicznych związanych z chorobą, można w ścieżkach poszukiwać genów docelowych. Nie muszą przy tym być to geny, których działanie w czasie choroby zmienia się najbardziej, ale takie, które są kluczowe dla funkcjonowania ścieżki. Do poszukiwania punktów uderzenia w obrębie wybranej ścieżki można użyć, o ile ścieżka jest dobrze przebadana, zaawansowanych narzędzi modelowania złożonych układów. Na przykład, dla ścieżki sygnalizacyjnej receptora nabłonkowego czynnika wzrostu (białko EGFR), istotnej w wielu chorobach, m. in. w nowotworach, zbudowano model pozwalający symulować przesyłanie sygnałów od receptorów, przez kaskadę kinaz, do czynników transkrypcyjnych (WILEY i współaut. 2003). Model taki umożliwia wybór i porównanie różnych alternatywnych punktów interwencji w ścieżce. Dzięki temu można próbować wybrać taki punkt interwencji, który z jednej strony będzie najskuteczniej hamował niepożądany efekt biologiczny, a z drugiej – będzie obciążony jak najmniejszym ryzykiem wystąpienia efektów ubocznych. Zarazem można próbować wybrać białko docelowe, które z chemicznego punktu widzenia rokuje największe szanse na szybkie znalezienie skutecznego modulatora funkcji, np. inhibitora.

#### OPRACOWANIE ODPOWIEDNIEGO ŚRODKA INTERWENCJI – CZYLI SZUKANIE KANDYDATA NA LEK

Gdy wybrany jest punkt interwencji, czyli białko istotne dla pewnego związanego z chorobą procesu biologicznego, przystępuje się do poszukiwania środka interwencji. Z reguły odbywa się to metodą wielkoskalowe-

go doświadczalnego testowania (ang. high-throughput screening) tysięcy związków chemicznych. Często jednak na tym etapie stosuje się metody obliczeniowe – chemii obliczeniowej i bioinformatyki strukturalnej. Jeżeli

trójwymiarowa struktura białka docelowego dla leku jest znana, można zastosować testowanie wirtualne (ang. virtual screening), polegające na poszukiwaniu metodami obliczeniowymi takich związków chemicznych, które najkorzystniej zwiążą się z białkiem docelowym. Procedura zwana dokowaniem (ang. docking), wymaga oszacowania fizycznych oddziaływań na poziomie atomowym dla każdej testowanej cząsteczki w wielu możliwych ułożeniach względem białka, z uwzględnieniem różnych możliwych konformacji samej cząsteczki oraz z uwzględnieniem możliwości lokalnego dopasowania struktury białka do „dokowanej” cząsteczki. Rozwinięciem tej metodologii jest dokowanie do wielu „celów” – w przypadku, gdy pożądanym jest lek, który blokowałby kilka białek docelowych. Taką strategię stosowano np. tworząc lek przeciwko wirusowi HIV. Dzięki temu lek skuteczniej uderza w proces chorobowy (tu, w wirusa) i zmniejsza się ryzyko, że wskutek szybkiej ewolucji patogenu pojawią się szczepy na lek odporne (JENWITHEESUK i współaut. 2008). Oczywiście znalezione metodami dokowania cząsteczki traktuje się jedynie jako punkty wyjściowe do testów doświadczalnych, metoda ta pozwala jednak zwykle na istotne skrócenie czasu badań przez znaczne zmniejszenie liczby testowanych związków chemicznych. Metody obliczeniowe są też rutynowo stosowane przy ulepszaniu znalezo-

nego doświadczalnie związku chemicznego, ponieważ pozwalają zaproponować modyfikacje struktury poprawiające wiązanie się do białka.

Jeżeli trójwymiarowa struktura białka docelowego dla leku nie jest znana, można spróbować zbudować model tej struktury, który, jeśli jest wystarczająco dokładny, może posłużyć do dokowania (CAVASOTTO i PHATAK 2009). Techniki przewidywania struktur białek omówiono szeroko np. w pracy BUJNICKIEGO (2005).

Budowa modelu struktury może też być użyta do przewidzenia lub zracjonalizowania mechanizmu działania białka, o którym z innych danych wiadomo, że jest zaangażowane w istotne procesy chorobowe (PROELL i współaut. 2008). Jeżeli o modelu struktury potencjalnego białka docelowego z góry wiadomo, że jest niezbyt dokładny, może mimo to być użyty do przewidzenia ogólnej funkcji molekularnej białka, co umożliwi ocenę, czy białko to może być przydatne jako „cel” dla leku. Analiza przewidzianej struktury pozwala również ocenić prawdopodobieństwo znalezienia skutecznego modulatora funkcji białka, innymi słowy, ocenić, czy w białko łatwo da się uderzyć lekiem. W związku z tym, modelowanie struktury ułatwia wskazanie potencjalnych białek docelowych dla leków do dalszych badań doświadczalnych.

