

## Wprowadzenie do biologii systemów

Stosowana dotychczas w naukach biologicznych redukcjonistyczna analiza komórki, jako podstawowej jednostki organizmów żywych, dostarczyła ogromnej ilości danych o strukturze i częściach składowych komórki oraz o ich funkcji. „Mimo tego wielkiego sukcesu [biologii molekularnej] staje się jednak coraz bardziej jasnym, że dyskretne funkcje biologiczne tylko w rzadkich przypadkach mogą być przypisane indywidualnym cząsteczkom” (Barabási i Oltvai, 2004). Wszystkie funkcje komórki dzieją się i zmieniają w czasie życia komórki, w określonych jej przedziałach i wynikają z aktywności i interakcji złożonych kompleksów różnorodnych cząsteczek (białka, DNA, RNA, metabolity itd.). Kompleksy takie tworzą wysoce złożone, interaktywne dynamiczne sieci i systemy. Uzyskiwanie informacji o ilościowych i jakościowych cechach sieci, o aktywności sub-sieci i modułów oraz ich rozmieszczeniu w komórce pozwoli na budowanie modeli przybliżających nas do pełniejszego rozumienia, jak funkcjonują systemy żywe. Redukcjonizm usiłował analizować ustroje (systemy) żywe jako „fizyczną rzeczywistość” i oczywiście odegrał znaczącą rolę w biologii. Istotnie, struktura i funkcja ustrojów żywych opierają się na fizyce i chemii, ale w miarę ich ewolucji narastają nowe poziomy złożoności i wyłaniają się (emergencja) nowe, zupełnie nieprzewidywalne właściwości, (Morowitz, 2002; Koj, 2006). Nauka zajmująca się badaniem ustrojów żywych, czyli jak mówimy, systemów żywych, przechodzi okres niepewności lub nawet kryzysu. Wynika to z pewnego rozczarowania, jakie w ostatnim półwieczu wiązano z postępami w zakresie badań nad sekwencją, strukturą i aktywnością genomów i genów (DNA), zwłaszcza z danych uzyskanych z Human Genome Project. Badania te tylko częściowo spełniły wielkie nadzieje, roztaczane jeszcze parę dziesiątków lat temu. Jeszcze niedawno sądzono, że poznanie genów i genomu wyjaśni nam wszelkie problemy dziedziczności, rozwoju osobniczego, ewolucji, umożliwi poznanie istoty chorób, pozwoli na ich diagnostykę i skuteczne leczenie. Poznanie genomu człowieka miało nam wyjaśnić istotę człowieczeństwa. Tymczasem okazało się, że genom szympansa różni się od genomu człowieka tylko o 1,5%, mimo bardzo istotnych różnic fenotypowych. Oznacza to, że geny „same z siebie” nie determinują fenotypu i że po za genami musi istnieć jeszcze inna „informacja”, „program” lub „system” (Fox Keller, 2005; Lewontin, 2001). W genocen-

trycznym modelu życia często pomijano wpływ na aktywność genową czynników epigenetycznych i środowiskowych (Gottlieb, 2000).

### **Interakcje cząsteczek są podstawą życia**

Wszelkie przestrzenie wnętrza komórki wypełnione są zawieszonymi w wodnym środowisku sieciami kompleksów białkowych i innych makromolekuł, „pływającymi” wśród bogactwa organelli komórkowych, między strukturami błonowymi, systemem cytoszkieletu oraz wewnątrz struktur, tworzących zamknięte przedziały (kompartamenty). Organelle i inne elementy strukturalne komórki są ściśle zintegrowane z systemem sieci. Niektóre kompleksy białkowe udało się już wyizolować i poznać wiele detali z ich struktury i funkcji. Inne kompleksy są bardzo labilne, przejściowe i dynamiczne, trudne do izolowania i analizy. Osobnym problemem jest lokalizacja takich kompleksów w komórce i ich zachowanie w „czasoprzestrzeni”, co oczywiście ma pierwszorzędne znaczenie dla procesów życiowych.

Kartezjański pogląd na systemy żywe głoszący, że organizm jest w istocie podobny do mechanizmu maszyny, zakładał, że złożoność istot żywych można będzie wyjaśnić przez rozłożenie ich na części, tak jak rozkłada się na części mechanizm zegara. Przypuszczano, że te mniejsze części „składowe” pozwolą się analizować, a wyniki złożyć w całość. Trudność takiego podejścia polega na tym, że systemy żywe przy każdej próbie analizy redukcjonistycznej zwykle ulegają rozsypaniu się na części, tracąc natychmiast swoje immanentne cechy charakteryzujące zjawisko życia. Ponowne „złożenie” ich wcale nie odtwarza pierwotnej funkcjonalnej całości. Poznanie części nie oznacza poznania całości, bo żywe organizmy nie są prostą sumą części, co wynika ze zjawiska emergencji.

„Życie po rozkładzie na części przestaje być życiem i jako takie nie może być dalej badane” (Cramer, 2001). Powszechnie stosowane w biochemii rozdrobnienie komórek na homogenną masę daje nam materiał, w którym fenomen życia już nie istnieje. Dla badania systemu żywego nie można wprost zastosować klasycznych pojęć, terminów i podejść fizycznych, takich jak przyczynowość, liniowość, ekstrapolacja lub przewidywalność, używanych w naukach technicznych i fizyce (Cramer, 2001). Myślenie kategoriami stosowanymi dla badań zjawisk fizykochemicznych w świecie nieożywionym nie może być przenoszone bez ograniczeń na istoty żywe. Redukcjonistyczne podejście do fenomenu życia narzuciło uproszczony pogląd na funkcjonowanie komórki i organizmów żywych i wydaje się odpowiedzialne za fakt, że obecnie naukę o życiu na poziomie molekularnym cechuje pewna naiwność (Strand, 2000).

Natomiast, redukcjonistyczna analiza fenomenu życia, poprzez stały rozwój i doskonalenie metod, doprowadziła opis struktur i funkcji molekularnych komórki niemal do perfekcji. Dzięki temu podejściu w ostatnim dwudziestoleciu biologia molekularna,

a zwłaszcza „genomika funkcjonalna”, przyniosła wielki ładunek precyzyjnych informacji o strukturze i funkcjach makrocząsteczek. Porządkowanie, bankowanie i obróbka tych informacji stała się możliwa jedynie dzięki równoległemu rozwojowi technik komputerowych.

Ale mimo nagromadzenia olbrzymiej ilości informacji o organizacji struktur komórkowych i ich funkcji oraz mimo poznania budowy i funkcji makrocząsteczek (DNA, RNA, białko) wciąż nie w pełni wiemy, jak działa system żywy (istota żywa). Uzyskanie pełnego obrazu fenomenu życia (nawet w organizmach jednokomórkowych) i opisanie jego fizycznych podstaw ujętych całościowo jest wciąż odległą, jeśli w ogóle osiągalną, perspektywą, nawet na poziomie komórki. Jednak niektóre złożone systemy molekularne, zwłaszcza dobrze poznane biochemiczne reakcje metaboliczne, kinetyka enzymów i metaboliczne mechanizmy kontrolne, działające w otoczeniu wielu zmiennych parametrów, mimo swojej wielkiej złożoności i mimo zjawisk emergentnych, dają się oceniać przez ilościową, numeryczną analizę, co stanowi inspirację do konstruowania modeli matematycznych oraz poszukiwania sposobów bardziej całościowego oglądu układów funkcjonujących w komórce. **W żywej komórce nie istnieje zjawisko autonomii genów, białek lub innych cząsteczek. Nie ma też prostego, liniowego przełożenia między strukturą i funkcją genów a fenotypem.** Fakty te są często pomijane w badaniach biomedycznych i agrobiologii.

Wszelkie procesy zachodzące w żywej komórce są ze sobą powiązane. Tysiące białek kodowanych przez geny, ale jednocześnie wciągniętych w proces kontroli ich aktywności, współdziałających ze sobą, nadzorujących różne procesy fizjologiczne w komórkach, tkankach i narządach, reagujących na zmieniające się w czasie zewnętrzne i wewnętrzne środowisko, stanowią gmach życia. **Podstawą życia** (także stanów patologicznych) **są dynamiczne interakcje między cząsteczkami**, odbywające się za pośrednictwem licznych, niekowalencyjnych wiązań (Bray, 2003).

Wyzwaniem dla współczesnej biologii jest zrozumienie, **jak interakcje wielkiej liczby cząsteczek determinują funkcje komórki i organizmu.** Interakcje typu białko-białko (PPI, ang. *protein-protein interactions*) istnieją zarówno w szlakach przekazywania sygnałów ze środowiska zewnętrznego do komórki, jak i w obrębie samej komórki, np. w dobrze już poznanych, enzymatycznych szlakach metabolicznych, procesach kontrolujących funkcje genów, w procesie podziału komórkowego czy sekrecji, w szlakach prowadzących do programowanej śmierci komórki, a także w wielu innych. Cząsteczki białka stanowią podstawowy składnik sieci, jako enzymy katalizują różne reakcje biochemiczne, wraz z innymi cząsteczkami budują struktury błoniaste i organelle komórkowe, są składnikiem chromatyny oraz pełnią wiele innych podstawowych funkcji. Są składnikiem maszyny transkrypcyjnej i translacyjnej, biorą udział w replikacji i naprawie DNA, w przekazywaniu sygnałów, są receptorami, transporterami, stanowią

pompy jonowe czy „motory molekularne”, są elementami układu immunologicznego. Białka na ogół nie działają „w pojedynkę”, lecz w zespole, zmieniają partnerów, tworząc z innymi białkami i makrocząsteczkami (DNA, RNA i in.) labilne, dynamiczne układy, w których występują jedynie słabe wiązania chemiczne.

Sieci i funkcjonalne systemy komórkowe są przedmiotem rozwijającej się nowej dyscypliny naukowej – biologii systemów. Biologia systemów wywodzi się z dwóch historycznych korzeni: z epokowego odkrycia natury materiału genetycznego (DNA) i wyjaśnienia szlaków biochemicznych, ich kontroli zwrotnej oraz teorii równowagi termodynamicznej (Westerhoff i Palsson, 2004).

Biologia systemów dąży do integracji informacji otrzymanych z doświadczeń, uzyskanych na drodze analizy redukcjonistycznej, a jednocześnie akcentuje mocno powstawanie złożoności (ang. *complexity*) na coraz wyższych hierarchicznych poziomach organizmów żywych. Terminu „system” używamy dla określenia grupy dwóch lub więcej części (komponent) współdziałających ze sobą, zebranych w sieć w taki sposób, że tworzą funkcjonalną całość (Trewaras, 2006). Terminy „system” i „sieć” używane są zamiennie. Termin „system” jest jednak pojemniejszy, obejmując sieci komórkowe, ale również wewnętrzną organizację organizmów i organizację całych społeczności organizmów (zwierząt, roślin). Mówimy zatem o systemach społecznych u ludzi, systemach w ekologii, systemach w gospodarce itd. Każdy system biologiczny i każda sieć składa się z podsystemów i ma układ hierarchiczny. Podsystemy są ze sobą powiązane i wzajemnie od siebie zależne, choć niekiedy ta zależność jest luźna.

