

PAWEŁ STACEWICZ*, RADOSŁAW SIEDLIŃSKI**

O INFORMATYCZNEJ KATEGORII ANALOGOWOŚCI I JEJ STOSOWALNOŚCI W BIOLOGII MOLEKULARNEJ***

Abstract

ON THE COMPUTER-SCIENCE CATEGORY OF ANALOGICITY
AND THE POSSIBILITY OF ITS APPLICATION IN MOLECULAR BIOLOGY

The main aim of this paper is to justify the thesis that in molecular biology – in the scope of phenomena fundamental for the functioning of the cell – a significant role is played by analog (non-discrete) information, which can be described in computational terms. It is a methodological thesis, indicating a certain direction of advancing new biological hypotheses. This aim is realized in two stages. In sections 1 and 2 we discuss the computer-science concept of analogicity, generally describing different concepts of analog-continuous and analog-empirical computations, as well as discussing the relationship between analogicity and digitality. In sections 3 and 4 we analyze some components of the process of protein formation, emphasizing that an adequate description of this process requires taking into account information of an analog nature, which, with a certain research attitude, can be described, but also used, computationally.

Keywords: analog computation, analogicity, computer science, empirical computation, molecular biology, philosophy of computing, protein folding

* Wydział Administracji i Nauk Społecznych, Politechnika Warszawska, Pl. Politechniki 1, 00-661 Warszawa, e-mail: p.stacewicz@ans.pw.edu.pl, ORCID: <https://orcid.org/0000-0003-2500-4086>.

** Polsko-Japońska Akademia Technik Komputerowych, ul. Koszykowa 86, 02-008 Warszawa, e-mail: rad_siedl@wp.pl, <https://orcid.org/0000-0002-7680-983X>.

*** Dziękujemy dwóm anonimowym recenzentom pracy, których cenne wskazówki pozwoliły usunąć wiele błędów z pierwotnej wersji artykułu. Dziękujemy także organizatorom i uczestnikom dwóch konferencji naukowych, VI Konferencji Filozofii Matematyki i Informatyki (Poznań 2018) oraz V Konferencji Filozofia w Informatyce (Lublin 2019), w trakcie których mieliśmy okazję omawiać wstępne fragmenty pracy. Za wszelkie pozostałe w artykule błędy, niejasności i niedociągnięcia ponosimy odpowiedzialność my sami i z góry przepraszamy za nie wszystkich Czytających.

We współczesnej praktyce badawczej informatyka i biologia coraz częściej się przenikają, wymieniając między sobą pojęcia, metody i wyniki. Dobitnym tego wyrazem są dobrze określone obszary badawcze, które sytuują się na pograniczu obydwu dyscyplin. Należą do nich bioinformatyka, a w niej informatyczne modele układów ożywionych (Luscombe, Greenbaun, Gerstein 2001), oraz rozmaite koncepcje obliczeń naturalnych inspirowanych bądź realizowanych biologicznie (Rozenberg, Back, Kok 2012)¹.

W artykule ograniczamy dziedzinę rozważań, po pierwsze, do dziedziny danych i obliczeń analogowych (przeciwstawianych niekiedy cyfrowym), po drugie, do procesów wewnątrzkomórkowych, przede wszystkim tych, które odpowiadają za powstawanie i funkcjonowanie białek. Koncentrując się na pojęciu analogowości, argumentujemy, że relacja między biologią a informatyką jest dwustronna, przypomina sprzężenie zwrotne, a obydwie dyscypliny mogą skorzystać na tym, że odwołują się do wyników drugiej strony. Ta pożyteczna dwustronność jest szczególnie widoczna wówczas, gdy pracuje się w paradygmacie obliczeniowym, to znaczy próbuje się rozpoznać w przyrodzie takie czy inne schematy przetwarzania odpowiednio kodowanych informacji. W artykule skupiamy się na informacji analogowej. Przy takim podejściu staje się jasne, że z jednej strony informatyka wnosi wiele do biologii, dostarczając jej chociażby odpowiednich narzędzi formalnych, takich jak matematyczne modele obliczeń analogowych. Z drugiej strony, biologia wspiera informatykę, wskazując i precyzyjnie opisując zjawiska (np. subkomórkowe), które przy pewnej interpretacji można wykorzystać do rozwiązywania problemów *stricto* obliczeniowych (np. optymalizacyjnych).

Głównym celem artykułu jest pokazanie na wybranych przykładach, że w biologii molekularnej – w zakresie zjawisk podstawowych dla funkcjonowania komórki – znaczącą rolę odgrywa informacja analogowa. Tę zaś warto ujmować w kategoriach informatycznych, wypróbując różne modele danych i obliczeń analogowych. Cel ten rozumiemy jako zgodny z zadaniami szeroko pojętej filozofii nauki, która winna wskazywać poszczególnym naukom obiecujące kierunki stawiania nowych hipotez, pełniąc tym samym ważną funkcję heurystyczną.

Cel pracy realizujemy dwustopniowo. W dwóch pierwszych rozdziałach omawiamy informatyczne pojęcie analogowości, opisując różne koncepcje obliczeń analogowych ciągłych oraz analogowych empirycznych, a także omawiając relację analogowości do cyfrowości (szerzej: dyskretności). W dwóch

¹ Osobnym obszarem są próby obliczeniowego ugruntowania jednego z fundamentów nowoczesnej biologii – teorii ewolucji w drodze doboru naturalnego (Kozłowski 2011). Należy do nich silnie osadzona w teorii obliczeń koncepcja metabiologii Gregory'ego Chaitina. W sprawie jej zwężłej prezentacji i krytyki por. Siedliński 2016.

kolejnych rozdziałach analizujemy proces powstawania i zwijania białek, podkreślając, że adekwatny opis tego procesu wymaga uwzględnienia informacji o charakterze analogowym, która przy pewnym badawczym nastawieniu może być opisywana, ale także wykorzystywana, obliczeniowo.

Przykłady, na które się powołujemy, zaczerpnięte zostały z biologii molekularnej, gdzie pojęcie informacji jest zazwyczaj utożsamiane z informacją genetyczną. Ten stan rzeczy ma powody natury historycznej, które pomijamy, koncentrując się na kwestiach metodologicznych. Zainteresowanego Czytelnika pozostaje nam odesłać do literatury przedstawiającej szczegóły procesu, który doprowadził do utożsamienia pojęcia informacji istotnej z biologicznego punktu widzenia z dyskretną oraz opartą na kodzie informacją, której nośnikiem jest DNA (por. np. Cobb 2017). To oczywiście nie zmienia faktu, że proces życia na wszystkich szczeblach organizacji – od subkomórkowego po ekologiczny – pełen jest istotnych czynników o charakterze analogowym. Naszym zamiarem jest wskazanie istotnych źródeł informacji analogowej komplementarnych wobec informacji genetycznej w obszarze tradycyjnie opisywanym głównie w kategoriach genocentrycznych.

1. ANALOGOWOŚĆ W PERSPEKTYWIE INFORMATYCZNEJ

1.1. W dobie wszechobecnych technologii cyfrowych pojęcie analogowości budzi skojarzenia raczej historyczne. Używa się go w sformułowaniach typu „sygnał analogowy”, „telefonía analogowa” czy „komputer analogowy”, słusznie przywołujących na myśl rozwiązania, które odchodzą w technologiczną przeszłość. Świadomi tych faktów przyjmujemy dalej nastawienie raczej teoretyczne, neutralne względem bieżących zastosowań. Koncentrujemy się na informatycznym, a dokładniej obliczeniowym, aspekcie analogowości, rozumiejąc analogowość jako pewną cechę obliczeń lub danych. Przyjmujemy zatem, że istota informatycznych układów analogowych – niezależnie od ich konkretnego przeznaczenia – zawiera się w tym, że mogą otrzymywać na wejściu dane analogowe, a następnie przetwarzać je w sposób analogowy (co jednak, jak pokazujemy dalej, bywa różnie rozumiane).

