

Wprowadzenie do biologii systemów. Część I

Mieczysław Chorąży

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.

NOWOTWORY Journal of Oncology 2012; 62, 5: 363–370

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

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 jej funkcji. „Mimo tego wielkiego sukcesu (biologii molekularnej, MCh-) staje się jednak coraz bardziej jasne, 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 subsieci 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 postęпами 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 geno-

mu człowieka miało nam wyjaśnić istotę człowieczeństwa. Tymczasem okazało się, że genom szympansa różni się od genomu człowieka tylko o 1,5%, mimo bardzo odmiennych różnic fenotypowych. Oznacza to, że geny „same z siebie” nie determinują fenotypu, i że poza genami musi istnieć jeszcze inna „informacja”, „program” lub „system” (Fox Keller, 2005; Lewontin, 2001). W genocentrycznym 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 (kompartymenty). 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, linijność, 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).

Z drugiej strony 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 dwudziestolecu biologia molekularna, a zwłaszcza „genomika funkcjonalna”, przyniosła wielki ładunek precyzyjnych informacji o strukturze i funkcjach makrocząsteczek. Porządkowanie, bankowanie i obróbka tych informacji stały się możliwe 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 poprzez 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, *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 i 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 maszyneryi transkrypcyjnej i translacyjnej, biorą udział w replikacji i naprawie DNA oraz w przekazywaniu sygnałów. Są receptorami i 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 uzyskanych z doświadczeń uzyskanych na drodze analizy redukcjonistycznej, a jednocześnie akcentuje mocno powstawanie złożoności (*complexity*) na coraz wyższych hierarchicznych poziomach organizmów żywych. Termin **system** używamy dla określenia grupy dwóch lub większej ilości części (komponentów) współdziałających ze sobą, zebranych w **sieć** w taki sposób, że tworzą funkcjonalną całość (Trevares, 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 opisanego 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, korzystaniu z informacji genetycznej oraz reakcji na sygnały środowiskowe (zewnętrzne i wewnętrzne). Gdy tylko organizacja biologiczna (system) 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 utrzy-

mane w ładzie. Funkcjonowanie żywej komórki nie daje się wyjaśniać prostymi liniowymi zależnościami między cząsteczkami. Biologia systemów dostarczyła wiele dowodów na to, 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 utorowała 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 zająć 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 od-

działywaniach białko-białko (PPI, *protein-protein interaction*), 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 dla uzyskania informacji o interakcjach makrocząsteczek, takie jak: mikromacierze peptydowe i DNA, spektrometria mas, technika dwuhybrydowa z użyciem drożdży (Y2H, *yeast two hybrids*), dedukcja oddziaływań na podstawie analizy badań genomicznych 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”, *quantum dots* — QD; „nano-łuski”, *nano-shells* (Nepal i Geckeler, 2006). Dla śledzenia losów makrocząsteczek w żywej komórce budowane są nowe typy mikroskopów optoelektronicznych, rozwijają się metody microPET (*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 bazach 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 wyod-

rębnią 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 i 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 domena SH3 (*Src homologic domain 3*), domena EH (*Eps homologic*) wiążąca się z rodziną małych cząsteczek GTPaz, domena WW zawierająca sekwencję ok. 40 aminokwasów, zwijająca się w trójpasmową 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, 2002), 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

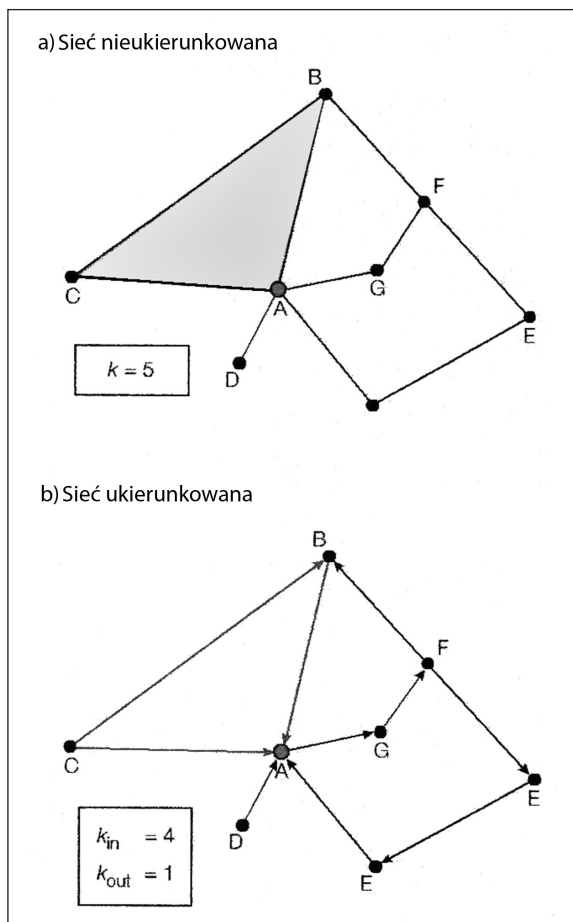
umożliwiają poznanie na poziomie molekularnym możliwych relacji pomiędzy występowaniem chorób wieloczynnikowych a kontekstem genetycznym bądź środowiskowym (Gohlke i in., 2009). Wraz z rozwojem biologii systemów pojawiają się także nowe koncepcje klasyfikacji chorób oraz szersze spojrzenie na powstawanie 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 (*node*; w teorii grafów „wierzchołek”, *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 L (od *link*; w terminologii grafów „krawędź”, *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 wiązaniach 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