Prawa fizyki są niewystarczające dla opisu zarówno kształtu, jak i organizacji systemów żywych, które wyłoniły się w toku ewolucji. Immanentną cechą systemów żywych jest zdolność do syntezy makrocząsteczek dzięki sprzężeniu reakcji endoergicznych i egzoergicznych, korzystanie z informacji genetycznej oraz reakcja na sygnały środowiskowe (zewnątrz- i wewnątrzkomórkowe). Gdy tylko organizacja (system) biologiczna się wyłoni, narzuca ona ograniczenia i zachowanie się jej komponentów (części). Systemy biologiczne („żywe”) działają z dala od stanu równowagi termodynamicznej (w niskiej entropii) dzięki stałemu dopływowi z zewnątrz energii wprost, jak w przypadku fosforylacji fotosyntetycznej czy w postaci materiału odżywczego. „Trajektorie” w złożonych systemach żywych są dynamiczne, podlegają rytmom, oscylacjom, sprzężeniom regulatorowym i są nieprzewidywalne, labilne, choć utrzymane w ładzie. Funkcjonowanie żywej komórki nie daje się wyjaśnić prostymi liniowymi zależnościami między cząsteczkami. Biologia systemów dostarczyła wielu dowodów, że w komórce funkcjonują układy dające się opisać teorią deterministycznego chaosu. Nawet podstawowy proces przepływu informacji z genów (DNA) do białek nie ma cech procesu liniowego (por. Chorąży, 2009), lecz daje się opisać przez nieliniowe modele, takie jak chaos deterministyczny (np. por. Jura i in., 2006).

Systemy biologiczne charakteryzuje dynamizm: mają one immanentną, organiczną zdolność do samoorganizacji. Ewolucja życia zaczęła się od prostych cząsteczek chemicznych, które w wyniku interakcji między sobą wytwarzały coraz bardziej złożone makrocząsteczki. Powstała przed paru dziesiątkami lat i intensywnie rozwijana chemia supramolekularna dostarcza nam modelowych przykładów, w jaki sposób cząsteczki chemiczne generują wielkie makrocząsteczki i ich kompleksy o złożonej architekturze. Klasyczna chemia organiczna poznawała i wykorzystywała właściwości atomów i silnych kowalencyjnych wiązań między innymi dla syntezy złożonych cząsteczek (Lehn, 2002; Lehn 2004). Supracząsteczki i ich różnorodne struktury powstają z cząsteczek chemicznych dzięki labilnym, niekowalencyjnym wiązaniom. Materia złożona z supracząsteczek jest w swej naturze dynamiczna, gdyż jej jednostki składowe (cząsteczki chemiczne, jony metali) podlegają stale procesom składania, rozpadu, reorganizacji i ponownego składania w zależności od warunków. Niekowalencyjne wiązania i siły biorące udział w interakcji supracząsteczek należą do oddziaływań elektrostatycznych i elektromagnetycznych, wiązań wodorowych, sił Van der Waalsa i in. Chemia supramolekularna uutorowała drogę dla koncepcji „informacji” w chemii, którą należy rozumieć jako immanentną wewnątrzcząsteczkową właściwość, dzięki której cząsteczki spontanicznie oddziałują między sobą i samorzutnie tworzą złożone struktury (Lehn 2004, Lehn 2007). Proces samoorganizacji supracząsteczek chemicznych jest analogiczny do zjawisk obserwowanych w żywej komórce. Głębsze poznanie zjawisk w chemii supracząsteczek przybliży nas do lepszego zrozumienia zjawisk toczących się w złożonej materii żywych organizmów.

Dynamiczne struktury sieci w komórce są pod wpływem działania sił (atraktora), które powodują utrzymanie ich w określonym stanie odległym od stanu równowagi termodynamicznej. W świecie ożywionym systemy stale się zmieniają i przystosowują do środowiska (warunków, kontekstu). Wytrącenie systemu z naturalnych, charakterystycznych dla niego stanów, może zajść w wyniku zadziałania nowych sił pochodzących np. z mikrośrodowiska i pojawienia się odmiennego atraktora.

### **Podstawy budowania modeli, grafy**

Graficzne obrazy sieci w biologii systemów budowane są na podstawie informacji o międzycząsteczkowych oddziaływaniach białko-białko (PPI), a także białko-DNA, białko-RNA, białko-metabolit, białko-substrat.

Mimo rozwinięcia w ostatnim dziesięcioleciu nowych metod doświadczalnych, biologia nie ma jeszcze odpowiednich narzędzi i metod badania złożonych systemów i dużych sieci. Dotychczasowe metody służące do uzyskania informacji o interakcjach makrocząsteczek, takie jak: mikromacierze peptydowe i DNA, spektrometria mas, technika dwuhybrydowa z użyciem drożdży (Y2H, ang. *yeast two hybrids*), dedukcja oddziaływań na podstawie analizy badań genomycznych i in., stale rozszerzają możliwości badawcze,

ale wymagają ciągłych uzupełnień (Bork, 2004). Doświadczalna biologia systemów chętnie stosuje technikę Y2H jako stosunkowo tania i ogólnie dostępną, a obecnie zautomatyzowaną. Opracowane są nowe metody badań procesów zachodzących w żywej komórce zmierzające do znakowania pojedynczych cząsteczek białkowych i śledzenia ich lokalizacji, interakcji i przemieszczeń w komórce. Duże nadzieje budzą węglowe nanocząsteczki, które sprzężone z metalami tworzą nowe typy znaczników (np. „kropki kwantowe”, ang. *quantum dots* – QD; „nano-łuski”, ang. *nano-shells* (Nepal i Geckeler, 2006). Dla śledzenia losów makrocząsteczek w żywej komórce budowane są nowe typy mikroskopów optoelektronicznych, rozwijają się metody microPET (ang. *positron emission tomography*) itp. Informacje o takich nowszych technikach znajdują się w przeglądowej pracy Kherlopian i in. (2008). Dotychczas zebrane informacje o strukturze, funkcji i interakcji makrocząsteczek zostały zebrane w dziesiątkach wielkich baz danych (np. STRING, PLEX, Bioverse, Predictome, BIND, GRID, Ospray, Swiss-Prot, WormBase i wielu innych), których powstanie było możliwe dzięki szybkiemu rozwojowi technik komputerowych. Bazy danych służą do konstrukcji sieci, gdy interesujące nas białka/geny zostaną wprowadzone do odpowiedniego programu komputerowego. Sieci (mat. grafy) podlegają pewnym rygorom matematycznym i uniwersalnym prawom, które opisują związki, relacje i połączenia między jednostkami „tkanki” sieciowej. Prawa te są uniwersalne, ponieważ obowiązują zarówno w sieciach molekularnych komórki, jak i w społecznościach zwierząt i ludzi, w ekologii i ekonomii. W społeczeństwie ludzkim reguły obowiązujące w strukturze i zachowaniach sieci (np. sieci gospodarcze, komunikacyjne) podlegają jednak modulacjom przez naturalne czynniki przypisane gatunkowi ludzkiemu, takie jak: świadomość, odpowiedzialność, altruizm, wola działania w kategoriach sprawiedliwości społecznej itd. Grafy (sieci) interakcji makrocząsteczek komórkowych uzyskane z baz danych są w pewnym sensie sztuczne i statyczne, ale są pomocne w formowaniu naszych pojęć o mechanizmach działających w żywych systemach w ujęciu syntetycznym, całościowym.

Wielkie sieci komórkowe są rozległe topologicznie i obejmują liczne złożone funkcje. Dla uproszczenia wyodrębnia się z nich podsieci i „funkcjonalne genowe moduły”, co pozwala poznać i porównać interakcję białek z poziomem ekspresji genów. U drożdży opisano np. sieć PPI z 2633 białek i związek jej z 520 genami. Spośród tych genów wyodrębniono moduł o 99 genach biorących udział w biogenezie rybosomów (Parkkinen i Kaski, 2010). Takie podsieci i moduły opracowywane są np. dla opisu metabolizmu, transkrypcji, szlaków sygnałnych. Opracowaniom struktury wielkich sieci towarzyszy rozkwit nowych terminów, różnych „-omics” (*proteomics, metabolomics, transcriptomics, interactomics, interferomics* itd.).

W białkach zostały zidentyfikowane krótkie sekwencje aminokwasowe (motywy, domeny) oddziałujące między sobą. Przykładem takich domen są: bogata w prolinę dome-

na SH3 (ang. *Src homologous domain 3*), domena EH (ang. *Eps homologous*), wiążąca się z rodziną małych cząsteczek GTPaz, domena WW zawierająca sekwencję ok. 40 aminokwasów, zwijająca się w trójpasnową wstążkę typu beta harmonijka (*beta sheet*), i chętnie wchodząca w interakcję z motywami proliny lub fosfoseryny różnych białek. Domeny takie służą do tworzenia kompleksów z innym białkiem (lub z odpowiednią sekwencją sygnałną np. w DNA, RNA) i biorą udział w konstrukcji sieci (Tong i in., 2002).

Informacje o interakcjach makrocząsteczek zawarte w dostępnych bazach danych nie są pełne oraz całkowicie wiarygodne, a dane pozostają często niejednoznaczne, ponieważ wyniki pochodzą z wielu laboratoriów, gdzie doświadczenia były prowadzone w niejednorodnych warunkach. Standaryzacja metod, uśrednianie wyników oraz programy sprawdzające i weryfikujące wyniki laboratoryjne przyniosą zapewne polepszenie jakości tych danych. Oczywiście stale jeszcze nie mamy pewności, jak dalece warunki oddziaływań w żywej komórce są podobne do warunków stworzonych w modelach doświadczalnych. Ponadto, co podkreśla Kitano (Kitano, 2000), budowanie diagramów sieci ilustrujących interakcje makrocząsteczek jest jedynie pierwszym krokiem i wstępem do poznania struktury dynamicznych układów i ich funkcji oraz zrozumienia mechanizmów regulacyjnych komórki.

Na podstawie danych o interakcji między jednostkami sieci, np. białka „x” z białkiem „y”, a białka „y” z białkiem „z” itd., matematycy i bioinformatycy opracowują matematyczne modele sieci i tworzą ich dwuwymiarowe obrazy graficzne. Sieci analizowane są pod względem swej struktury, powiązań, oddziaływań i możliwych regulacyjnych mechanizmów, zachodzących między poziomami hierarchicznej złożoności oraz podsystemami działającymi w komórce. Stale proponowane są i rozwijane nowe matematyczne modele sieci (Pan i Sinha, 2008). Matematyczne modelowanie sieci pozwala na badanie i przewidywanie, w jakim przedziale komórkowym przebiega określony proces molekularny oraz czy i jakie zorganizowane podsieci funkcjonalne są w tym układzie możliwe (Shin i in., 2009). Modele budowy sieci umożliwiają poznanie na poziomie molekularnym możliwych relacji pomiędzy występowaniem chorób wieloczynnikowych a kontekstem genetycznym bądź środowiskowym (Gohlke i in., 2009). Wraz z rozwojem biologii systemów pojawiają się także nowe koncepcje klasyfikacji chorób oraz szersze spojrzenie na powstawanie chorób o podłożu wieloczynnikowym, takich jak np. rak, co wyraża się również w nowej terminologii biomedycznej – „choroby sieci”.