Najbardziej precyzyjne znaczenia analogowości uzyskuje się wówczas, gdy cechę tę definiuje się w ramach konkretnego modelu obliczeń opisującego matematycznie elementarne operacje, które można wykonywać na danych. Jednym z nich jest model rzeczywistych funkcji rekurencyjnych, w którym za elementarny matematyczny odpowiednik danej analogowej uznaje się liczbę rzeczywistą (Costa, Graça 2003). Ponieważ jednak, co wyjaśniamy dalej, po-

wołanie się na ścisły model matematyczny nie zawsze jest możliwe, niekiedy trzeba poprzestać na wskazaniu pewnego sposobu obliczania. Zakłada się przy tym, że sposób ten jest na tyle charakterystyczny, że pozwala wyodrębnić pewną szczególną klasę technik przetwarzania danych.

Niezależnie od tego, którą strategię się przyjmuje — twardą (odniesienie do modelu) czy miękką (odniesienie do sposobu) — w metodologii informatyki trzeba rozróżnić dwie rodziny znaczeń terminu „obliczenia analogowe” (Stacewicz 2017b, 2020). Z jednej strony, za analogowe uznaje się obliczenia ciągle pozwalające przetwarzać dane ciągle, które na najniższym poziomie matematycznego opisu są charakteryzowane za pomocą liczb rzeczywistych z pewnego *continuum* (np. z przedziału $[0,1]$) (Moore 1996). Z drugiej strony, mianem analogowych określa się obliczenia empiryczne (inaczej: naturalne) polegające na realizacji specjalnie dobranych procesów fizycznych będących naturalnymi analogonami² pewnych operacji matematycznych. Do kategorii pierwszej, nazwijmy ją AN-C, należą tradycyjne obliczenia analityczne służące np. do różniczkowania, całkowania czy rozwiązywania równań różniczkowych. Były one rozpowszechnione przed erą technologii cyfrowych, a implementowano je fizycznie m.in. za pomocą układów elektronicznych opartych na wzmacniaczach operacyjnych (Kulka, Nadachowski 1979, 1982). Do kategorii drugiej, nazwijmy ją AN-E, trzeba zaliczyć zarówno pewne metody historyczne służące np. do rozwiązywania równań liniowych za pomocą wyspecjalizowanych układów mechanicznych (Ifrah 2006: 654-655), jak i metody współczesne polegające m.in. na realizacji obliczeń za pomocą pewnych procesów biochemicznych (por. np. Kari, Rozenberg 2008). To właśnie do tych ostatnich odwołujemy się w drugiej części pracy.

1.2. Obliczenia analogowe typu AN-C, czyli ciągle, mają wiele dobrze określonych modeli teoretycznych, które w taki lub inny sposób precyzują podstawowe dla nich pojęcie ciągłości (danych i obliczeń). Przy tym można zaryzykować stwierdzenie, że najbardziej elementarnym wzorcem tego pojęcia jest własność ciągłości liczb rzeczywistych, czyli — mówiąc poglądowo — własność polegająca na tym, że dwie liczby rzeczywiste mogą różnić się dowolnie mało, wobec czego każdy przedział liczb rzeczywistych stanowi gęste *continuum*³.

² Nazwa „analogowe” jest w tym przypadku usprawiedliwiona właśnie dlatego, że obliczanie odwołuje się do (uzasadnionej empirycznie) analogii między strukturami i operacjami matematycznymi a obiektami i procesami fizycznymi.

³ Na pojęciu liczby rzeczywistej są matematycznie „nadbudowane” kolejne pojęcia istotne z punktu widzenia teorii obliczeń, takie jak ciągła funkcja zmiennej rzeczywistej czy funkcja różniczkowalna.

Ze względu na sposób realizacji obliczeń w czasie przedstawione modele dzieli się na modele obliczeń ciągłych wykonywanych w czasie ciągłym oraz modele obliczeń ciągłych wykonywanych w krokach dyskretnych. Do pierwszych należy niewątpliwie prekursorski model GPAC (Shannon 1941) oraz jego rozmaite rozszerzenia, takie jak np. EAC (Rubel 1993). Konstrukcje tego typu nie wymuszają żadnego rodzaju dyskretyzacji (ani sygnałów, ani trybu ich przesyłania czy przetwarzania), choć oczywiście przewidują możliwość przetwarzania sygnałów dyskretnych jako szczególnych wielkości z dozwolonego przez model *continuum*. W ich przypadku dopuszczalne na poziomie implementacji operacje przetwarzania danych opisuje się za pomocą dobrze określonego minimum funkcji ciągłych oraz operacji na tych funkcjach, takich jak dodawanie, mnożenie przez stałą, dodawanie stałej do funkcji, różniczkowanie czy całkowanie (Mycka 2010). Ponadto każdy z modeli ma aspekt strukturalny, to znaczy pozwala określić, np. w formie grafu zorientowanego, złożoną strukturę operacji wiodących od danych do wyniku końcowego. Określenie takiej struktury, czyli np. zaprojektowanie grafu oraz przypisanie elementarnych operacji funkcyjnych do jego węzłów, można potraktować jako stworzenie algorytmu realizacji pewnego zadania. Co najważniejsze jednak, w przypadku opisywanych modeli — mimo dyskretnych schematów przetwarzania danych — zakłada się, że poszczególne operacje następują po sobie w sposób ciągły (mówiąc obrazowo: przechodzą jedna w drugą).

Modele drugiego rodzaju, do których należą np. rekurencyjne sieci neuronowe (Siegelmann 1998) lub model BSS (Blum, Shub, Smale 1989), postulują dyskretyzację czasu, a zatem w ich przypadku kolejne operacje na danych ciągłych są od siebie wyraźnie oddzielone. Wynik jednej, uzyskany w pewnym możliwym do wyróżnienia etapie, staje się daną wejściową kolejnej. Przykładowo, model BSS przewiduje dyskretną listę rozkazów, z których każdy stanowi wywołanie oddzielnej operacji realizowanej w określonym czasie po zakończeniu operacji poprzedniej. Własność ciągłości dotyczy tutaj pojedynczych operacji, które — po pierwsze — mogą być wykonywane na danych reprezentowanych przez liczby rzeczywiste (zakłada się przy tym, że dane takie są całościowo zapisywane w specjalnych rejestrach), a po drugie, mogą polegać na zastosowaniu do danej ciągłej pewnej funkcji wymiernej. Krótko mówiąc, obliczenia są tutaj ciągłe lokalnie w ramach pojedynczych operacji na wielkościach ciągłych (por. Blum, Shub, Smale 1989).

Należy zaznaczyć, że celem tego artykułu nie jest dokładny opis matematyczny przedstawionych modeli, lecz podkreślenie faktu, że konstrukcje takie istnieją i mają solidne podstawy matematyczne. Z tego nie wynika, rzecz jasna, że modele takie są dostatecznie bliskie rzeczywistości, to znaczy opisują typy danych lub operacji, którym odpowiadają wielkości albo procesy rzeczywiście

ciągłe występujące w rzeczywistości fizycznej, np. w świecie ożywionym. Niemniej, jeśli z badań empirycznych, w tym biologicznych, miałyby wynikać, że zjawiska o charakterze ciągłym rzeczywiście występują w przyrodzie (albo słabiej: że opisując je, musimy stosować matematykę ciągłą), to modele te byłyby dla nauk przyrodniczych (w tym biologii) bardzo wartościowe. Co więcej, byłyby one istotniejsze poznawczo niż modele klasyczne, czyli dyskretne (które są, rzecz jasna, bardzo użyteczne w technice).