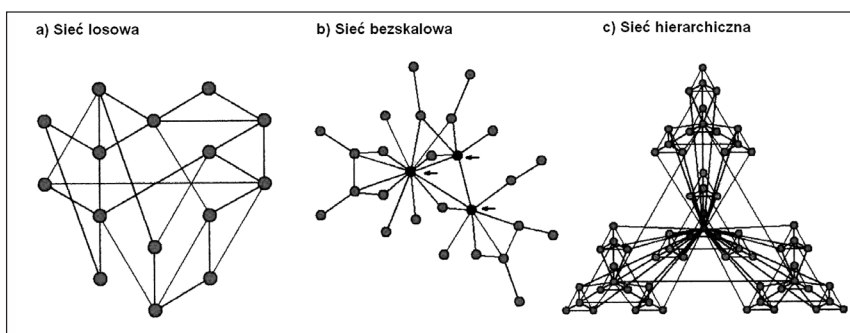


Rycina 1. Podstawowe elementy i miary sieci. a) Sieć nieukierunkowana. Węzeł A posiada stopień $k = 5$, bo połączony jest bezpośrednio łącznikami z 5 innymi węzłami. Łączniki nie posiadają określonego kierunku. b) Sieć ukierunkowana. Łączniki posiadają strzałki wskazujące na kierunek oddziaływania. Węzeł 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 AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nature Rev Genet* 2004; 5: 101–113

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 podaję za Barabási i Oltvai (2004) oraz Barabási (2009). Rycina 1 przedstawia prosty układ małej sieci zbudowany z węzłów (A, B, C, D, E, F).

Cechą charakteryzującą węzeł jest jego stopień (*degree*), czyli liczba powiązań (*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 pierw-



Rycina 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 odstępują 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 wskazane strzałką). 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 AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nature Rev Genet* 2004; 5: 101–113

szym 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.

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

- średnia wartość k , oznaczana symbolem $\langle k \rangle$ dla sieci nieukierunkowanej, mieści się w granicach 2–3 i wynika z wzoru $\langle k \rangle = 2L/N$;
- dystrybucja stopnia $P(k)$ 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 ryciny 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 (*clustering coefficient*) pozwala na oszacowanie możliwości istnienia grup węzłów (*node cluster*) i wykazanie hierarchicznych zależności w sieci.

Dystrybucja stopnia 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 izolowa-

ny, peryferyjny w sieci PPI odpowiada białku mającemu „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 miałaby 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 „dystrybucję stopnia”, czyli oszacować prawdopodobieństwo, z jakim dany węzeł będzie miał określoną liczbę połączeń (czyli ustalić stopień węzła, k). Dystrybucja stopnia podlega prawu potęgowemu, co oznacza, że rozkład stopni węzłów w sieci odbiega od normalnego rozkładu Poissona. W większości sieci eksponent potęgi wynosi od 2 do 3, nie ma zatem charakterystycznej wartości skali. Stąd pochodzi nazwa sieci: sieć bezskalowa (*scale-free network*; ryc. 2). 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** (*growth and preferential attachment*) (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 bezrefleksyjnie nawet na systemy gospodarcze (bogatszy zawsze się bogaci kosztem biednego). W biologii jest dużo przykładów obecności sieci bezskalowych, inaczej tzw. sieci „małego świata” (*small-world network*), tworzących podsieci, cechujących się obecnością połączeń prawie wszystkich węzłów poprzez małą liczbę