### Struktura i rozwój sieci

Jak już wspomniano, informacje o interakcjach makromolekuł komórkowych są porządkowane i prezentowane w postaci sieci. Graficznie każdy rodzaj sieci (biologiczny, ekonomiczny, społeczny) jest przedstawiany jako zbiór „węzłów” – punktów ilustrujących jednostki składowe sieci oraz linii łączących te punkty („łączniki”). Sieć może mieć

łączniki nieukierunkowane (sieć symetryczna, graf niezorientowany) lub ukierunkowane (sieć asymetryczna, graf zorientowany). W życiu społecznym analogicznym przykładem są sieci ilustrujące interakcje i zależności międzyludzkie występujące np. w prostej czynności przywitania. Dwie osoby („węzły”, por. niżej) o takim samym statusie społecznym witają się przez podanie rąk („łącznik”, por. niżej) z równoważnym sygnałem emocjonalnym (zależność symetryczna), zaś przykładem drugim jest sytuacja, gdy szef korporacji wita się z pracownikami lub powszechnie znany celebryta podaje rękę, ale ledwie zauważa grupę zwykłych ludzi (zależność asymetryczna).

Podstawową jednostką obrazu sieci graficznej jest węzeł, oznaczany symbolem N (ang. *node*; w teorii grafów „wierzchołek”, ang. mat. *vertex*). W sieciach biologicznych każdy węzeł przedstawia jedną cząsteczkę, najczęściej jest to cząsteczka białka lub „gen” strukturalny (DNA). Nie spotyka się jeszcze modeli ilustrujących oddziaływanie różnych izoform białka.

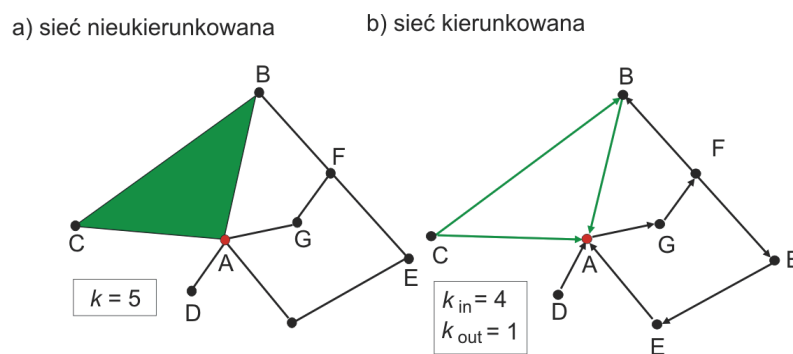
Dwa węzły powiązane łącznikiem o symbolu L, (od ang. *link*; w terminologii grafów „krawędź”, ang. *edge*) przedstawiają oddziaływanie między węzłami. W sieciach biologicznych jest to najczęściej oddziaływanie białko-białko lub białko-gen, gen-gen (w domyśle – produkty genów). Łączniki reprezentują oddziaływania funkcjonalne między makrocząsteczkami znajdujące w kompleksach żywej komórki, a wykazane doświadczalnie na podstawie danych funkcjonalnych, np. drogą dedukcji z badań genomicznych i proteomicznych, oznaczone techniką Y2H, spektroskopią powinowactwa i innymi metodami (por. wyżej). Interakcje międzycząsteczkowe oparte są zwykle na niekowalencyjnych, np. słabych wiązaniach wodorowych. Interakcje są labilne i dynamiczne. W sieciach metabolicznych węzeł przedstawia substrat lub metabolit, a łącznik – aktywny enzym prowadzący reakcję. Gdy struktura lub funkcja cząsteczki stanowiącej węzeł jest znana lub przynajmniej znany jest „obszar” jej funkcjonowania, węzły na diagramach przedstawiane są w postaci znaków (kółko, trójkąt, kwadrat, romb itp.), często oznaczonych kolorami. Kolor zwykle ilustruje funkcjonalny moduł lub podsieć. Węzły mogą też być „etykietowane”, czyli posiadać pełną nazwę identyfikującą lub akronim, np. symbol genu lub białka.

Teoria sieci jest złożona i nie będzie omawiana w tym artykule, gdyż autor nie czuje się kompetentny do jej komentowania. Podstawowe właściwości prostych sieci podaje za Barabási i Oltvai (2004) oraz Barabási (2009). Rycina 1 przedstawia prosty układ malej sieci zbudowany z węzłów (A, B, C, D, E, F).

Cechą charakteryzującą węzeł jest jego stopień (ang. *degree*), czyli liczba powiązań (ang. *connectivity*) z innymi węzłami. Stopień węzła oznacza się symbolem  $k$ . Na rycinie 1a węzeł A ma stopień  $k = 5$ , bo oddziałuje z węzłami C, B, G, E, i D. Oddziaływanie może być symetryczne, bez kierunku oddziaływania (jak na rycinie 1a) lub asymetryczne, czyli ukierunkowane, ilustrowane łącznikami ze strzałkami (jak na rycinie 1b), co



jednocześnie określa typ sieci. W pierwszym przypadku jest to sieć nieukierunkowana (lub symetryczna), w drugim – sieć ukierunkowana (asymetryczna). W ukierunkowanym typie sieci rozróżnia się stopień węzła dla oddziaływania wchodzącego  $k_{in}$  oraz oddziaływania wychodzącego  $k_{out}$ . Dla węzła A w sieci ukierunkowanej (ryc. 1b)  $k_{in} = 4$ , czyli do A wchodzi 4 sygnały od C, B, D i E. Jednocześnie węzeł A charakteryzuje  $k_{out} = 1$ , bo z węzła A wychodzi sygnał tylko do węzła G.



Ryc. 1. Podstawowe elementy i miary sieci: a) Sieć nieukierunkowana. Węzeł A posiada stopień  $k = 5$ , bo połączony jest bezpośrednio łącznikami z 5 innymi węzłami. Łączniki nie posiadają określonego kierunku; b) Sieć ukierunkowana. Łączniki posiadają strzałki wskazujące na kierunek oddziaływania. Węzeł A posiada 4 stopnie wejściowe ( $k_{in} = 4$ ) i 1 stopień wyjściowy ( $k_{out} = 1$ ). Szczegóły w tekście. Reprodukowano za zgodą Macmillan Publishers Ltd: Barabási A.L., Oltvai Z.N. (2004): *Network biology: understanding the cell's functional organization*. „Nature Rev. Genet.” 5, 101-113

Dla charakterystyki sieci podawane są również jej inne właściwości i miary (ang. *measures*):

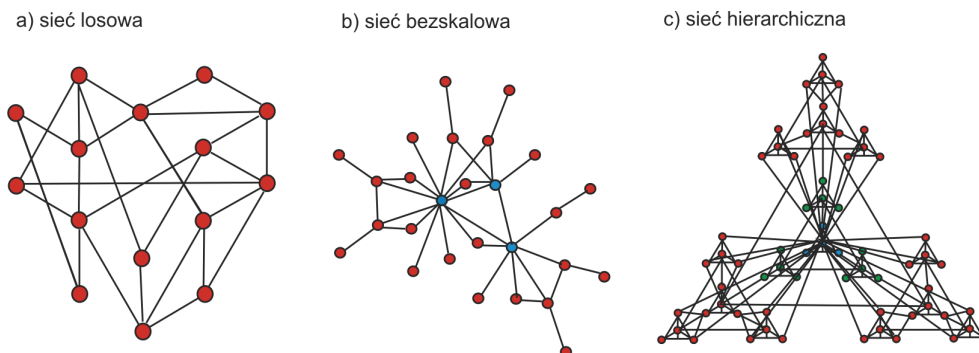
- średnia wartość  $k$ , oznaczana symbolem  $\langle k \rangle$  dla sieci nieukierunkowanej, mieści się w granicach 2-3 i wynika z wzoru  $\langle k \rangle = 2L/N$ ,
- stopień dystrybucji  $P(k)$  wskazuje na prawdopodobieństwo oszacowania, ile łączników określanych przez  $k$  posiada wybrany węzeł,
- większość sieci biologicznych jest sieciami nieskalowanymi (por. niżej),
- w sieciach ukierunkowanych odległość ścieżki między dwoma węzłami może być różna w zależności od kierunku łącznika; na przykładzie ryc. 1b widać, że odległość ( $l$ ) od B do A ( $l_{BA}$ ) jest krótka, posiada tylko jeden łącznik (wynosi:  $l_{BA} = 1$ ), gdyż na swej drodze nie ma żadnego węzła, natomiast odległość od A do B wynosi 3 ( $l_{AB} = 3$ ), ponieważ ma na swej drodze dodatkowe dwa węzły G i F (ścieżka z A do G, G do F i F do B),

- współczynnik grupowania (ang. *clustering coefficient*) pozwala na oszacowanie możliwości istnienia grup węzłów (ang. *node cluster*) i wykazanie hierarchicznych zależności w sieci.

Stożenie dystrybucji i współczynnik grupowania nie zależą od wielkości sieci, co daje możliwość uchwycenia podstawowych cech sieci i ich klasyfikacji, np. porównania sieci u organizmów na różnych etapach ewolucji, porównania sieci metabolicznej do sieci kontroli ekspresji genów itp. Złożoność podsieci szacuje się według skali. Węzeł izolowany, peryferyjny w sieci PPI odpowiada białku mającego „samodzielną” funkcję, a jego wartość na skali wynosi „0”. Gdy węzeł oddziałuje z innymi partnerami – jego stopień rośnie.

Oddziaływanie węzła z innymi partnerami nie jest losowe. Gdyby tak było, większość węzłów miałoby taką samą liczbę połączeń i ich rozkład byłby normalny (rozkład Poissona). W sieciach rzeczywistych dystrybucja połączeń nie ma rozkładu normalnego.

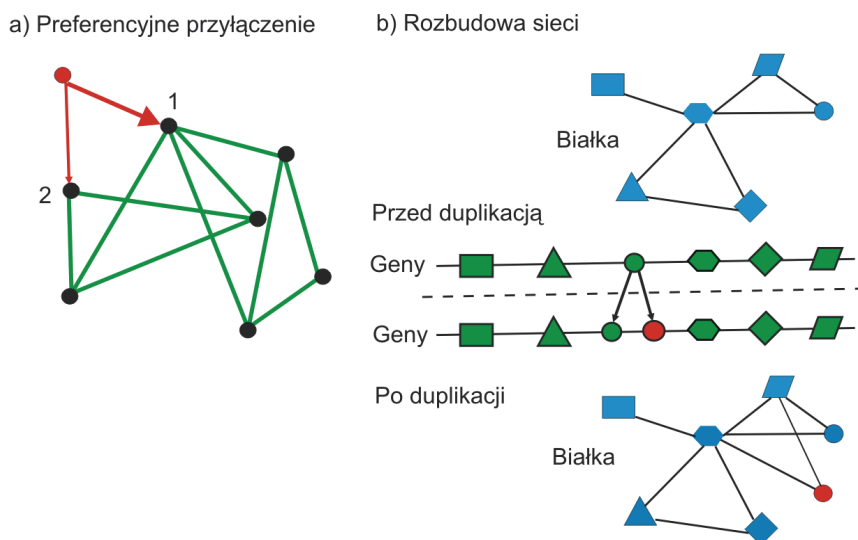
Na podstawie teorii grafów można wyliczyć „stopień dystrybucji”, czyli oszacować prawdopodobieństwo, z jakim dany węzeł będzie miał określoną liczbę połączeń (czyli ustalić stopień węzła,  $k$ ). Stopień dystrybucji podlega prawu potęgowemu, co oznacza, że rozkład stopni węzłów w sieci odbiega od normalnego rozkładu Poissona. W większości sieci eksponent potęgi wynosi między 2 i 3, nie ma zatem charakterystycznej wartości skali. Stąd pochodzi nazwa sieci: sieć bezskalowa (ang. *scale-free network* (ryc. 2).