1.3. Drugi z typów obliczeń analogowych — obliczenia AN-E, czyli empiryczne — zawdzięcza swoje istnienie teoriom nauk empirycznych (takim jak fizyka czy biologia), w których dla pewnych procesów fizycznych istnieje opis formalny przy użyciu struktur i operacji matematycznych. Jeśli na teorii takie spojrzymy odwrotnie, niż to czyni się zazwyczaj, tzn. nie pod kątem sformalizowanego opisu rzeczywistości, lecz pod kątem fizycznej realizacji postulowanych przez nie operacji matematyczno-algorytmicznych, to uzyskamy interesującą nas koncepcję obliczeń.

Rozważane w ramach tej koncepcji formy obliczania polegają na realizacji specjalnie dobranych, istniejących w przyrodzie, niejako gotowych procesów fizycznych (Kari, Rozenberg 2008). Zakłada się przy tym, że efektywność tych procesów przewyższa efektywność tradycyjnych rozwiązań algorytmicznych. A zatem, aby zróżniczkować pewną funkcję, można odnaleźć w przyrodzie — kierując się określoną teorią sformalizowaną — efektywny proces różniczkujący i go zastosować. Aby scałkować funkcję, wystarczy wykorzystać pewien naturalny układ całkujący. Aby rozwiązać trudny problem optymalizacyjny, można zdać się na istniejący w przyrodzie, być może ewolucyjnie wytworzony, układ optymalizujący. Co ważne jednak, przydatność takiego czy innego układu wynika z adekwatności teorii opisującej ten układ (jako element szerszej klasy zjawisk pewnego typu) w kategoriach matematycznych. Podsumowując i uogólniając: obliczenia typu AN-E polegają na realizacji pewnych odpowiednio dobranych procesów fizycznych, które na gruncie uzasadnionej empirycznie teorii odpowiadają określonym operacjom matematycznym.

Ponieważ obliczenia empiryczne są odpowiednio dobierane, trudno mówić o istnieniu pewnego uniwersalnego ich modelu, który na podobieństwo teorii obliczeń typu AN-C (ciągłych) określałby elementarne (matematyczne) operacje i struktury, pozwalające zdefiniować dowolne obliczenia typu AN-E. W ich przypadku można powołać się jedynie na modele lokalne, będące właściwie fragmentami teorii uzasadniających przypisanie pewnemu procesowi fizycznemu określonej (prostej lub złożonej) operacji matematycznej (por. Stacewicz 2020). Na przykład, modelem obliczeń wykonywanych przez pewien mechaniczny układ różniczkujący byłby ten fragment mechaniki teoretycznej, który

opisuje ruch takich a takich ciał z wykorzystaniem rachunku różniczkowego i całkowego.

Wyjaśnijmy jeszcze jedną kwestię. Chociaż, historycznie rzecz biorąc, obliczenia typu AN-E odwoływały się do fizyki klasycznej i były realizowane za pomocą układów (mechanicznych lub elektronicznych) możliwych do opisanego z pomocą fizyki klasycznej (Ifrah 2006), to współczesne obliczenia empiryczne (choć rzadko już nazywane analogowymi) wykraczają poza wskazaną dziedzinę. Wykraczają w dwóch kierunkach. Z jednej strony, czerpią swoje uzasadnienie z fizyki kwantowej, która „podpowiada” między innymi, w jaki sposób można wykorzystać tzw. paralelizm kwantowy (Deutsch 1985). Z drugiej strony, do czego nawiążemy w dalszej części pracy, odwołują się do biologii (w naszym ujęciu jest to biologia molekularna i biochemia).

2. ANALOGOWOŚĆ A CYFROWOŚĆ

2.1. Obliczenia analogowe typu AN-C, których wyróżnikiem jest ciągłość przetwarzanych danych, definiuje się w opozycji do obliczeń cyfrowych. Te drugie, we współczesnej praktyce dominujące, są określone m.in. za pomocą modelu uniwersalnej maszyny Turinga⁴, który na poziomie najbardziej podstawowym i abstrahującym od konkretnej reprezentacji danych i programów wyjaśnia ogólnie, na czym polega przetwarzanie odpowiednio zakodowanych informacji przez programowalne urządzenie cyfrowe. Model ten pozwala uchwycić istotę cyfrowości. Polega ona na tym, że zarówno same dane, jak i schematy ich przetwarzania (programy) mają postać kodu symbolicznego złożonego z rozróżnialnych, dyskretnych elementów. Następujące składniki rozważanego modelu odpowiadają dyskretności: podział taśmy abstrakcyjnej maszyny na odrębne klatki, skończona (choć dowolnie duża) liczba odrębnych stanów maszyny, skokowe ruchy głowicy, symboliczny alfabet zapisu danych i rozkazów. Na jeszcze głębszym poziomie liczbowej reprezentacji danych i programów cyfrowość wyraża się w konieczności reprezentowania i jednych, i drugich za pomocą liczb obliczalnych w sensie Turinga⁵. Zbiór

⁴ W tekście głównym użyto sformułowania „między innymi”, ponieważ istnieje wiele innych modeli obliczeń cyfrowych równoważnych ze względu na klasę rozwiązywanych problemów modelowi Turinga — chociażby opis obliczeń w kategoriach rachunku lambda Alonso Churcha.

⁵ W praktyce jest to skończony podzbiór liczb naturalnych (reprezentowanych np. przez skończoną długość ciągu binarne), ale w teorii — przewidującej nieograniczone, coraz większy przyrost zasobów realnych maszyn — jest to zbiór liczb obliczalnych (por. Turing 1936).

tych liczb jest ziarnisty i ma strukturę podobną do zbioru liczb wymiernych⁶, co oznacza, że jego elementy (odpowiadające danym) są dyskretne.

Obliczenia typu AN-C, niezależnie od tego, w jakim modelu się je definiuje, łamią zasadę dyskretności. Dopuszczają przetwarzanie sygnałów ciągłych opisywanych matematycznie za pomocą liczb rzeczywistych (najczęściej z pewnego przedziału), umożliwiają także ciągłe ich przetwarzanie, np. w taki sposób, jak przewiduje to model GPAC (por. 1.2). Co ciekawe, jeśli odniesiemy się ponownie do liczbowej reprezentacji danych, to obliczenia typu AN-C pozwalają operować także na liczbach nieobliczalnych w sensie Turinga. Wynika to z matematycznego faktu, że ich zbiór stanowi takie dopełnienie zbioru liczb obliczalnych, które zapewnia zbiorowi liczb rzeczywistych własność ciągłości (Stacewicz 2019). Przypomnijmy przy tym, że z definicji dziedziną obliczeń ciągłych są sygnały (a w ujęciu matematycznym: liczby) ciągłe.

Fakty te przesądają o tym, że ze względów teoretycznych (a nie np. historycznych czy technologicznych) obliczeń typu AN-C nie można traktować jako przeciwstawnych względem obliczeń cyfrowych. Są one raczej ich rozszerzeniem — podobnie jak zbiór liczb rzeczywistych stanowi rozszerzenie zbioru liczb naturalnych, a przedział $(0,1)$ stanowi nieskończone liczne rozszerzenie zbioru binarnego $\{0,1\}$. Mówiąc wprost, obliczeniowe rozszerzenie polega na tym, że każda z odmian obliczeń ciągłych pozwala operować na istotnie szerszym zbiorze danych niż taki a taki zbiór dyskretny.