#### INNE ZASTOSOWANIA METOD OBLICZENIOWYCH W PROCESIE PROJEKTOWANIA LEKÓW

Poza ramy niniejszego artykułu, a także poza ramy „klasycznej” bioinformatyki wykacza wiele zastosowań metod obliczeniowych, np. metody przewidywania toksyczności leku, metody przewidywania jego metabolizmu czy dostępności w tkance.

Jednym z nowszych zastosowań bioinformatyki w przemyśle farmaceutycznym jest badanie zależności między zmiennością międzyosobniczą a skutecznością leku. Bioinformatyka jest np. nieodłącznym narzędziem farmakogenomiki, która zależności te analizuje na poziomie polimorfizmów nukleotydowych w genie docelowym dla leku oraz na poziomie zmian ekspresji genów w reakcji na lek. Badając ekspresję genów po podaniu leku, można lepiej zrozumieć molekularny mechanizm działania leku, a także mechanizmy ewentualnych szkodliwych efektów ubocznych (GRESHAM i McLEOD 2009). Te same badania, a także analizy mutacji w

genie docelowym, pozwalają na zrozumienie, dlaczego niektóre leki są skuteczne tylko w subpopulacjach pacjentów. Na przykład lek przeciwnowotworowy gefinitib, uderzający w receptor nabłonkowego czynnika wzrostu (białko EGFR), jest, jak się okazało, bardzo skuteczny, ale tylko w niewielkiej podgrupie pacjentów (ok. 10% w niektórych populacjach). Mechanizm odpowiedzi na ten lek udało się zrozumieć, gdyż koreluje się ona z występowaniem pewnych mutacji w białku docelowym, a ich znaczenie dla funkcji białka można zrozumieć analizując warianty struktury związane z mutacjami (HEIST i CHRISTIANI 2009). Takie badania prowadzą do tzw. medycyny spersonalizowanej, czyli leczenia z uwzględnieniem genetycznej tożsamości pacjenta. Lekarz, przepisując lek spersonalizowany, kieruje się wynikami testu diagnostycznego, który pozwala ocenić prawdopodobieństwo, że pacjent odpowie pozy-

tywnie na lek. Już dziś stosuje się takie podejście w terapiach przeciwnowotworowych

nakierowanych na receptor HER2 (DOWSETT i DUNBIER 2008).

## PODSUMOWANIE

Już od ponad dekady bioinformatyka jest nieodłącznym elementem procesu tworzenia nowych leków. Trudno ocenić w mierzalny sposób znaczenie jednej dziedziny nauki w tak złożonym procesie, ale może tu pomóc uproszczona analiza bibliometryczna. Z ponad dwunastu tysięcy publikacji dotyczących odkrywania leków, jakie można znaleźć w głównej literaturowej bazie danych Medline (<http://www.ncbi.nlm.nih.gov/pubmed/>), 15% dotyczy (wg tytułu lub streszczenia) bioinformatyki bądź metod obliczeniowych. Jest to oszacowanie z dołu, gdyż bardzo często nie wymienia się wprost istotnych metod badawczych.

Proces odkrywania leków, jak cała biologia, podlega bardzo szybkiej ewolucji, związanej z niezwykle szybkim rozwojem wiedzy i metod. Trudno przewidzieć wszystkie wyzwania przyszłości, jakie stoją przed bioinformatyką, ale niewątpliwie należy do nich postęp farmakogenomiki oraz analizy wielu genomów człowieka i mikrobów w powiązaniu z problemami medycznymi. Farmakogenomika będzie wymagała analiz danych wielkoskalowych, np. danych ekspresji genów w kontekście danych klinicznych dla dużych

populacji pacjentów (GRESHAM i MCLEOD 2009). Podobnie, sekwencjonowanie wielu genomów ludzkich (PENNISI 2008) będzie umożliwiało analizy podatności na choroby, odpowiedzi na leki i inne analizy typowo medyczne w kontekście całego genomu. Ponadto, setki gatunków bakterii żyjących w organizmie nawet zdrowego człowieka (O'KEEFE 2008), ma niewątpliwie wpływ na stan zdrowia gospodarza, a zarazem jest obrazem tego stanu. Kolejnym wielkim wyzwaniem biologii, a w tym i bioinformatyki, jest zrozumienie zależności między stanem tej populacji mikroorganizmów (mikrobiomu) a zdrowiem ludzkiego gospodarza (MAI i DRAGANOV 2009) i wykorzystanie tej wiedzy do celów terapeutycznych i diagnostycznych. Ostatnim wyzwaniem dla bioinformatyki, o jakim tu wspomniemy, jest zastosowanie biologii systemowej w procesie tworzenia nowych leków. Będą to nie tylko wspomniane wcześniej i stosowane już dziś analizy danych wielkoskalowych w kontekście sieci zależności między obiektami biomedycznymi, ale uwzględnianie w analizach danych biomedycznych całości komórki oraz populacji komórek (SMITH i współaut. 2009).