Ryc. 2. Podstawowe modele sieci: a) Sieć losowa. Rozkład stopni węzłów jest losowy i podlega rozkładowi normalnemu (Poissona), co wskazuje, że większość węzłów ma mniej więcej tę samą liczbę połączeń. Węzły, które odstępają od tej reguły, występują niezmiernie rzadko; b) Sieć bezskalowa, zwana też siecią Barabási-Alberta (sieć BA). Rozkład stopni węzłów nie jest rozkładem normalnym i podlega prawu potęgowemu. Sieć zawiera niewielką liczbę węzłów typu piasta o wielu połączeniach (węzły o kolorze niebieskim). Jest to sieć uniwersalna występująca najczęściej w biologii. Szczegóły w tekście; c) Sieć hierarchiczna. Sieć posiada moduły i małe zgrupowania, które układają się w hierarchiczne struktury. Reprodukowano za zgodą Macmillan Publishers Ltd.: Barabási A.L., Oltvai Z. N. (2004): *Network biology: understanding the cell's functional organization*. „Nature Rev. Genet.” 5, 101-113

W sieciach naturalnych znajdują się też fragmenty sieci (subsieci) mające układ hierarchiczny. W sieciach rzeczywistych węzeł, który ma już dużą liczbę połączeń, wykazuje większe prawdopodobieństwo interakcji, czyli przyłączenia („doczepienia”) nowego węzła. Taki rozkład zachodzi dzięki działaniu w sieci dwóch mechanizmów: **wzrostu i preferencyjnego łączenia** (ang. *growth and preferential attachment*), (Barabási i Oltvai, 2004; Barabási, 2009). Odkrycie reguły wzrostu i preferencyjnego łączenia ma wielkie znaczenie dla biologii. Zdumiewającym było też odkrycie, że wiele sieci rzeczywistych ma podobną architekturę i wspólny układ podstawowy. Sieć bezskalowa jest zatem siecią uniwersalną, powszechnie występującą. Jest to niezwykle ważna reguła sieci, przenoszona nawet bezrefleksyjnie na systemy gospodarcze (bogatszy zawsze się bogaci kosztem biednego).

W biologii jest dużo przykładów obecności sieci bezskalowych, inaczej tzw. sieci „małego świata” (ang. *small-world network*), tworzących podsieci, cechujących się obecnością połączeń prawie wszystkich węzłów przez małą liczbę kroków (czyli sieci „nasyconych” węzłami typu piasty (por. niżej). Sieci bezskalowe stale się zmieniają, bo stale dochodzą nowe, nieprzypadkowe połączenia.

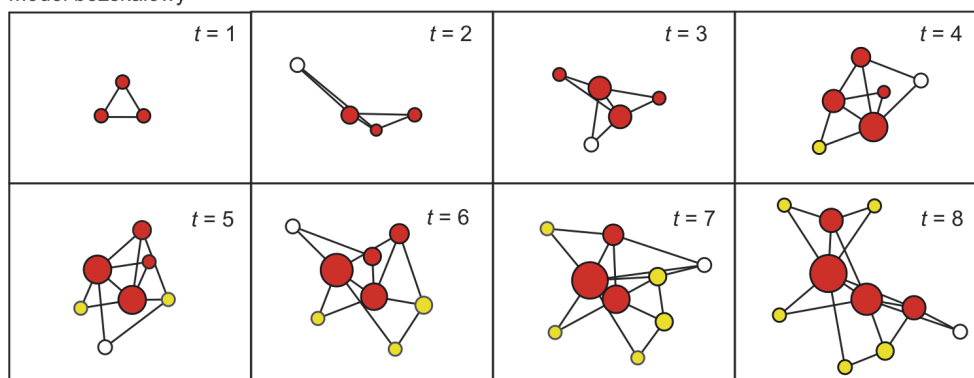


Ryc. 3. Powstawanie topologii bezskalowej w sieci, w trakcie ewolucji. Przykład działania prawa wzrostu i przyłączania: a) Preferencyjne przyłączenie: nowy węzeł (kolor czerwony) preferuje przyłączenie do węzła 1, gdyż ten posiada najwyższy stopień; b) Rozbudowa sieci w trakcie duplikacji genu. Sieć białek (kolor niebieski) przed duplikacją genu (kolor zielony). Po duplikacji powstał nowy gen (kolor czerwony), kodujący nowe białko (kolor czerwony), które dołącza się do istniejącej już sieci (niebieski) do węzła o najwyższym stopniu stosownie do reguły wzrostu i przyłączania. Reprodukowano za zgodą Macmillan Publishers Ltd: Barabási A.L., Oltvai Z.N. (2004): *Network biology: understanding the cell's functional organization*. „Nature Rev. Genet.”5, 101-113

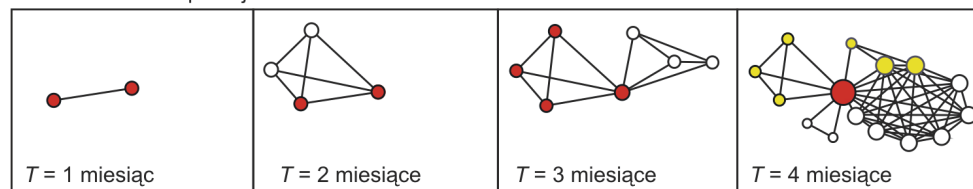
Proces taki zachodzi zapewne w ewolucji, gdy pojawiają się nowe geny w wyniku duplikacji (ryc. 3). Jeśli w układzie dodawany jest nowy węzeł (np. nowe białko kodowane przez gen zduplikowany), to ów nowy węzeł ma tendencję do interakcji z węzłem, który już ma dużo połączeń; w ten sposób w sieci wzrasta liczba węzłów typu „piasty” (por. niżej). Jest to ważny element zjawiska samoorganizacji sieci. Stąd też wynika wniosek, że ewolucja i struktura są procesami nierozłącznymi: w ewolucji stale przyrasta liczba węzłów i ich połączeń.

Inne, niebiologiczne sieci rozwijają się i nabywają nowe funkcje w czasie według podobnych reguł (ryc. 4).

Model bezskalowy



Naukowa sieć kooperacji



Ryc. 4. Powstawanie sieci. Część górna: model bezskalowy – rozwój w czasie. Począwszy od trzech połączonych węzłów początkowych, każdy nowy węzeł (oznaczony kółkiem pustym) przyłącza się do węzła o dużym stopniu (posiadających wiele połączeń) według reguły wzrostu i przyłączania. Rozmiary węzła są na tym schemacie proporcjonalne do jego stopnia. Część dolna: rozwój sieci kooperacji fizyków. Sieć ilustruje współautorstwo publikacji fizyków. Każdy węzeł oznacza pojedynczego autora, a łącznik oznacza współautorstwo. Ostatni schemat (dolny, prawy róg) przedstawia wyłanianie się węzła typu piasty i struktury typu „kliki” (każdy węzeł jest połączony z każdym węzłem). Według Barabási A.-L. (2009): *Scale-free networks: a decade and beyond*. „Science” 325, 412-413. Reprodukowano za zgodą AAAS

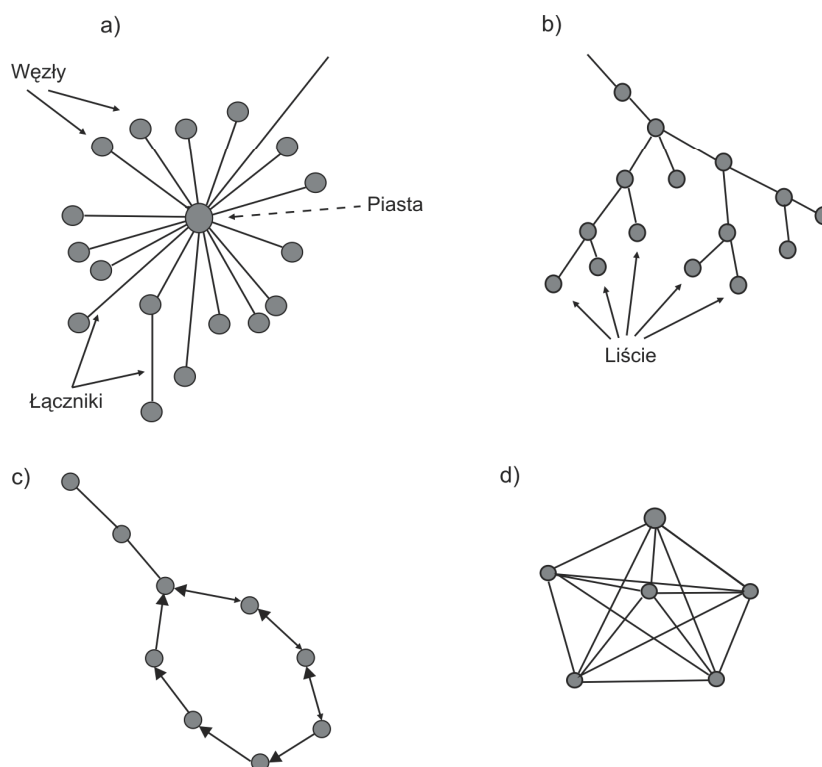
Obecnie rozwijane są teoretyczne metody bioinformatyczne pozwalające na przewidywanie, które białka mają wysoki potencjał wzajemnej interakcji w sieciach (Hsing i in.,

2008), oraz oszacowanie „współczynnika klikowości” (por. niżej). Reguła wzrostu i preferencyjnego łączenia ma swój wyraz w strukturze sieci. W trakcie rozwoju sieci powstają struktury zwane węzłami „centralnymi”, powiązаныmi licznymi łącznikami z wieloma innymi węzłami (por. np. ryc. 8 i 9). Węzeł centralny nazywany jest też „piastą” (ang. *hub*), gdyż ta struktura podobna jest do koła wozu/roweru z centralną piastą i szprychami. W sieci wykrywamy również połączenia między węzłami centralnymi (piastami). Mają one układ hierarchiczny: można wyróżnić węzły typu piasty o wielkim stopniu (takich jest niewiele) i węzły o małym stopniu (te są liczne). Taki układ nadaje sieci oporność. Prawdopodobieństwo, że czynnik uszkodzający zniszczy najpierw węzły o małym stopniu, jest duże, a zatem wzrasta szansa, że węzły o dużym stopniu utrzymają integralność sieci, a zatem życie komórki. Sieci bezskalowe są bardziej odporne na uraz niż sieci losowe. Ale gdy w sieci bezskalowej węzły o dużym stopniu zostaną wybiórczo usunięte, sieć taka rozpada się. Białka znajdujące się w piaście są przedmiotem zainteresowania w terapii lekowej jako tzw. białka docelowe (ang. *target proteins*). Nie zawsze działanie potencjalnych leków na tak „wysokim” poziomie jest jednak selektywne i bezpieczne dla pacjenta. Przykładem może być próba stosowania inhibitorów konwertaz prohormonów i probiałek. Wiedza o tkankowo-swoistej regulacji oraz obróbce probiałek do aktywnych molekuł jest jeszcze niepełna. Wiele z takich białek prekursorowych po ograniczonej, kontrolowanej proteolizie i konwersji przez rodzinę konwertaz probiałkowych (PC, ang. *pro-protein convertases*) staje się aktywnymi cząsteczkami (metaloproteinyazy, czynniki wzrostowe, cząsteczki adhezyjne i in.) biorącymi następnie udział w licznych funkcjach komórkowych związanych także z procesem nowotworowym (Khatib i in., 2002). Wydawało się, że konwertazy mogą być białkami „targetowymi”. Blokada aktywności konwertaz inhibitorami PC nie okazała się jednak skutecznym podejściem terapeutycznym. W obrazowaniu systemów czasami stosowane są graficzne sieci semiotyczne, które ilustrują zależności między „znakami” (ang. *sign*), odzwierciedlające określoną koncepcję lub złożoną funkcję, a nie pojedynczą cząsteczkę. Sieci semiotyczne są przydatne dla modelowania układów złożonych, np. ekosystemów, systemów społecznych.