Fakt „bycia rozszerzeniem”, chociażby na poziomie zbioru akceptowalnych danych i możliwych do wygenerowania wyników, skłania do przypuszczenia, że niektóre przynajmniej z obliczeń AN-C mają status hiperobliczeń. Znaczą to, że pozwalają one rozwiązywać problemy nierozwiązywalne w ramach klasycznego modelu Turinga, w tym problem stopu czy równań diofantycznych (Harel 2000, Shagrir 2004). I tak właśnie jest: dla niektórych modeli obliczeń ciągłych, np. modelu rzeczywistych funkcji rekurencyjnych, istnieją matematyczne dowody możliwości rozwiązania wskazanych problemów (por. Mycka, Piekarczyk 2004). Matematyczne rozumowania nie dowodzą, rzecz jasna, że odpowiednie rozwiązania mogą zostać fizycznie zaimplementowane. Stanowią jednak silną przesłankę na rzecz tezy, że jeśli w jakiejś dziedzinie (np. biologicznej) wielkości ciągłe rzeczywiście występują, to w dziedzinie tej faktycznie istnieją — być może naturalne, a być może czekające na skonstruowanie — układy rozwiązujące te problemy.

⁶ Zbiór ten ma moc *alef zero*, a więc jest równoliczny ze zbiorem liczb wymiernych, a także naturalnych.

2.2. Relacja obliczeń typu AN-E (empirycznych), tak jak rozumiemy je w tej pracy, do obliczeń cyfrowych pozostaje kwestią otwartą. Przede wszystkim na poziomie teorii uzasadniających przydatność określonych zjawisk naturalnych do realizacji obliczeń empirycznych (por 1.3) nie przesądza się, czy zjawiska te mają charakter dyskretny, czy ciągły. Odnosząc się zaś do konkretnych układów naturalnych, trzeba zauważyć, że nawet jeśli właściwe im obliczenia służą do realizacji typowych operacji analitycznych (czyli ciągłych), jak różniczkowanie⁷, to i tak nie można przyjąć z całą pewnością, że odpowiednie obliczenia (naturalne) są faktycznie ciągłe. Mogą być takimi tylko na poziomie teorii opisującej układ. Spostrzeżenie to prowadzi wprost do pewnej ważnej kwestii z zakresu filozofii informatyki, a mianowicie problemu zasadności tezy Churcha–Turinga (tezy C–T).

Jak wiadomo, teza ta głosi, że efektywnie obliczalne są tylko takie funkcje, które są obliczalne za pomocą maszyn Turinga, a zatem takie, dla których istnieje program uniwersalnej maszyny Turinga (UMT) opisujący sposób wyznaczania ich wartości. Ponieważ model UMT dotyczy obliczeń cyfrowych (a nie ciągłych), to teza C–T postuluje *de facto*, że obliczenia empiryczne AN-E, podobnie jak wszelkie inne obliczenia, mają ostatecznie charakter cyfrowy. A zatem, nawet jeśli wydaje nam się, że do zrealizowania pewnej operacji matematycznej używamy specjalnego procesu fizycznego, który jest ciągły (ponieważ np. opisuje go jakiś dział analizy matematycznej), to i tak sam fakt jego skutecznej realizacji przesądza o tym, że tak naprawdę nasze obliczenie ma charakter turingowski, czyli cyfrowy.

Rozważając tezę C–T w kontekście obliczeń empirycznych, które niejako z definicji polegają na wykorzystywaniu określonych właściwości przyrody, można wskazać dwa możliwe argumenty na rzecz jej (domniemanej) prawdziwości. Zgodnie z pierwszym teza C–T jest prawdziwa, ponieważ świat fizyczny (w którym dokonują się obliczenia) jest w istocie dyskretny i dlatego też zawsze istnieje jakiś wystarczająco efektywny opis realizacji procesu obliczeniowego w kategoriach dyskretnych, a dokładniej w postaci programu dla UMT⁸. Drugi

⁷ Historycznie rzecz biorąc, tak właśnie było. Używane do lat pięćdziesiątych XX w. urządzenia analogowe — np. analizatory różniczkowe czy integratory polowe — były konstruowane z myślą o obliczeniach ciągłych, dokonywanych na obiektach ciągłych, jak liczby rzeczywiste czy funkcje zmiennej rzeczywistej (Ifrah 2006).

⁸ Warto zauważyć, że rezultaty fizyki współczesnej nie wskazują jednoznacznie na prawdziwość (względnie fałszywość) tezy C–T, a w szczególności na niemożność (względnie możliwość) fizycznej realizacji obliczeń typu AN-C. Dostarczają zarówno argumentów za, jak i przeciw.

Do pierwszych trzeba zaliczyć argumenty nawiązujące do zasady holograficznej Bekensteina (1981), zgodnie z którą ilość informacji przechowywana w ograniczonym rejonie przestrzeni jest skończona. Może to stanowić przeszkodę w zapisie w pamięci urządzeń fi-

argument jest niezależny od struktury świata fizycznego: teza C–T miałyby być prawdziwa, ponieważ żadna realizowalna inżynieria obliczeniowa, w tym sposób odczytywania wyników obliczeń, nie może wykroczyć poza taką czy inną dyskretyzację. Z tego względu jedyne, co możemy osiągnąć, wykorzystując przyrodę obliczeniową, to urządzenia i programy turingowskie (z właściwymi im ograniczeniami).

Kwestię tę tylko sygnalizujemy, uważając jednak, że nie ma jakichś zasadniczych powodów po temu, aby tezę C–T przyjmować i tym samym negować przydatność kategorii analogowości w dziedzinie biologii czy obliczeń biologicznych. Teza ta nie została udowodniona (na podstawie jakichś oczywistych aksjomatów), a z punktu widzenia praktyki informatycznej ma uzasadnienie zaledwie indukcyjne, a więc nieostateczne. Polega ono na tym, że wszelkie skonstruowane dotychczas (ale nie wszelkie możliwe) efektywne urządzenia liczące podpadają pod model Turinga, a nawet pod modele prostsze, które nie przewidują jakichkolwiek nieskończonych zasobów maszyn (Harel 2000).

3. INFORMACJA W BIOLOGII MOLEKULARNEJ: DYSKRETNA I CIĄGŁA

3.1. Aby odnieść się do kategorii analogowości we współczesnej biologii molekularnej, niezbędne jest omówienie roli, którą pełni w niej pojęcie informacji. Z informatycznego punktu widzenia, omówionego w pierwszej części pracy, pojęcie informacji wiąże się nieodłącznie z pojęciami danych i obliczeń na nich⁹. Bez niego trudno wyobrazić sobie dziś uprawianie wielu istotnych

zycznych sygnałów czy kodów ciągłych (opisywanych matematycznie za pomocą liczb rzeczywistych o nieskończonych i nieregularnych rozwinięciach). Za wskazówkę tę autorzy dziękują jednemu z anonimowych recenzentów pracy.

Przeciw tezie C–T przemawia z kolei obliczeniowa interpretacja wyników Pour-El i Richardsa (1989), zgodnie z którymi pewne fale trójwymiarowe (opisywane szczególnymi równaniami różniczkowymi) mogą uzyskiwać stany charakteryzowalne wyłącznie za pomocą liczb nieobliczalnych w sensie Turinga (Turing 1936). To zaś sugeruje, że w przyrodzie istnieją wielkości nieobliczalne, które być może da się wykorzystać dla potrzeb obliczeń fizycznych (dopowiedzmy w nawiasie, że liczby nieobliczalne w sensie Turinga dopełniają dyskretny zbiór liczb obliczalnych do ciągłego *continuum*, w warstwie teoretycznej zatem stanowią matematyczny fundament obliczeń typu AN-C). W podobnym kierunku prowadzą analizy Johna Doyle'a, który proponuje obliczeniowe wykorzystywanie występujących w przyrodzie procesów osiągnięcia równowagi (np. termodynamicznej). Procesów tych nie można opisać za pomocą liczb i funkcji obliczalnych (Copeland 2002: 470). Szersze analizy wskazanych wyników zawiera np. praca (Stacewicz 2019).