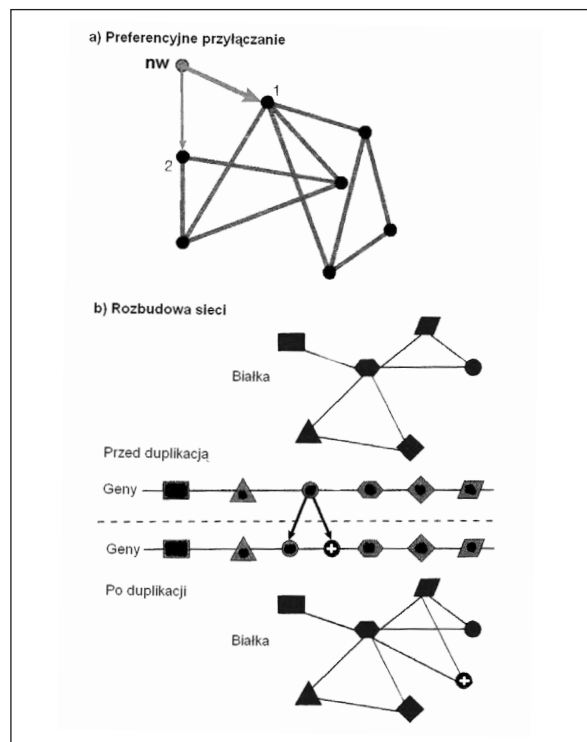
kroków (czyli sieci „nasyconych” węzłami typu piasty; por. niżej). Sieci bezskalowe stale się zmieniają, bo stale dochodzą nowe, nieprzypadkowe połączenia. Proces taki zachodzi zapewne w ewolucji, gdy pojawiają się nowe geny w wyniku duplikacji (ryc. 3).

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 ma już 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.

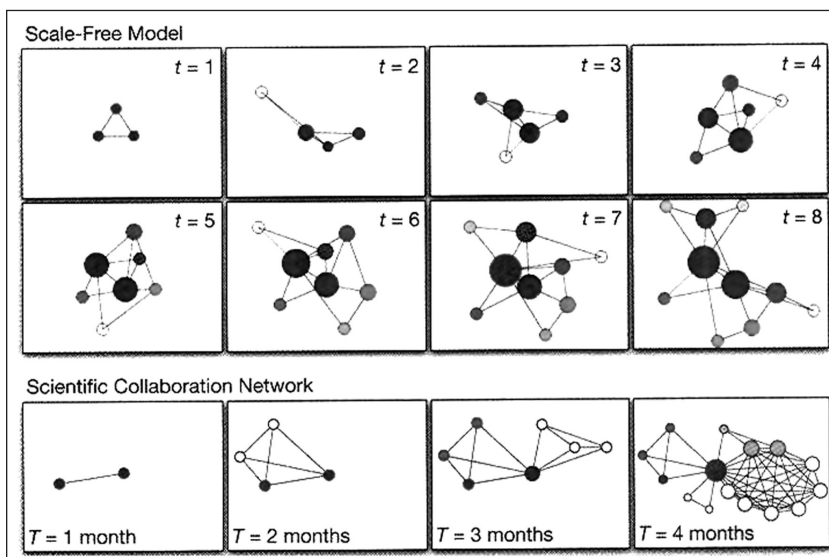
Inne, niebiologiczne sieci, rozwijają się i nabywają nowe funkcje w czasie według podobnych reguł (ryc. 4).

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”, powiązany licznymi łącznikami z wieloma innymi węzłami (por. np. ryc. 8 i 9). Węzeł centralny nazywany jest też „piastą” (*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. Prawdopodob-



Rycina 3. Powstawanie topologii bezskalowej w sieci, w trakcie ewolucji. Przykład działania prawa wzrostu i przyłączania. a) Preferencyjne przyłączenie: nowy węzeł (nw) preferuje przyłączenie do węzła 1, gdyż ten posiada najwyższy stopień. b) Rozbudowa sieci w trakcie duplikacji genu. Sieć białek przed duplikacją genu. Po duplikacji powstał nowy gen (oznaczony symbolem „+”) kodujący nowe białko (oznaczone symbolem „+”), które dołącza się do istniejącej już sieci, do węzła o najwyższym stopniu — stosownie do reguły wzrostu i przyłączania. Reprodukowano za zgodą Macmillan Publishers Ltd: Barabási AL, Oltvai ZN. Network biology: understanding the cell’s functional organization. *Nature Rev Genet* 2004; 5: 101–113



Rycina 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 (posiadającego 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. Scale-free networks: a decade and beyond. *Science* 2009; 325: 412–413. Reprodukowano za zgodą AAAS

bieństwo, że czynnik uszkodzający zniszczy najpierw węzły o małym stopniu, jest duże, a zatem wzrasta szansa, że węzły o dużym stopniu utrzymają integralność sieci, czyli 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 (*target proteins*). Nie zawsze jednak działanie potencjalnych leków na tak „wysokim” poziomie jest selektywne i bezpieczne dla pacjenta. Przykładem może być próba stosowania inhibitorów konwertaz pro-hormonów i pro-białek. Wiedza o tkanowo-swoistej regulacji oraz obróbce pro-białek do aktywnych molekuł jest jeszcze niepełna. Wiele z takich białek prekursorowych po ograniczonej, kontrolowanej proteolizie i konwersji przez rodzinę konwertaz pro-białkowych (PC, *pro-protein convertases*) staje się aktywnymi cząsteczkami (metaloproteiny, 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” (*sign*), odzwierciedlające określoną koncepcję lub złożoną funkcję, a nie pojedynczą cząsteczkę. Sieci semiotyczne są przydatne dla modelowania układów złożonych, np. ekosystemów i systemów społecznych.