### Rodzaje, funkcje, modele i oporność sieci

Wspomniano wyżej, że w sieci powstają węzły centralne lub typu piasty mające dużo połączeń z innymi węzłami (ryc. 5a). W sieciach biologicznych znane są też inne rodzaje podsieci. Niektóre fragmenty sieci (grafy) przypominają drzewa (pień → konary → gałęzie → liście), (ryc. 5b). Skupiska drzew tworzą las. Grafy w postaci drzew binarnych używane są między innymi w biologii dla ilustracji proliferacji i powstawania komórek potomnych z komórki wyjściowej, używane są w grafach rodowodów, w ekologii dla ilustracji łańcuchów pokarmowych. Te grafy są zwykle acykliczne, ukierunkowane.

W innych fragmentach występują cykle, to jest taki układ, gdy ukierunkowane łączniki tworzą pętlę prowadzącą do początkowej ścieżki. Mówiąc inaczej: ostatni wierzchołek pętli jest jednocześnie jej pierwszym wierzchołkiem (ryc. 5c). Mamy liczne przykłady takich sieci cyklowych w sieciach biologicznych (regulacje zwrotne). Takie substruktury grafu, gdzie wszystkie węzły łączą się ze sobą, nazywamy „kliką” (termin często używany również w sieciach społecznych) (ryc. 5d).



Ryc. 5. Niektóre rodzaje podsieci (grafy) występujące w sieciach biologicznych: a) piasta lub węzeł centralny, b) drzewo, c) cykl lub pętla, d) klika

Architektura sieci oparta jest na podstawowych nawracających wzorach zwanych motywami, mających formę trójkąta, kwadratu, pięciokąta, pętli itp. Opracowane są teoretyczne wzorce motywów, istniejące w sieciach biochemicznych (regulacja transkrypcji, sieci metaboliczne), neurobiologii, ekologii (łańcuchy pokarmowe) i technice. Sieci regulujące transkrypcję genów oparte są zwykle o trzy węzły o różnym kierunku połączeń. Taka triada, zwana modelem Heidera, zapewnia utrzymanie układu w stanie równowagi (Milo i in., 2002).

Pierwowzorem sieci molekularnych, jakimi posługuje się współczesna biologia systemów, były mapy metaboliczne opracowywane przez biochemików i najczęściej wyda-

wane przez firmy produkujące odczynniki laboratoryjne. Takie mapy metaboliczne, bardzo cenne dla celów dydaktycznych, ilustrowały przemiany cząsteczek, np. glukozy w cyklu kwasów trójkarboksylowych. W sieci metabolicznej węzły stanowią substraty lub pośrednie metabolity, a łącznikami są ukierunkowane reakcje katalizowane przez enzymy. Na przykładzie metabolizmu glukozy w architekturze graficznej tej sieci molekularnej koenzym A lub kwas pirogronowy można rozpatrywać jako węzły „centralne” (piasty), funkcjonalnie związane poprzez enzymy z innymi węzłami (metabolity). Innym przykładem cząsteczki białkowej stanowiącej węzeł „centralny” (piastę), mającej takie liczne powiązania, jest białko p53 (TP53).

Poznanie architektury i funkcji złożonych sieci w żywej komórce może poszerzyć nasze możliwości badawcze i zrewolucjonizować badania nad mechanizmami powstawania chorób uwarunkowanych wieloczynnikowo (np. nowotwory złośliwe). Jednak dane niezbędne dla budowania wielkich sieci genomicznych czy proteomicznych są jeszcze niekompletne. Takie spojrzenie ogólne na sieć (system) z pominięciem detali pozwoli nam na budowanie koncepcji działania całego systemu i postrzeganie bardziej holistyczne, bez wchodzenia w szczegóły, podobnie jak chemia organiczna rozwija się, pomimo że nie wchodzimy w subtelne prawa i detale fizyki kwantowej. Wielkie sieci obejmujące tysiące genów kodujących białko i genów regulatorowych będą wymagały wielkiej liczby równań dla opisu ich struktury i działania (Bernholdt, 2005). Każdy bowiem węzeł ilustrujący np. białko musiałby uwzględniać swoiste właściwości i zachowanie takiej cząsteczki – jej stabilność, syntezę i rozpad, jej konformację przestrzenną w różnych stanach funkcjonalnych komórki, w warunkach wahanía parametrów mikrośrodowiska, towarzyszących pojawieniu się w pobliżu innych cząsteczek, np. jonów itd. W dodatku cząsteczki mogą przyjmować różne konformacje przestrzenne i pełnić różne funkcje w zależności od lokalizacji, interakcji i „kontekstu”.

W sieciach regulacji genów istnieją obwody bistabilne typu „on/off”, a także moduły, takie jak np. autoregulowane pozytywne lub negatywne sprzężenia zwrotne, które mogą zmienić sygnał przejściowy w sygnał stały, a zatem służyć jako narzędzie przechowywania, czy też hamujące sprzężenia zwrotne, które tłumią niestabilność wskutek szumu oraz sprzężenia typu *feed-forward loops*, które mogą wzmacniać odpowiedź genową (wzrost ekspresji genów).

Analizując sieć metaboliczną, możemy zauważyć istnienie w niej wyspecjalizowanych podsystemów, modułów i motywów oraz dość dobrze zdefiniowanych kierunków działania. Sieci metaboliczne dają nam dość dobre uogólnione pojęcie o ich działaniu w systemie. Nie zawsze i nie wszystkie rodzaje sieci wskazują jednak zdefiniowane wejście (np. sygnału) w układ oraz wynik końcowy, czyli wyjście. Takie sieci nazywamy sieciami „izotropowymi” (Bray, 2003). Sieci, podsieci, moduły i motywów o różnej złożoności wykazują organizację hierarchiczną (Pan i Sinha, 2008; Milo i in., 2002).

Szlaki metaboliczne i sygnałowe są kształtowane przez sieć PPI, a ta z kolei jest regulowana przez sieć kontrolującą ekspresję genów (TRN, ang. *transcription regulatory network*). W obu typach sieci znaleziono mniej połączeń między centralnymi węzłami (piastami) w porównaniu do połączeń par piasta-węzeł peryferyjny o niskiej wartości. To wskazuje na małe prawdopodobieństwo komunikacji typu „cross-talk” między różnymi funkcjonalnymi modułami sieci, jednocześnie podnosząc oporność sieci na perturbacje (Maslov i Sneppen, 2002).

Jedną z podstawowych właściwości systemów biologicznych jest ich **oporność** na wewnętrzne i zewnętrzne perturbacje (Kitano, 2004). System musi działać w nieoczekiwanych warunkach środowiskowych. Bez mechanizmów oporności i pewnej plastyczności systemu nie byłaby możliwa ewolucja. Mechanizmy kontrolne, takie jak zduplikowanie funkcji, modularność sieci, rozprzęganie i eliminowanie procesów zmierzających do unicestwienia systemu, są rozpowszechnione w świecie organizmów żywych. Istnieje pewien rodzaj równowagi między opornością a wydajnością systemu i jego potrzebami (np. zapotrzebowaniem pokarmowym). System (sieć) może mieć miejsca bardziej stabilne (oporne), a inne bardziej „wątle”. Wyszukiwanie w sieci takich miejsc wątlących może być przydatne dla planowania nowych środków terapeutycznych ukierunkowanych na „białka docelowe” (tarczowe, „targetowe”), np. w leczeniu raka. Jednak trudność leży w tym, że choroby bardziej złożone najpewniej nie są spowodowane nieprawidłową funkcją jednego białka, lecz nieprawidłową interakcją różnych komponentów sieci (np. wielu białek, wielu miejsc bardziej „wątlących”).

Według Semple i in. sieci są na ogół bardziej odporne na wzrost ekspresji genów kodujących białka, gdy geny te aktywują się w procesach hamujących wzrost komórek. Sieć jest natomiast wrażliwa na spadek aktywności takich genów (Semple, Vavouri i Lehner, 2008).

Usunięcie krytycznej liczby węzłów sieci powoduje jej destrukcję i rozpad na grupy węzłów niekomunikujących się ze sobą. Prawdopodobnie działa tak agresywna inwazja wirusów lub obcych substancji toksycznych. Jednak oporność sieci na destrukcyjny atak jest zdumiewająca. Nawet usunięcie 80% przypadkowych węzłów może jeszcze utrzymać w sieci pozostałe 20% w postaci zwartego kompleksu (Barabási i Oltvai, 2004). Tłumaczyć to można tym, że przypadkowa, losowa destrukcja węzłów dotyczy zwykle wielu licznych, ale małych węzłów (węzłów z niewieloma połączeniami, węzłami o niskiej wartości skalowej), które nie przyczyniają się do rozpadu sieci, a jej „ogólna” integralność jest nadal utrzymana poprzez nieliczne białka o wielu połączeniach (piasty). Selektowny atak na wybrane białka centralne mające wiele połączeń typu białko-białko (czyli białko tworzące piastę) jest bardziej destrukcyjny. Tak na przykład u drożdży (*S. cerevisiae*) analiza delecyjna pokazuje, że te białka, które mają więcej połączeń między sobą są bardziej niezbędne dla życia. Tylko 10% białek niezbędnych to białka, które



mają 5 połączeń, ale aż 60% białek niezbędnych to takie, które mają 15 połączeń w sieci, czyli takie, które stanowią piastę (Barabási i Oltvai, 2004). Rozpad węzłów o typie piasty potwierdzono na innych gatunkach metodą symulacji komputerowej.