## BIOINFORMATICS IN SEARCH OF NOVEL DRUGS

### Summary

In the process of novel drugs creation, today's pharmaceutical industry applies a whole spectrum of advanced research methods for elucidation of disease mechanisms, finding and selection of the right points of therapeutic intervention, designing a proper means of intervention (drug candidate), evaluation of safety and efficacy of the candidate and evaluation of the drug in clinical trials. At each

stage of this complicated process, the computational techniques of biology, herein collectively called "bioinformatics", are an indispensable element of the research toolkit. In this article, are discussed the most common applications of bioinformatics, and challenges posed by the development of biology and medicine, as well as the evolving model of the drug discovery process

## LITERATURA

- BUJNICKI J. M., 2005. *Przewidywanie struktury białek: podejście boltzmannowskie i darwinowskie*. Kosmos 54, 155-162.
- CAVASOTTO C. N., PHATAK S. S. 2009. *Homology modeling in drug discovery: current trends and applications*. Drug Discov Today.
- COUZIN-FRANKEL J., 2009. *Genetics. The promise of a cure: 20 years and counting*. Science 324, 1504-1507.
- DOWSETT M., DUNBIER A. K., 2008. *Emerging biomarkers and new understanding of traditional markers in personalized therapy for breast cancer*. Clin. Cancer Res. 14, 8019-8026.
- EKINS S., NIKOLSKY Y., BUGRIM A., KIRILLOV E., NIKOLSKAYA T., 2007. *Pathway mapping tools for analysis of high content data*. Methods Mol. Biol. 356, 319-350.

- GRESHAM V., MCLEOD H. L., 2009. *Genomics: applications in mechanism elucidation*. *Adv. Drug Deliv. Rev.* 61, 369–374.
- HACKETT N. R., HEGUY A., HARVEY B. G., O'CONNOR T. P., LUETTICH K., FLIEDER D. B., KAPLAN R., CRYSTAL R. G., 2003. *Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers*. *Am. J. Respir. Cell. Mol. Biol.* 29, 331–343.
- HEIST R. S., CHRISTIANI D., 2009. *EGFR-targeted therapies in lung cancer: predictors of response and toxicity*. *Pharmacogenomics* 10, 59–68.
- HOFFMANN R., VALENCIA A., 2003. *Life cycles of successful genes*. *Trends Genet.* 19, 79–81.
- HOLBROOK J. D., SANSEAU P., 2007. *Drug discovery and computational evolutionary analysis*. *Drug Discov. Today*. 12, 826–832.
- JENWITHEESUK E., HORST J. A., RIVAS K. L., VAN VOORHIS W. C., SAMUDRALA R., 2008. *Novel paradigms for drug discovery: computational multitarget screening*. *Trends Pharmacol. Sci.* 29, 62–71.
- JONES S., ZHANG X., PARSONS D. W., LIN J. C., LEARY R. J., ANGENENDT P. i współaut., 2008. *Core signaling pathways in human pancreatic cancers revealed by global genomic analyses*. *Science*. 321, 1801–1806.
- MAI V., DRAGANOV P. V., 2009. *Recent advances and remaining gaps in our knowledge of associations between gut microbiota and human health*. *World J. Gastroenterol.* 15, 81–85.
- O'KEEFE S., J. 2008. *Nutrition and colonic health: the critical role of the microbiota*. *Curr. Opin. Gastroenterol.* 24, 51–58.
- OSBORNE J. D., ZHU L. J., LIN S. M., KIBBE W. A., 2007. *Interpreting microarray results with gene ontology and MeSH*. *Methods Mol. Biol.* 377, 223–242.
- PAWŁOWSKI K., 2008. *Uncharacterized/hypothetical proteins in biomedical 'omics' experiments: is novelty being swept under the carpet?* *Brief Funct. Genomic Proteomic* 7, 283–290.
- PAWŁOWSKI K., LEPISTO M., MEINANDER N., SIVARS U., VARGA M., WIESLANDER E., 2006. *Novel conserved hydrolase domain in the CLCA family of alleged calcium-activated chloride channels*. *Proteins* 63, 424–439.
- PENNISI E., 2008. *Personal genomics. Number of sequenced human genomes doubles*. *Science* 322, 838.
- PLEWCZYŃSKI D., RYCHLEWSKI L. 2009. *Meta-basic estimates the size of druggable human genome*. *J Mol Model.* 15, 695–699
- PROELL M., RIEDL S. J., FRITZ J. H., ROJAS A. M., SCHWARZENBACHER R., 2008. *The Nod-like receptor (NLR) family: a tale of similarities and differences*. *PLoS One* 3, e2119.
- SAMS-DODD F., 2006. *Drug discovery: selecting the optimal approach*. *Drug Discov. Today* 11, 465–472.
- SHEKHAR C., 2008. *In silico pharmacology: computer-aided methods could transform drug development*. *Chem. Biol.* 15, 413–414.
- SMITH P. J., KHAN I. A., ERRINGTON R. J., 2009. *Cytomics and cellular informatics-coping with asymmetry and heterogeneity in biological systems*. *Drug Discov. Today* 14, 271–277.
- WILEY H. S., SHVARTSMAN S. Y., LAUFFENBURGER D. A., 2003. *Computational modeling of the EGF-receptor system: a paradigm for systems biology*. *Trends Cell. Biol.* 13, 43–50.