Prof. dr hab. n. med. Mieczysław Chorąży

członek rzeczywisty PAN

Zakład Biologii Nowotworów

(Obecnie: Centrum Badań Translacyjnych

i Biologii Molekularnej Nowotworów)

Centrum Onkologii — Instytut im. Marii Skłodowskiej-Curie

Oddział w Gliwicach

e-mail: chorazy@io.gliwice.pl

Piśmiennictwo

Barabási AL. Scale-free networks: a decade and beyond. *Science* 2009; 325: 412–413.

Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nature Rev Genet* 2004; 5: 101–113.

Bork P, Jensen LJ, von Mering C i wsp. Protein interaction networks from yeast to human. *Current Opinion in Struct Biol* 2004; 14: 292–299.

Bray D. Molecular networks: the top-down view. *Science* 2003; 301: 1864–1865.

de Chassey B, Navratil V, Tafforeau L i wsp. Hepatitis C virus infection protein network. *Molecular Systems Biology* 2008; 4: 230.

Chorąży M. Gen strukturalny — ewolucja pojęcia i dylematy. *Nauka* 2009: 57–108.

Cramer F. Gene technology in humans: can the responsibilities be borne by scientists, physicians, and patients? *Interdisciplinary Science Review* 2001; 26: 1–4.

Fox Keller E. The century beyond the gene. *JBiosc* 2005; 30: 3–10.

Gohlke JM, Thomas R, Zhang Y i wsp. Genetic and environmental pathways to complex diseases. *BMC Systems Biology* 2009; 3: 46.

Gottlieb G. Environmental and behavioural influences on gene activity. *Current Directions in Psychol Sci* 2000; 9: 93–97.

Hsing M, Byler KG, Cherkasov A. The use of gene-ontology for predicting highly-connected hub nodes in protein-protein interaction networks. *MBC Systems Biology* 2008; 2: 80. <http://www.biomedcentral.com/1752-0509/2/80>.

Jura J, Węgrzyn P, Jura J i wsp. Regulatory mechanisms of gene expression: complexity with elements of deterministic chaos. *Acta Bioch Pol* 2006; 53: 1–9.

Khatib A-M, Siegfried G, Chrétien M i wsp. Protein conversion in tumor progression and malignancy. *Am J Path* 2002; 160: 1921–1935.

Kherlopian AR, Song T, Duan Q i wsp. A review of imaging techniques for systems biology. *BMC Systems Biology* 2008; 2: 74. <http://www.biomedcentral.com/1752-0509/2/74>

Kitano H. Systems biology: a brief overview. *Science* 2002; 295: 1662–1664.

Koj A. Zjawiska emergencji w biologii. W: *Struktura i emergencja*, Heller M, Mączka J (red.). Kraków: Byblos; 2006, 153–160.

Lehn J-M. Toward complex matter: Supramolecular chemistry and self-organization. *PNAS* 2002; 99: 4763–4768.

Lehn J-M. Supramolecular chemistry: from molecular information towards self-organization and complex matter. *Rep Prog Phys* 2004; 67: 249–265.

Lehn J-M. From supramolecular chemistry towards constitutional dynamic chemistry and adaptive chemistry. *Chem Soc Rev* 2007; 36: 151–160.

Lewontin R. *The triple helix*. London: Harvard University Press, Cambridge; 2001, 1–136.

Morowitz HI. *The emergence of everything*. London: Oxford University Press; 2002.

Pan RK, Sinha S. Modular network with hierarchical organization: the dynamical implications and complex structure. *Prahana J Phys* 2008; 71: 331–340.

Parkkinen JA, Kaski S. Searching for functional gene modules with interaction component models. *MBC Systems Biol* 2010; 4: 4.

Shin CJ, Wong S, Davis MJ i wsp. Protein-protein interaction as a predictor of subcellular location. *BMC Systems Biology* 2009; 3: 28.

Strand R. Naivety in the molecular life sciences. *Futures* 2000; 32: 451–470.

Tong AHY, Drees B, Nardelli G i wsp. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* 2002; 295: 321–324.

Trevaras A. A brief history of systems biology. *The Plant Cell* 2006 18: 2420–2430.

Westerhoff HV, Palsson BO. The evolution of molecular biology into systems biology. *Nature Biotech* 2004; 22: 1249–1252.

Przedruk artykułu „Wprowadzenie do biologii systemów” z kwartalnika *Nauka* 2011; nr 1: 59–84 za zgodą Autora i Redakcji *Nauki*.