W sieci można wyróżnić połączenia, które charakteryzuje wysoka aktywność interakcyjna (*hot links*), zatopionych w sieci mniej aktywnych interakcji (Barabási i Oltvai, 2004). Rozwinięcie metod bioinformatycznych, które pozwolą na przewidywanie, gdzie znajdują się takie „gorące” połączenia między białkami typu piasta w całej sieci białek, może być pomocne w opracowywaniu potencjalnych leków (Hsing, Byler i Czerkasov, 2008). Dużym wyzwaniem współczesnej biologii jest zrozumienie budowy i działania sieci związanych z regulacją ekspresji genów, zwłaszcza sieci kontrolującej rozwój zarodka i różnicowanie komórek. Taką sieć genomyczną złożoną z 40 genów działających w czasie rozwoju jeżowca postulował Davidson (Davidson i in., 2002). Sieć ta posiada pozornie skomplikowaną strukturę, która miałaby kontrolować realizację genetycznego planu budowy ciała (*body plan*). Stan, aktywność i lokalizacja sieci w rozwijającej się zygocie zmienia się w czasie rozwoju. Sieć ta jest jednak bardzo uproszczona, statyczna, podobna do prostych instalacji elektrycznych działających na zasadzie „on/off”. Taki model dopuszcza zatem jedynie dwie wartości konfiguracji genu (tj. aktywny/nieaktywny). Sieć nie uwzględnia wielu swoistych właściwości w oddziaływaniach białko-białko, białko-DNA (genów) i ukierunkowania łączników. Aktywność genu nie przypomina wyłącznika elektrycznego, lecz podobna jest raczej do działania brzęczyka lub buczałki, oscyluje i zmienia się w małych przedziałach czasu. Oczywiście, regulacje binarne (*on/off*) działają również w obwodach regulujących funkcję genów.

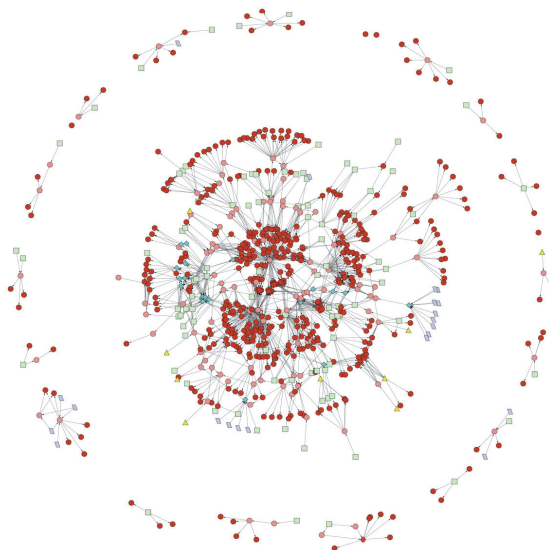
Modele sieci nie uwzględniają dynamiki, są raczej statyczne, bo same opierają się na statycznych danych, np. sieci regulatorowe genów używają informacji na podstawie mierzenia uśrednionych profili mRNA wziętych z tkanek i komórek znajdujących się w różnym stanie funkcjonalnym. Stąd wynika trudność w wyciąganiu wniosków co do przyczynowej relacji między genami (Andreucut, Huang i Kauffman, 2008).

Istotne znaczenie dla biologii i ewolucji będą miały informacje o podobieństwie węzłów centralnych u różnych, ewolucyjnie odległych gatunków.

Można przypuszczać, że poznanie sieci biorących udział w dojrzewaniu komórki jajowej, w przygotowaniu i rozmieszczeniu w oocycie RNA (mRNA, inne drobnocząsteczkowe RNA?), jak i w „dojrzałym” oocycie w przygotowaniu polarności komórki zygotycznej wymagać będzie nowych przemyśleń na temat tego czym jest i jak działa osobniczy „plan rozwojowy” od chwili poczęcia. Czy to nie system sieci modulowanej (jak?) w czasie dojrzewania komórki jajowej i następnie działającej w oocycie stoi również na straży realizacji planu rozwoju i odrębności gatunkowej? Czy swoisty dla każdego gatunku plan budowy, jako kardynalna, dziedziczona właściwość jest umocowany wyłącznie w numerycznym zapisie sekwencji DNA? Czy istnieją i działają tu także inne formy informacji?

### Przykłady sieci w różnych organizmach

Pierwsze analizy struktury sieci zostały opracowane dla drożdży (Gavin i in., 2002). Analiza objęła ponad 1700 genów drożdży (*S. Cerevisiae*), w tym ponad 1 100 ortologów genów ludzkich. Wśród oczyszczonych 589 białek autorzy znaleźli 232 wielocząsteczkowe kompleksy białkowe zawierające ponad 200 białek o nowych funkcjach. Poszerzone badania na drożdżach prowadziło równoległe wiele zespołów (por. Tong i in., 2004). Następnie opisano „interaktomy” bakterii, muszki owocowej nicienia (*C. elegans*), człowieka (np. Stelzl i in., 2005) i wielu innych gatunków. Najczęściej używanym modelem dla porównawczych badań interakcji białek są drożdże i muszka owocowa. Stwierdzenie homologii sekwencji dwóch białek ortologicznych u dwóch różnych gatunków niekoniecznie oznacza, że białka te w obu organizmach pełnią te same funkcje. Teoria sieci pozwala odpowiedzieć na pytanie, czy funkcje ortologicznych białek są istotnie identyczne. Są opracowane modele matematyczne dla porównania subsieci obu gatunków, co daje możliwość identyfikacji funkcji (Bandyopadhyay, Sharan i Ideker, 2006).

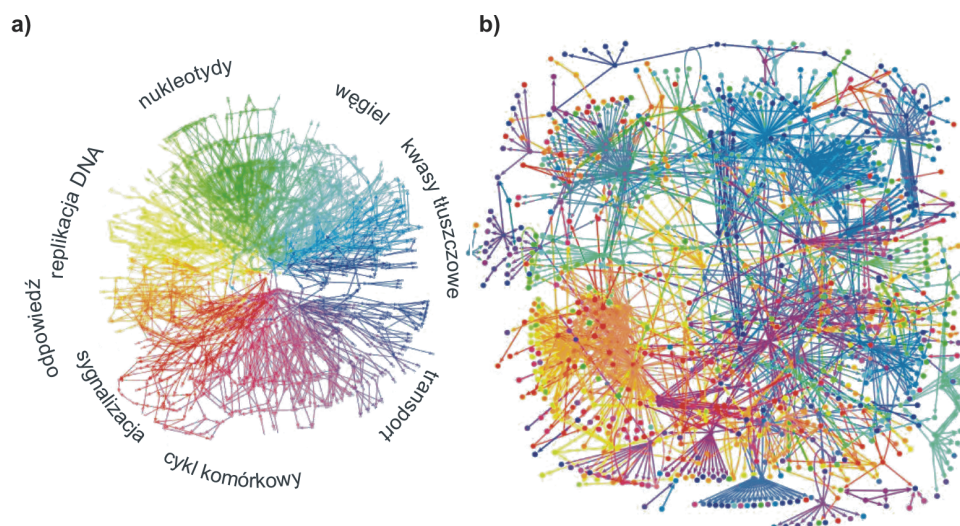


Ryc. 6. Fragment sieci regulacji transkrypcji kontrolujący metabolizm u *E. Coli*. Geny kodujące czynniki transkrypcyjne – punkty różowe, geny kodujące enzymy – punkty brązowe, metabolity zewnętrzne – kwadraty zielone, sygnały – żółte trójkąty. Według Samal A., Jain S. (2008): *The regularity network of E.coli metabolism as a Boolean dynamical system exhibits both homeostatic and flexibility of response*. „BMC Systems Biology” 2: 21

Samal i Jain (2008) skonstruowali model transkrypcyjnej regulatorowej sieci, kontrolującej metabolizm u *E. coli*. Ma ona postać hierarchicznej, modułarnej, acyklicznej sieci. System ten zachowuje zdolność homeostazy mimo zaburzeń w funkcji genów,

a jednocześnie jest plastyczny w swej odpowiedzi na zmiany środowiska (np. warunki odżywiania), zachowując swoją normalną architekturę i funkcję. Symulacja wpływu usuwania z genomu *E. coli* dowolnie wybranych 583 genów na wydajność metaboliczną (mierzoną zdolnością do przeżycia w 81 mediach minimalnych) wykazała znaczną trwałość (oporność) sieci. Sieć jest zatem układem elastycznym, ale zadziwiająco opornym, powracającym do stanu pierwotnego po chwilowych „odkształceniach”. Na tym polega jego siła i moc.

Kontrola ekspresji genów i kontrola przetwarzania informacji pierwotnej, cyfrowej na analogową informację wtórną (por. Chorąży, 2009) są odmienne i mają swe odniesienie do sieci. U bakterii *E. coli* architektura sieci działających w procesach regulacji transkrypcji (TRN) pozwoliła wyróżnić: dwa odmienne typy kontroli – numeryczny i analogowy oraz relacje tych form do konformacji przestrzennej (superskręty, forma zrelaksowana) DNA (Marr i in., 2008).



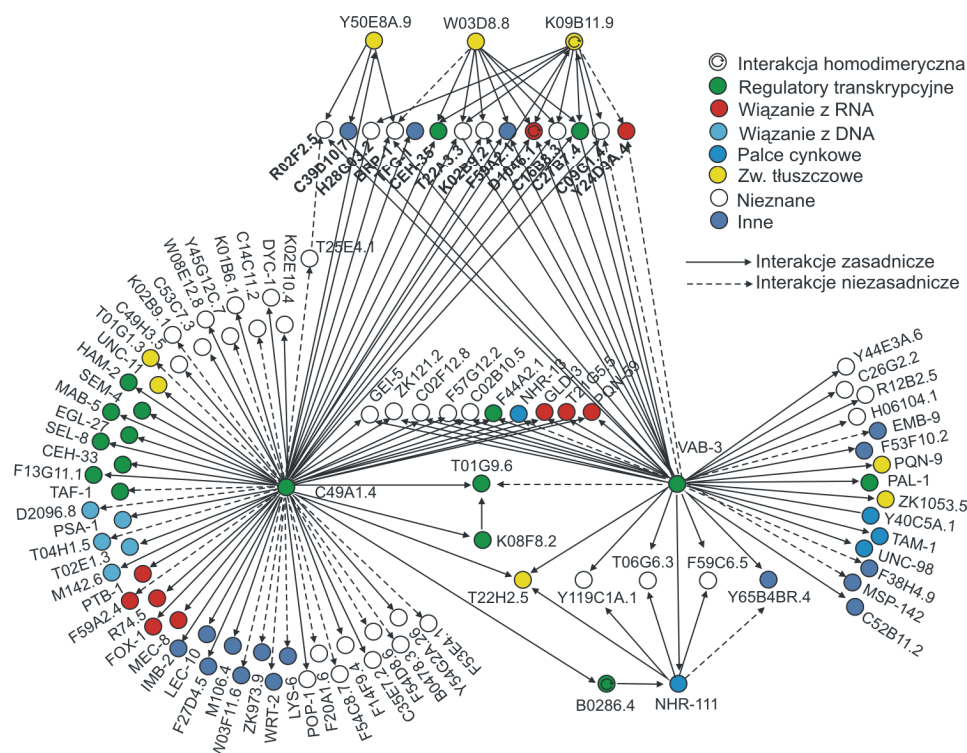
Ryc. 7. Sieć regulatorowa u drożdży. a) Sieć zbioru genów przypisanych tym samym funkcjom w różnych organizmach – geny ontologiczne (GO, ang. *gene ontology*). b) Sieć regulatorowa transkrypcji i enzymatycznych interakcji u drożdży *S. Cerevisiae*. Węzły sieci przedstawiają białka, oznaczone kolorami stosownie do sieci genów ontologicznych (GO) na ryc. 7a). Według Axelsen J.B., Bernhardsson S., Sneppen K. (2008): *One hub – one process: a tool based view on regulatory network topology*. „BMC Systems Biology” 2: 25

Jednym z kluczowych problemów biologii jest poznanie współzależności między systemem regulacji i funkcjonalną siecią molekularną. Sieć opracowana dla proteomu drożdży (*S. Cerevisiae*) wykazuje, że białka, których synteza była regulowana wspólnie, są związane z podobnymi zadaniami. Moduł przedstawiony jako grupa białek zgromadzonych wokół białka-piasty (węzła centralnego) najpewniej odpowiada jednemu zhar-

monizowanemu procesowi (Axelsen i in., 2008), (ryc. 7). Procesy zbliżone mają na rycinie podobny kolor.