⁹ W informatyce teoretycznej pojęcie danych jest zawsze osadzone w jakimś modelu obliczeń (np. cyfrowym), por. Stacewicz 2017a. W dalszej części artykułu koncentrujemy się

dyscyplin biologicznych, takich jak genetyka, biologia molekularna czy biologia komórki (Szathmáry, Maynard Smith 1995). Pojęcia zapożyczone z języka informatyki, takie jak „kod genetyczny” lub „program genetyczny”, zadomowiły się na stałe w języku potocznym i tekstach popularnonaukowych z zakresu biologii. O ile samo użycie pojęcia informacji w biologii nie budzi większych kontrowersji (aczkolwiek ma krytyków, np.: Sarkar 1996, Moffat 2009), to już powszechny sposób jego rozumienia i traktowania domaga się refleksji, a może nawet rewizji.

Punktem szczególnie spornym jest powszechnie przyjmowane (często bez jakiegokolwiek uzasadnienia) założenie, że informacja relewantna biologicznie – bioinformacja – ma charakter dyskretny, a jej nośnikami są kwasy nukleinowe DNA i RNA. Ze względów natury historycznej (siatka pojęciowa nowoczesnej biologii molekularnej ukształtowana została w latach pięćdziesiątych XX w. pod wpływem inspiracji płynących ze strony informatyki i teorii komunikacji) informacja determinująca rozwój i funkcjonowanie organizmu utosamiona została z molekularną informacją genetyczną (Cobb 2017, Downes 2006, Łuków, Żekanowski 2005). Założenie to było dodatkowo wspierane przez dynamiczny rozwój i spektakularne sukcesy genetyki, której fundamentem biochemicznym jest właśnie biologia molekularna. Mimo powszechności użycia pojęcia informacji ustalenie jednak, co dokładnie się pod nim kryje, nastrocza biologom wielu kłopotów i wciąż jest przedmiotem debat (por. Maynard Smith 2000 oraz odpowiedzi na ten artykuł publikowane na łamach *Philosophy of Science*).

W dalszej części artykułu wskazujemy trzy, naszym zdaniem istotne, przykłady informacji biologicznej o charakterze analogowym (czyli niemającej charakteru dyskretnego oraz nieopartej na kodzie). Pierwszym jest mechanizm splicingu alternatywnego oraz jego regulacja, drugim – proces zwijania białek, a trzecim – zjawisko allosterii. Termin „informacja” traktowany będzie przy tym w jego pierwotnym łacińskim znaczeniu (rzeczownik *informatio* od czasownika *informare*) jako nadawanie formy, kształtowanie. A zatem argumentujemy, że w procesach formowania matrycowego RNA, a następnie w procesach zwijania oraz funkcjonowania białek istotną rolę pełnią czynniki o charakterze analogowym. Przy czym, odnosząc się do przedstawionych wyżej rozróżnień informatycznych, czynniki te są najbliższe kontekstowi AN-C, czyli obliczeń analogowych ciągłych, polegających na przetwarzaniu wielkości ciągłych. Co więcej, pokazujemy również, że co najmniej jeden z tych procesów (zwijanie białek) potencjalnie może zostać wykorzystany jako narzędzie do

przede wszystkim na danych i ich nośnikach, formułując jednak pewne hipotezy dotyczące możliwości obliczeń realizowanych biologicznie.

rozwiązywania problemów NP-zupełnych, które wciąż leżą poza możliwościami obliczeniowymi współczesnych technik cyfrowych zgodnych z turinowskim modelem obliczeń. Te obserwacje z kolei nawiązują do przedstawionego wcześniej kontekstu obliczeń AN-E, czyli obliczeń empirycznych polegających na przetwarzaniu danych (i rozwiązywaniu problemów) za pomocą procesów naturalnych.

3.2. Z pewnością najszerszej opisanym i najlepiej poznanym mechanizmem informacyjnym działającym w świecie ożywionym jest mechanizm tzw. kodu genetycznego (Godfrey-Smith, Sterelny 2016). Termin ten służy do opisanie stałej zależności między kolejnością nukleotydów tworzących łańcuch DNA a kolejnością aminokwasów tworzących łańcuch białkowy (zwany też polipeptydowym od nazwy wiązania amidowego łączącego aminokwasy; dalej określenia te będą używane wymiennie). Dokładniej rzecz ujmując, łańcuch DNA zbudowany jest z czterech typów nukleotydów, w których nośnikami informacji są zasady azotowe: adenina, guanina, cytozyna oraz tymina. Ich kolejne trójki (tzw. tryplety lub kodony) odpowiadają poszczególnym 20 aminokwasom będącym cegiełkami, z których zbudowane są białka. Istnieje jednoznaczne przyporządkowanie między liniową sekwencją trypletów a liniową sekwencją aminokwasów w łańcuchu białkowym. Z tego względu struktura DNA traktowana jest jako materialny nośnik informacji o tzw. pierwszorzędowej strukturze białek.

Aby jednak informacja zapisana w DNA mogła zostać wykorzystana do stworzenia struktury białka, musi zajść pewna liczba procesów. Szczegóły tego mechanizmu zostały dobrze poznane, a używanie do jego opisu terminologii informacyjnej nie budzi kontrowersji. Proces odczytania informacji z kodującego białko fragmentu łańcucha DNA (czyli genu), a następnie etap jej obróbki oraz ostatecznie zbudowania łańcucha białkowego z aminokwasów nazywany jest ekspresją genu. Proces ten nie jest bynajmniej prosty i składa się z kilku podprocesów. Najważniejszymi z nich są transkrypcja oraz translacja. Sednem procesu transkrypcji jest przepisanie informacji z dwuniciowego łańcucha DNA na jednoniciowy łańcuch mRNA (matrycowego RNA). Główna różnica między obydwoma kwasami polega na tym, że w nukleotydach tworzących łańcuch RNA w miejscu tyminy występuje inna zasada azotowa — uracyl. Za proces transkrypcji odpowiedzialny jest enzym o nazwie polimeraza RNA. Istotą procesu translacji jest z kolei odczytanie informacji z mRNA i zbudowanie na jej podstawie łańcucha białkowego z odpowiednich aminokwasów. Łańcuch ten jest tożsamy z pierwszorzędową strukturą białka. Za translację odpowiedzialne są skomplikowane struktury wewnątrzkomórkowe zbudowane z kilkudziesięciu typów białek oraz rRNA (rybosomalnego RNA) zwane rybosomami.

Informacyjna interpretacja roli pełnionej przez kwasy nukleinowe (w szczególności DNA) w komórce nie ogranicza się do kodowania pierwszorzędowej struktury białek. Te ostatnie są bowiem fundamentalnymi składnikami zarówno strukturalnymi, jak i funkcjonalnymi komórek. Jako takie zaangażowane są w budowę kompletnego fenotypu. Tym samym DNA staje się niejako przyczyną całego żyjącego organizmu. W taki właśnie, skrajnie zresztą uproszczony sposób rola DNA przedstawiana bywa zarówno przez naukowców, jak i przez laików nauką zainteresowanych.

Kłopot w tym, że redukcja komórki do zestawu białek jest uproszczeniem. Do jej prawidłowego funkcjonowania niezbędny jest szereg związków chemicznych (w tym kwasy tłuszczowe, cukry, jony), które nie są bezpośrednio kodowane w DNA¹⁰. Dopiero niezwykle skomplikowana sieć wzajemnych zależności chemicznych między nimi tworzy kompletny fenotyp komórkowy lub organizm. Dlatego trudno go uznać za bezpośrednie następstwo odczytu informacji genetycznej zapisanej w DNA. Żywy organizm jest złożonym systemem współzależnych reakcji chemicznych, którego stabilność i funkcjonalność zależą od wielu rozmaitych czynników wewnętrznych i zewnętrznych (Roterman, Konieczny, Spólnik 2017).