Często wykorzystywanym modelem do badania sieci interakcji typu PPI jest nicien *Caenorhabditis elegans*. Ten organizm złożony z niespełna 1000 komórek jest doskonałym obiektem badawczym o dobrze poznanej organizacji anatomicznej i molekularnej. Sieć PPI dla *C. elegans* opracowała grupa M. Vidala. Analizie interakcji poddano 3024 białka. Wraz z wcześniejszymi opracowaniami analiza ta zaowocowała stworzeniem modelu wielkiej sieci WI5 (*worm interactom*, wersja 5) tego nicienia (Li i in., 2004). Pięć lat później ta sama grupa opracowała gigantyczną sieć dla tego samego organizmu o jeszcze wyższej złożoności, obejmującą około 116 000 dwuczłonowych interakcji białko-białko (WI8) (Simonis i in., 2009).

Jako przykład użyteczności sieci w celu śledzenia ewolucyjnych relacji między gatunkami może posłużyć przykładowa analiza dwóch białek nicienia, które oddziaływały z wieloma innymi białkami i wykazały wysoki współczynnik grupowania (ang. *clustering coefficient*) (ryc. 8).



Ryc. 8. Graficzna ilustracja bardzo licznych interakcji wokół węzła C49A1.4 i węzła VAB-3 u nicienia *C. elegans*. Oba te węzły są węzłami typu piasty; pierwszy jest węzłem o wysokim stopniu. Szczegóły w tekście. Wg Li S., Armstrong C. M., Bertin N. i in. (2004). *A map of the interactome network of the Metazoan C. elegans*. „Science” 303, 540-543. Reprodukowano za zgodą AAAS

Są to białka VAB-3 (ang. *Variable ABnormal morphology*), których ekspresja jest zlokalizowana w systemie nerwowym nicienia i białko C49A1.4, które ulega ekspresji w jego przewodzie pokarmowym i ma związek z rozwojem komórek hipodermalnych w części głowowej.

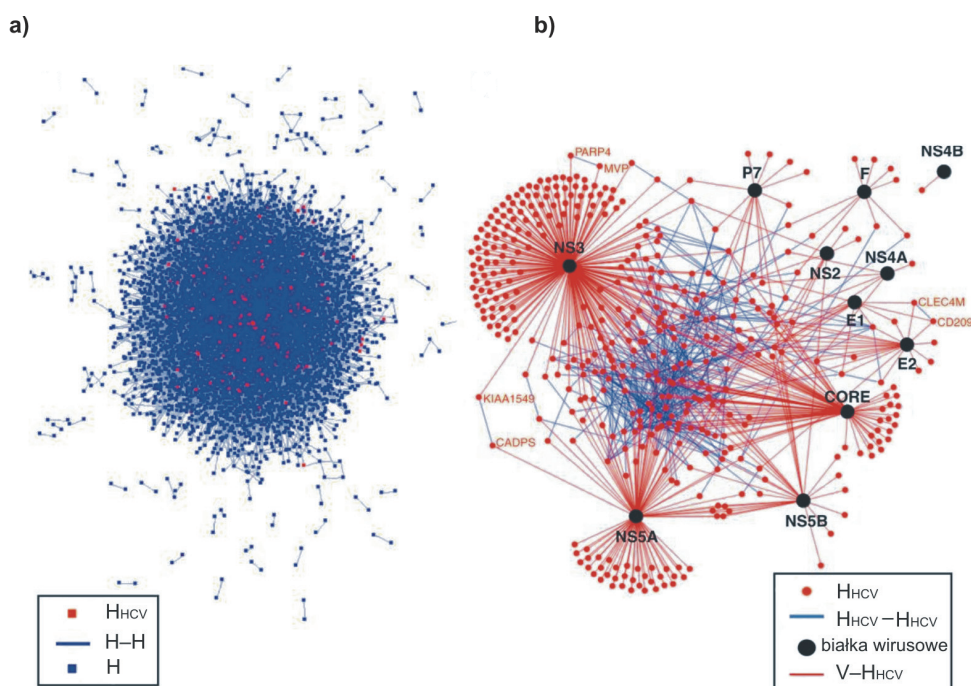
Oba te białka wykazują silne podobieństwo strukturalne do białek muszki owocowej (*Drosophila*). Białko VAB-3 nicienia to homolog białka EY, produktu genu *ey* (*eyeless*) muszki owocowej, a białko C49A1.4 to homolog EYA.A4, produktu genu *eya* (*eyes absent*) muszki. Białka EY i EYA.A4 są związane z morfogenezą oczu owada. Mutacje genów *ey* i *eya* muszki owocowej powodują głębokie patologie w rozwoju oczu, niedorozwój lub kompletne zaburzenia ich lokalizacji (np. powstawanie oka na odnóżach owada). Zarówno białko VAB-3, jak i C49A1.4 są elementami podsięci wykazującymi wysoki stopień powiązań z innymi białkami o różnych funkcjach u tego nicienia bądź z ortologami u innych organizmów (Li S. i in., 2004).

Porównawcze badania struktury i funkcji sieci, zwłaszcza stopień podobieństwa kompleksów białkowych sieci u przedstawicieli gatunków na różnych etapach ewolucji, dają nadzieje na głębsze poznanie molekularnych podstaw ewolucji na różnych jej etapach (Hirsh i Sharan, 2006; Yosef i in., 2009).

Problem funkcjonowania sieci PPI w stanach patologicznych zasługuje na osobne omówienie. Tu podam jeden przykład użyteczności podejścia oferowanego przez biologię systemów dla modelowania procesu infekcji ludzkim wirusem typu C zapalenia wątroby (HCV, ang. *hepatitis C-type virus*), wzięty z obszernej publikacji autorów francuskich (de Chassey i in., 2008). Wirusowe zapalenie wątroby stanowi poważny problem medyczny i społeczny; dotyczy wielu milionów ludzi, prowadzi do przewlekłych stanów chorobowych, głębokich zmian patologicznych w wątrobie prowadzących do steatozy, fibrogenyzy i marskości wątroby wiodącej do innych powikłań (np. zaburzenia krążenia), wreszcie jest czynnikiem współindukującym raka wątroby. Mapa ludzkiego interactomu skonstruowana przez wymienionych autorów objęła ponad 44 000 interakcji PPI między różnymi 9520 białkami, obejmującymi ok. 30% ludzkiego proteomu. Kodowane przez HCV białka strukturalne (CORE, E1, E2 i P7) oraz niestrukturalne (NS2, NS3, NS4A, NS4B, NS5A i NS5B) wykazują zdolność do interakcji z ok. 280 białkami ludzkimi, z tego ponad 90% to białka ulegające ekspresji w wątrobie. Dane te wraz z danymi z innych prac uzupełnionymi przez autorów dają w sumie interaktom HCV-człowiek złożony z 11 białek HCV i 421 białek człowieka (ryc. 9).

Spośród białek HCV, białka NS3, NS5A i CORE wykazują wysoki stopień, a zatem największą zdolność do interakcji odpowiednio z 214, 96 i 76 białkami komórkowymi człowieka. Z tej grupy 45 białek człowieka wchodzi w interakcję z więcej niż z jednym białkiem HCV, co sugeruje, że mają one podstawową rolę w procesie infekcyjnym (cyt. za de Chassey i in., 2008). Wśród swoistych perturbacji wywołanych chroniczną infekcją

sześciu wariantów genotypowych HCV przytoczę przykład zaburzeń przypuszczalnie indukowanych białkami NS3 i NS5A wirusa HCV, a związanych potencjalnie z indukcją raka wątroby. Oba te białka wykazują interakcję z kompleksem białek biorących udział w adhezji komórek do macierzy pozakomórkowej, a z kolei te związane z aktynami i cytoszkieletem pełnią funkcję w mobilności i migracji komórek. Rozregulowanie tych funkcji przez białka NS3 i NS5A może prowadzić do zaburzeń mobilności komórek odczepienia się ich od pozakomórkowej macierzy, a następnie do zapoczątkowania kancerogenezy i progresji nowotworu.



Ryc. 9. Graficzna ilustracja interakcji między białkami HCV i białkami człowieka. a) Sieć interakcji białek człowieka (H-H; kolor niebieski), punkty w kolorze czerwonym oznaczają cząsteczki białka po zakażeniu HCV. b) Sieć interakcji V-H. Węzły w kolorze czarnym – białka wirusowe; w kolorze czerwonym – białka ludzkie; łączniki czerwone – interakcja między białkami ludzkimi i wirusowymi; węzły niebieskie – interakcje między białkami ludzkimi. Największy komponent sieci (środek ryciny) obejmuje 196 cząsteczek białkowych. Wg de Chasse B., Navratil V., Tafforeau L. (2008): *Hepatitis C virus infection protein network*. „Molecular Systems Biology”, 4: 230

Oczekuje się, że poznanie architektury i ogólnych właściwości sieci komórkowych i interakcji białko-białko pozwoli na budowanie uogólnionych hipotez o molekularnych podstawach życia, rodzajach regulacji ekspresji genów, interakcji między podsieciami, stabilności i elastyczności sieci. Celem biologii systemów jest także nowa klasyfikacja chorób, diagnostyka i terapia. Możliwe, że uda się wskazać takie węzły i połączenia w sie-



ci, które są krytyczne dla procesu chorobowego, co umożliwi w przyszłości stworzenie nowej strategii planowania leczenia i opracowania nowych leków. Biologia systemów być może pozwoli wyjaśnić, dlaczego mutacja w genie nie musi dawać jednoznacznie zdeterminowanego efektu (fenotypu).

Sieci działają w środowisku wodnym, stanowiącym główny składnik komórki. Zjawiska życia zachodzą jedynie w określonej fazie płynnej wody. Gwałtowne przejście wody z fazy płynnej do stałej (zamrożenie) zabija komórkę, rozrywa nie tylko architekturę organelli i błon komórkowych (przez kryształki lodu), ale zapewne także rozrywa układ sieci i blokuje wszelkie dynamiczne procesy. Sterowane zamrażanie (np. przez monitorowanie tempa schładzania) powoduje powstanie z wody amorficznej fazy stałej, sprawia, że pomimo pozornego zamarcia przejawów życia komórki architektura struktur komórkowych i układ przestrzenny sieci pozostają nienaruszone, a powrót do życia jest możliwy. W ziarnach roślin naturalny kontrolowany proces odbierania wody („wysuszenie”) pozostawia nienaruszone struktury anatomiczne i cały układ sieci makromolekuł zarodka ziarna, które to ziarno nawet po długim okresie „stanu suchego” rozwija się, gdy zostanie odpowiednio nawodnione.