3.3. Aby wskazać analogowy wymiar procesów obróbki informacji molekularnej, należy nieco bliżej przyjrzeć się ich szczegółom. Spróbujmy zatem skomplikować, a tym samym urealnić, naszkicowany wyżej obraz. Rozpoczniemy od niezwykle ciekawego procesu tzw. splicingu (składania mRNA) zachodzącego między procesami transkrypcji a translacji. Aby zrozumieć, czym jest, należy mieć na uwadze, że łańcuch RNA powstały w wyniku transkrypcji zawiera średnio (w przypadku człowieka) zaledwie kilka procent sekwencji kodujących białka (tzw. eksonów). Cała reszta to sekwencje niekodujące (tzw. introny). Ten pierwotny, „zaśmiecony” sekwencjami niekodującymi transkrypt nazywany jest pre-mRNA (prekursorowym mRNA)¹¹. Aby mógł on posłużyć za wzorec do budowy łańcucha białkowego, musi zostać uprzednio oczyszczony z intronów. Za proces ich usuwania odpowiedzialne są złożone kompleksy białek i snRNA (małego jądrowego RNA) zwane spliceosomami. W wyniku wycięcia intronów powstaje łańcuch RNA składający się wyłącznie z eksonów. Aby jeszcze bardziej zbliżyć ten obraz do rzeczywistości, wspomnijmy, że

¹⁰ Aczkolwiek można by je uznać za pośredni efekt odczytu informacji zapisanej w DNA (np. nieorganiczne jony z otoczenia wnikają do komórki w sposób selektywny — tzn. przenoszone są przez białka tworzące kanały w błonie komórkowej).

¹¹ Opisana sytuacja występuje wyłącznie w przypadku „zaawansowanych” ewolucyjnie organizmów eukariotycznych. U prokariotów (bakterii i archeonów) transkrypt jest pozabawiony intronów.

eksony pozostałe po usunięciu intronów mogą (choć nie muszą) zostać uporządkowane na wiele sposobów. Jeżeli tak się dzieje, to proces ten nazywamy alternatywnym splicingiem. Rekordzistą jest tu gen DSCAM muszki owocówki (*D. melanogaster*), dla którego zidentyfikowano ponad 38 tys. wariantów splicingowych (Schmucker i in. 2000). Szacuje się, że u człowieka alternatywnemu splicingowi podlega co najmniej 35% wszystkich genów (Mironov, Fickett, Gelfand 1999). Po odpowiednim uporządkowaniu eksonów powstaje dojrzały łańcuch mRNA mogący służyć jako matryca, na podstawie której rybosomy zbudują odpowiednie białko¹².

Mechanizm alternatywnego splicingu pozwala na kodowanie większej liczby białek, niż wynikałoby to z liczby genów obecnych w komórce. Tym samym jest niezwykle ważnym źródłem różnorodności białek umożliwiających poprawne funkcjonowanie organizmów. Z punktu widzenia rozważanych tu kwestii niezwykle ciekawym przykładem analogowych aspektów funkcjonowania żywej komórki są mechanizmy regulowania splicingu. Nie ma tu miejsca na choćby krótką próbę charakterystyki szczegółów rozmaitych mechanizmów regulacyjnych, dlatego skupiamy się wyłącznie na czynnikach, których analogowość nie budzi, w naszej ocenie, wątpliwości.

Zacznijmy od tego, że funkcje regulacyjne pełni już sama sekwencja nukleotydowa łańcucha pre-mRNA. Łańcuch taki zawiera miejsca splicingowe, czyli fragmenty, w których może zostać zainicjowana operacja wycinania intronów. Istotą alternatywnego splicingu jest wybór spośród konkurencyjnych miejsc splicingowych. W wyborze odpowiedniego miejsca biorą udział fragmenty łańcucha określane mianem SRE (*Splicing Regulatory Elements*). Wpływają one na proces splicingu, oddziałując z białkami zaangażowanymi w proces wycinania intronów. Poza nimi w regulacji procesu alternatywnego splicingu biorą udział również inne czynniki o charakterze zarówno wewnątrzkomórkowym, jak i pozakomórkowym. Do tych pierwszych możemy zaliczyć tak fundamentalne procesy komórkowe, jak transkrypcja i mitoz. Ważną rolę regulacyjną w procesach alternatywnego splicingu odgrywa zarówno stopień ufosforylowania pewnych istotnych białek (polimerazy RNA II), jak i stopień defosforylacji innych (Srp38) (Szcześniak, Szweykowska-Kulińska 2009). Wybór jednego z wariantów splicingowych może być także modulowany przez analogowe czynniki zewnętrzne: hormony, czynniki wzrostu, odpowiedź immunologiczną, stres komórkowy. Jednym z najdokładniej opisa-

¹² Porządek eksonów, czyli ich kolejność, nie ulega zmianie. Alternatywny splicing polega na zachowaniu — bądź wycięciu wraz z sąsiednimi intronami — danego eksonu w dojrzałym RNA. Również intron może pozostać niewycięty, stając się tym samym częścią eksonu. Możliwe jest także włączenie części (jednego z końców) eksonu do sąsiedniego intronu i usunięcie jej wraz z nim (analogiczna sytuacja może zajść dla fragmentu intronu).

nych przykładów regulacji alternatywnego splicingu przez czynniki zewnątrz-komórkowe jest wpływ insuliny na wybór wariantów splicingowych dla kinazy białkowej C (Szcześniak, Szweykowska-Kulińska 2009).

Okazuje się zatem, że informacja niezbędna do zbudowania konkretnego łańcucha białkowego nie jest zakodowana wyłącznie w kwasach nukleinowych. W wielu przypadkach musi być czerpana skądinąd, a te inne źródła dostarczają jej w postaci ciągłej. Można zatem zaryzykować stwierdzenie, że każda próba opisu alternatywnego splicingu w kategoriach obliczeniowych powinna uwzględniać fakt, że jest to jakiś rodzaj obliczeń typu AN-C, a więc obliczeń dostosowanych do przetwarzania wielkości ciągłych (aczkolwiek stwierdzeniu temu daleko do pewności, jako że wciąż zbyt mało wiemy o mechanizmach regulacji splicingu alternatywnego).

4. CZYNNIKI ANALOGOWE W PROCESIE FORMOWANIA I FUNKCJONOWANIA BIAŁEK

W trakcie opisanego wyżej procesu translacji powstaje łańcuch polipeptydowy zbudowany z aminokwasów. On sam oraz w pełni funkcjonalne białko nie są ze sobą tożsame. Aby zrozumieć, dlaczego tak się dzieje, należy uświadomić sobie znaczenie, jakie w powstaniu w pełni funkcjonalnego białka pełni organizacja przestrzenna łańcucha polipeptydowego.

4.1. Częsteczka białka jest obiektem o niezwykle złożonej, hierarchicznej strukturze. Wyróżnić w niej można nadbudowane nad sobą kolejne poziomy organizacji przestrzennej: od tzw. struktury I-rzędowej, czyli liniowej kolejności aminokwasów w łańcuchu polipeptydowym, aż do struktury IV-rzędowej, czyli organizacji dużych białek składających się z kilku osobnych łańcuchów (np. hemoglobina czy polimeraza RNA). Dalej przyjrzymy się bliżej strukturze III-rzędowej, która z wielu powodów uznana może być za istotny poziom organizacji przestrzennej białek. Aby zrozumieć, dlaczego tak jest, przypatrzmy się przestrzennej organizacji cząsteczki białka. Struktura I-rzędowa to po prostu liniowa kolejność aminokwasów połączonych wiązaniami peptydowymi. Kolejność aminokwasów tworzących łańcuch odtwarza kolejność kodonów (trypletów) w mRNA powstałym na bazie DNA zgodnie ze sztywnym przyporządkowaniem – kodem genetycznym. Białko w tej pierwotnej postaci nie jest jeszcze aktywne chemicznie, tzn. nie pełni w komórce przypisanych mu funkcji. Aby tak się stało, nowo utworzony łańcuch aminokwasowy musi uzyskać odpowiednią strukturę przestrzenną determinującą jego reaktywność

chemiczną — musi ulec zwinięciu. Pierwszym krokiem do tego jest przybranie przez niego struktury II-rzędowej. Mianem tym określa się sposób uorganizowania poszczególnych fragmentów łańcucha polipeptydowego za sprawą powstawania wiązań wodorowych (między atomami tlenu jednych grup peptydowych a atomami wodoru innych grup peptydowych). W wyniku powstania tych wiązań w łańcuchu wykształcają się charakterystyczne struktury przestrzenne: α -helisy oraz β -karkty.

Proces zwijania białka wciąż nie jest ukończony. Charakterystyczne II-rzędowe struktury przestrzenne muszą bowiem zostać jeszcze spójczone względem siebie. Dokonuje się to za sprawą oddziaływań międzycząsteczkowych: wodorowych, jonowych, hydrofobowych oraz sił van der Waalsa (Alberts i in. 2005). W wyniku tych procesów łańcuch aminokwasowy uzyskuje ostateczną formę przestrzenną — strukturę III-rzędową, która warunkuje jego funkcjonalność.

Większość białek dopiero po jej przyjęciu może stać się aktywnymi uczestnikami rozmaitych procesów chemicznych zachodzących w żywej komórce. Topologia zwiniętego łańcucha jest zatem czynnikiem warunkującym pożądaną funkcjonalność białka. Zwinięcie łańcucha aminokwasowego w sposób choćby nieco odmienny od wzorcowego może skutkować brakiem aktywności białka. Najczęściej występującą *in vivo* strukturę III-rzędową danego białka nazywa się strukturą natywną; tym terminem będziemy się posługiwać dalej w tekście.

4.2. Zobaczymy teraz, co determinuje topologię białka w postaci natywnej. Po raz pierwszy odpowiedzi na pytanie o to, czy wystarczające są czynniki o charakterze dyskretnym (kolejność zasad w łańcuchu DNA), czy też należy również uwzględnić czynniki inne (a jeżeli tak, to jakie) udzielił w latach sześćdziesiątych XX w. Christian Anfinsen, który wraz ze współpracownikami badał procesy denaturacji i renaturacji krótkiego (124 aminokwasy) białka — rybonukleazy A. W wyniku tych badań sformułował hipotezę głoszącą, że struktura III-rzędowa determinowana jest w całości przez strukturę I-rzędową. Przyjął dodatkowo, że struktura natywna białka jest również formą o najniższej energii swobodnej możliwej do przyjęcia przez łańcuch polipeptydowy. Cała informacja niezbędna do jej osiągnięcia zawarta jest już w liniowej sekwencji aminokwasów tworzących dany łańcuch. Hipoteza ta funkcjonuje dziś pod nazwą dogmatu Anfinsena. Jedną z jej konsekwencji jest przekonanie, że zasadniczo dla dowolnego łańcucha polipeptydowego można na podstawie samej tylko znajomości liniowego uporządkowania tworzących go aminokwasów jednoznacznie określić właściwą dla niego strukturę natywną (Anfinsen 1973). Interpretując to zagadnienie obliczeniowo, uzyskujemy twierdzenie, że aby od-

należć natywną strukturę dowolnego białka, wystarczy wygenerować wszystkie możliwe konformacje jego łańcucha polipeptydowego, następnie obliczyć energię swobodną dla każdej z nich oraz wskazać strukturę o najniższej energii swobodnej, która będzie szukaną postacią natywną. Problem z taką procedurą polega na tym, że choć jest ona wykonalna teoretycznie, to pozostaje niewykonalna praktycznie (przynajmniej w modelu obliczeń cyfrowych)¹³.

Rozważmy nieduże białko o długości 100 aminokwasów. Przyjmując, że każdy z tworzących je aminokwasów może być w trzech różnych stanach (konformacjach), otrzymujemy olbrzymią przestrzeń stanów całego układu wyrażaną się liczbą $3^{100} = 5 \times 10^{47}$ możliwych konformacji. Gdyby przeszukiwać tę przestrzeń z prędkością 1 konformacji na 10^{-13} sekundy (0,1 pikosekundy – czas, w którym światło pokonuje drogę ok. 0,03 mm), to odkrycie stanu o najniższej energii swobodnej zajęłoby $5 \times 10^{47} \times 10^{-13} \text{s} = 1,6 \times 10^{27}$ lat, co jest czasem wielokrotnie dłuższym niż znany nam czas istnienia Wszechświata¹⁴ (por. Zwanzig, Szabo, Bagchi 1992). Tymczasem każda żywa komórka syntetyzuje, a następnie zwija do postaci natywnej wiele rozmaitych łańcuchów białkowych w każdej minucie swego istnienia – najkrótsze znane czasy zwijania białek lokują się w skali mili-, a nawet mikrosekund (Kubelka, Hofrichter, Eaton 2004). Fakt zasadniczej rozbieżności między czasem potrzebnym (teoretycznie) do znalezienia postaci natywnej a rzeczywistą szybkością i powszechnością zwijania białek został po raz pierwszy opisany przez amerykańskiego biologa molekularnego Cyrusa Levinthala (1968, 1969) i od jego nazwiska nazywany jest paradoksem Levinthala.

Pod koniec ubiegłego stulecia izraelski matematyk Aviezri Fraenkel wykazał, że problem zwijania białek jest NP-zupełny, uznając jednocześnie zdolność przyrody do rozwiązywania niezwykle złożonych obliczeniowo problemów za „niepokojącą” (*uncanny*) (Fraenkel 1993). Zauważa, że badacze próbujący uporać się z paradoksem Levinthala stają zazwyczaj na jednym z następujących dwóch stanowisk: (1) przyroda potrafi rozwiązywać problemy NP-zupełne w czasie wielomianowym, (2) przyroda funkcjonuje w ramach turingowskiego paradygmatu obliczeń. Przyjęcie pierwszego poglądu (i traktowanie procesu zwijania białka jako znajdowania minimum energetycznego) nie musi być

¹³ Warto zauważyć, że dla każdego modelu obliczeń (w tym cyfrowego) należy odróżnić (1) problemy nierozwiązywalne zasadniczo, w przypadku których nie istnieją uniwersalne algorytmy do rozwiązywania wszelkich przypadków szczególnych danego zagadnienia (w modelu cyfrowym takim jest np. problem stopu maszyn Turinga), od (2) problemów nierozwiązywalnych praktycznie, w przypadku których algorytmy takie istnieją, ale wszystkie one wymagają użycia nieosiągalnych praktycznie zasobów systemu, jak pamięć czy czas (Harel 2000). Zagadnienie opisane w tekście głównym należy do drugiej kategorii.