Kierunki przyszłych badań będą obejmowały głębsze poznanie architektury sieci, ich topologii w komórce, zachowania sieci w czasie rozwoju osobniczego, zmiany w odniesieniu do cyklu komórkowego i różnych funkcji komórek oraz stopnia ich zróżnicowania, a także porównanie sieci w różnych komórkach i narządach organizmów wielokomórkowych. Musimy jednak pamiętać, że sieci interakcyjne są w bezustannym ruchu. Makrocząsteczki wchodzą w interakcje, kompleksy powstają i rozpadają się, tworzą się nowe z innymi partnerami lub z innymi domenami, chemiczne cząsteczki takie jak grupy metylowe lub fosforowe modyfikują białka, modulując przez to ich interakcje itp. Stale biegnie synteza i rozpad makrocząsteczek, do gry wchodzi nowe komponenty. Złożoność tych procesów przekracza ludzką wyobraźnię i stanowi wielkie wyzwanie dla nauk o życiu. Do niedawna, np. szlaki sygnałne wydawały się nam proste i miały, jak sądziliśmy, liniowy charakter, rola białka p53 wydawała się oczywista i poznana dogłębnie. Nowe odkrycia i nowe podejścia rujną ten uproszczony, a często naiwny obraz. Szlaki sygnałne wchodzi w złożone interakcje z siecią, liniowy przekaz sygnału staje się procesem bardziej złożonym, odkrywa się nowe funkcje białka p53, np. jego udział w generowaniu mikroRNA, a sama cząsteczka p53 wchodzi w niezwykle złożone interakcje z innymi białkami i siecią.

Poznanie endogennych i zewnętrznych mechanizmów uszkodzenia sieci być może pozwoli na głębsze poznanie istoty wielu chorób i otworzy nowe możliwości diagnostyczne i terapeutyczne. Porównanie sieci u różnych gatunków, zwłaszcza sieci kontrolujących i regulujących genom, pozwoli na prześledzenie rozwoju i zachowania sieci w procesie ewolucji.

### Podziękowania

Bardzo serdecznie dziękuję pani dr Dorocie Butkiewicz i pani dr hab. Katarzynie Lisowskiej, panu prof. Jerzemu Silberringowi, a szczególnie panu prof. Aleksandrowi Kojowi za merytoryczne uwagi. Pani Beacie Bęben składam podziękowanie za przygotowanie tekstu i rycin do druku.

### Literatura

- Andrecut M., Huang S., Kauffman S.A. (2008): *Heuristic Approach to sparse approximation of gene regulatory networks*. „J. Computational Biol.” 15: 1173-1186.
- Axelsen J.B., Bernhardsson S., Sneppen K. (2008): *One hub – one process: a tool based view on regulatory network topology*. „BMC Systems Biology” 2: 25 (<http://www.biomedcentral.com/1752-0509/2/25>)
- Bandyopadhyay S., Sharan R., Ideker T. (2006): *Systematic identification of functional orthologs based on protein network comparison*. „Genome Res.” 16: 428-435.
- Barabási A.L. (2009): *Scale-free networks: a decade and beyond*. „Science” 325: 412-413.
- Barabási A.L., Oltvai Z.N. (2004): *Network biology: understanding the cell's functional organization*. „Nature Rev. Genet.” 5: 101-113.
- Bernholdt S. (2005): *Less is more in modelling large genetic networks*. „Science” 310: 449-451.
- Bork P., Jensen L.J., von Mering C. et al. (2004): *Protein interaction networks from yeast to human*. „Current Opinion in Struct. Biol.” 14: 292-299.
- Bray D. (2003): *Molecular networks: the top-down view*. „Science” 301: 1864-1865.
- de Chasseay B., Navratil V., Tafforeau L. i in. (2008): *Hepatitis C virus infection protein network*. „Molecular Systems Biology”, 4: 230.
- Chorąży M. (2009): *Gen strukturalny – ewolucja pojęcia i dylematy*. „Nauka” 3/2009: 57-108.
- Cramer F. (2001): *Gene technology in humans: can the responsibilities be borne by scientists, physicians, and patients?* „Interdisciplinary Science Review” 26: 1-4.
- Davidson E.H., Rast J.P., Oliveri P. et al. (2002): *A genomic regulatory network for development*. „Science” 295: 1669-1678.
- Fox Keller E. (2005): *The century beyond the gene*. „J. Biosc.” 30: 3-10.
- Gavin A.-C. i in. (2002): *Functional organization of the yeast proteome by systematic analysis of protein complexes*. „Nature” 415: 141-147.
- Gohlke J.M., Thomas R., Zhang Y. et al. (2009): *Genetic and environmental pathways to complex diseases*. „BMC Systems Biology” 3: 46.
- Gottlieb G. (200): *Enironmental and behavioural influences on gene activity*. „Current Directions in Psychol. Sci.” 9: 93-97.
- Hirsh E., Sharan R. (2006): *Identification of conserved protein complexes based on model of protein network evolution*. „Bioinformatics” 23, e170-e176.
- Hsing M., Byler K.G., Cherkasov A. (2008): *The use of gene-ontology for predicting highly-connected „hub” nodes in protein-protein interaction networks*. „MBC Systems Biology” 2: 80. (<http://www.biomedcentral.com/1752-0509/2/80>).
- Jura J., Węgrzyn P., Jura J., Koj A. (2006): *Regulatory mechanisms of gene expression: complexity with elements of deterministic chaos*. „Acta Bioch. Pol.” 53: 1-9.
- Khatib A.-M., Siegfried G., Chrétien M. et al. (2002): *Protein conversion in tumor progression and malignancy*. „Am. J. Path.” 160: 1921-1935.
- Kherlopian A.R., Song T., Duan Q. et al. (2008): *A review of imaging techniques for systems biology*. „BMC Systems Biology” 2: 74. (<http://www.biomedcentral.com/1752-0509/2/74>).



- Kitano H. (2004): *Biological robustness*. „Nature Rev. Gen.” 3: 826-827.
- Kitano H. (2002): *Systems biology: a brief overview*. „Science” 295: 1662-1664.
- Koj A.: *Zjawiska emergencji w biologii*. [w:] *Struktura i emergencja*, red. M. Heller, J. Mączka, Byblos, Kraków 2006, str. 153-160.
- Lehn J.-M. (2002): *Toward complex matter: Supramolecular chemistry and self-organization*. „PNAS” 99: 4763-4768.
- Lehn J.-M. (2004): *Supramolecular chemistry: from molecular information towards self-organization and complex matter*. „Rep. Prog. Phys.” 67: 249-265.
- Lehn J.-M. (2007): *From supramolecular chemistry towards constitutional dynamic chemistry and adaptive chemistry*. „Chem. Soc. Rev.” 36: 151-160.
- Lewontin R.: *The triple helix*. Harvard University Press, Cambridge, London, 2001, str. 1-136.
- Li S., Armstrong C.M., Bertin N. et al. (2004). *A map of the interactome network of the Meta-zoan C. elegans*. „Science” 303: 540-543.
- Marr C., Geertz M, Hütt M.-T., Muskhelishvili G. (2008): *Detecting the logical types of network control in gene expression profiles*. „BMC Systems Biology” 2: 18. (<http://www.biomedcentral.com/1752-0509/2/18>)
- Maslov S., Sneppen K. (2002): *Specificity and stability in topology of protein network*. „Science” 296: 910-913.
- Milo R., Shen-Orr S., Itzkovitz S. et al. (2002): *Network motives: simple building blocks of complex network*. „Science” 298: 824-827.
- Morowitz H.I.: *The emergence of everything*, Oxford University Press, Oxford New York 2002.
- Nepal D., Geckeler K.E. (2006): *Functionalization of carbon nanotubes*, w „Functional Nanomaterials”, red. K.E. Geckeler, E. Rosenberg, wyd. American Scientific Publishers, Stevenson Ranch, Cal., 2006, str. 57-79.
- Pan R.K., Sinha S. (2008): *Modular network with hierarchical organization: the dynamical implications and complex structure*. „Prahana- J. Phys.” 71: 331-340.
- Parkkinen J.A., Kaski S. (2010): *Searching for functional gene modules with interaction component models*. „MBC Systems Biol.” 4: 4.
- Samal. A., Jain S. (2008): *The regulatory network of E.coli metabolism as a Boolean dynamical system exhibits both homeostatic and flexibility of response*. „BMC Systems Biology” 2: 21. (<http://www.biomedcentral.com/1752-0509/2/21>)
- Semple J.I., Vavouri T., Lehner B. (2008): *A simple principle concerning the robustness of protein complex activity to change in gene expression*. „MBC Systems Biology” 2:1.
- Shin C. J., Wong S., Davis M.J., Ragan M.A. (2009): *Protein-protein interaction as a predictor of subcellular location*. „BMC Systems Biology” 3: 28.
- Simonis N., Rual J.-F., Carvunis A.-R. i in. (2009): *Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network*. „Nature Methods” 6: 47-54.
- Stelzl U., Worm U., Lalowski M. et al. (2005): *A human protein-protein interaction network: a resource for annotating the proteome*. „Cell” 122: 957-968.
- Strand R. (2000): *Naivety in the molecular life sciences*. „Futures” 32: 451-470.
- Tong A.H.Y., Drees B., Nardelli G. et al. (2002): *A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules*. „Science” 295: 321-324.
- Tong A.H.Y., Lesage G., Bader G.D. et al. (2004): *Global mapping of the yeast genetic interaction network*. „Science” 303: 808-813.

- Trewaras A.: *A brief history of systems biology*. „The Plant Cell” 18: 2420-2430.
- Westerhoff H.V., Palsson B.O. (2004): *The evolution of molecular biology into systems biology*. „Nature Biotech.” 22: 1249-1252.
- Yosef N., Kupiec M., Ruppin E., Sharan R. (2009): *A complex-centric view of protein network evolution*. „Nucleic Acids Research” 37, e88, doi:10.1093/nar/gkp414.

### **An introduction to systems biology**

A gigantic amounts of data and information on molecules that constitute the very complex cell machinery have been collected, classified and stored in data banks. Although we possess enormous amount of knowledge about the properties and functions of thousands of molecular entities, we are still far from understanding how they do work in a living cell. It is clear now that these molecules (genes, proteins) are not autonomous, that there is no direct linear relation between genotype and phenotype, and that the majority of functions are carried and executed by concerted molecular activity, and that the majority of diseases are multifactorial. A basic property of the matter in a living cell (both normal and pathologic) is an interaction between variety of macromolecules, mainly proteins, genes (DNA) etc. In a process of self-organization they are able to form an active molecular biologic system – a complex, labile and dynamic network which integrity is secured by non-covalent bounds. In this essay some basic properties of network structure and the universal rules that govern them are described. Network or system biology is promising new research approach in biology and medicine.

**Key words:** systems biology, reductionism, structure of networks, graph