¹⁴ Obecnie przyjmuje się, że ten równy jest w przybliżeniu 13,85 miliarda lat.

jednak tożsame z przyjęciem, że odbywa się to przez przeszukanie całej jego przestrzeni stanów (można np. przyjąć, że odbywa się tu losowe przeszukiwanie przestrzeni stanów, tak jak rozwiązuje się niektóre problemy optymalizacyjne w praktyce informatycznej)¹⁵. Znacznie płodniejsze poznawczo, a na pewno bardziej użyteczne, jest zdaniem Fraenkela potraktowanie takich układów fizycznych, jakimi są białka — niezależnie od tego, jakie dokładnie są mechanizmy ich błyskawicznego zwijania — jako obliczeniowych „czarnych skrzynek” potrafiących rozwiązywać niezwykle złożone problemy w zaskakująco krótkim czasie. Podejście takie prowadzi z kolei do pytania o praktyczną możliwość opracowania narzędzi i interfejsów pozwalających kodować, a następnie wprowadzać do owych „skrzynek” przykłady problemów NP-zupełnych celem ich rozwiązania w krótkich skalach czasowych. Sam proces obliczania byłby w takim wypadku tożsamy z termodynamicznym procesem znajdowania minimum energetycznego dla danego łańcucha polipeptydowego umieszczonego w konkretnym środowisku. Zgodnie z rozróżnieniami omówionymi w pierwszej części artykułu byłyby to biologiczne obliczenia typu AN-E (empiryczne), czerpiące swoje uzasadnienie z wiarygodnego biochemicznie opisu procesu zwijania białek (będącego fragmentem pewnej teorii biologicznej)¹⁶. Obliczenia te można byłoby stosować do rozwiązywania wszelkich problemów, których matematyczny opis byłby izomorficzny z opisem problemu zwijania jakiegoś konkretnego

¹⁵ Warto w tym miejscu przedstawić hipotezy tłumaczące zdaniem Fraenkela przyjęcie stanowiska drugiego. Wymaga ono wyjaśnienia zdolności przyrody do rozwiązywania problemów NP-zupełnych bez konieczności utożsamiania klasy problemów decyzyjnych rozwiązywalnych w czasie wielomianowym (P) z klasą trudniejszych problemów decyzyjnych, dla których rozwiązania można tylko weryfikować w takim czasie (NP). Jako możliwe wyjaśnienia tej kwestii Fraenkel podaje: (a) hipotezę, że zwijając białka, natura nie osiąga globalnego minimum energetycznego, zadowalając się odpowiednio bliskim minimum lokalnym; (b) hipotezę, że duże łańcuchy białkowe nie są zwijane w całości, lecz dzielone na mniejsze części, które wymagają znacząco mniejszych nakładów czasu do zwinięcia; (c) hipotezę, że nie wszystkie postaci problemu zwijania białek muszą wymagać czasu wielomianowego do weryfikacji rozwiązania (aby dany problem został uznany za NP-zupełny, nie musi być on uniwersalnie NP-zupełny), być może zatem ewolucja w jakiś sposób sprzyja zachowaniu wyłącznie białek o wielomianowych czasach potrzebnych do znalezienia rozwiązań problemu zwijania, a odrzuca pozostałe; (d) hipotezę, że mechanizm zwijania danego białka może być jakoś zakodowany w odnośnej sekwencji DNA, lecz ów kod pozostaje wciąż nieznanym (Fraenkel 1993).

¹⁶ Oczywiście opis w kategoriach chemicznych nie może uchodzić za ścisły w tym samym sensie, w jakim ścisłe są modele kwantowo-mechaniczne (które jak dotąd stosowane były do przewidywania trójwymiarowych struktur bardzo krótkich, zaledwie kilkuaminokwasowych peptydów, tzw. 3₁₀ helis; por. Bour, Kubelka, Keiderling 2002. Niemniej, zarówno ze względu na powszechne użycie narzędzi matematycznych, jak i osadzenie w paradygmacie mechaniki kwantowej, modele i opisy chemiczne można uznać za sformalizowane (tym samym zaś „użyteczne” dla obliczeń typu AN-E).

białka. Byłyby one najbardziej użyteczne wówczas, gdyby w ramach wspomnianych wyżej interfejsów można było manipulować pewnymi parametrami fizykochemicznymi, uzyskując w ten sposób różne uszczegółowienia danego problemu optymalizacyjnego. Z metodologicznego punktu widzenia warto zauważyć, że przedstawione rozumowanie stanowi silny (aczkolwiek wciąż hipotetyczny) argument na rzecz tezy, że biologia może pomóc informatyce w pokonaniu pewnych trudności tradycyjnych obliczeń cyfrowych¹⁷.

4.3. Powróćmy do pytania o czynniki determinujące topologię postaci natywnej białka. Hipoteza sformułowana przez Anfinsena podkreśla wystarczającą rolę informacji zawartej w strukturze pierwszorzędowej. W momencie jej powstania biolodzy nie wiedzieli jednak o istnieniu chaperonów (białek opiekuńczych) — wyspecjalizowanej grupy białek pełniących w komórce niezwykle ważną rolę. Jej odkrycie i opisanie dokonane zostało dopiero w latach osiemdziesiątych XX w. (Ellis 1996). Zanim odpowiemy na pytanie, czym są chaperony i dlaczego odgrywają tak istotną rolę w procesie formowania struktury natywnej białek, przyjrzyjmy się bliżej mechanizmowi zwijania białek. Pozwoli to zrozumieć rolę, jaką odgrywają w nim białka opiekuńcze.

Każdy z 20 aminokwasów jest zbudowany w podobny sposób: w jego centrum tkwi atom węgla połączony z grupą aminową i karboksylową oraz łańcuchem bocznym. Łańcuchy boczne mogą mieć charakter niepolarny (hydrofobowy) lub polarny (hydrofilowy). Aminokwasy tworzące łańcuch białkowy są połączone wiązaniami kowalencyjnymi (tzw. wiązaniami peptydowymi), które umożliwiają swobodną rotację atomów nimi połączonych. Z tego względu nawet krótkie łańcuchy polipeptydowe mogą związać się na olbrzymią liczbę sposobów. Podstawową siłą wymuszającą zwijanie łańcucha białkowego jest polarność łańcuchów bocznych. Aminokwasy o charakterze hydrofobowym mają tendencję do „uciekania” od wody i skupiania się we wnętrzu zwijającej się cząsteczki białka, natomiast aminokwasy hydrofilowe eksponują swoje łańcuchy boczne na zewnątrz — ku środowisku wodnemu wypełniającemu wnętrze żywej komórki. Innymi słowy, hydrofobowość części aminokwasów tworzących łańcuch białkowy jest siłą napędową procesu zwijania, niejako wymuszając na białku przyjęcie odpowiedniej topologii.

¹⁷ Na marginesie tych rozważań można postawić dodatkową hipotezę, że być może za siłę obliczeń analogowych typu AN-E (czyli za fakt, że pozwalają one szybko rozwiązać problem NP-zupełny) odpowiada to, że są to jednocześnie obliczenia ciągłe. Gdyby hipoteza ta okazała się prawdziwa, wówczas moglibyśmy wzmocnić nasz postulat badawczy, że biologia potrzebuje informatycznego pojęcia analogowości—ciągłości i opisu pewnych procesów biologicznych za pomocą takich czy innych modeli obliczeń ciągłych.

Białka opiekuńcze są z wielu względów bardzo ważne dla normalnego funkcjonowania komórki: rozwijają zwinięte łańcuchy białkowe w celu ich transportu przez błony wewnątrzkomórkowe, zapobiegają również denaturacji białek w wyniku działania wysokich temperatur (tzw. białka szoku cieplnego — HSP). Przede wszystkim jednak umożliwiają innym białkom sprawne zwinięcie do postaci natywnej. W trakcie procesu zwijania łańcuchowi polipeptydowemu grozi wielokrotne uwięzienie w lokalnych minimach energetycznych. Rolą białek opiekuńczych jest zadbanie, aby do tego nie doszło.

Co więcej, dla pewnych białek istnieją wyspecjalizowane typy chaperonów (tzw. chaperony steryczne) aktywnie wpływające na konformację zwijanego białka (dzieje się tak np. w przypadku bakteryjnych proteaz: α LP odkrytej u *Lysobacter enzymogenes* oraz SGBP odkrytej u *Streptomyces griseus*; Jaswal i in. 2005). Dołączają one do wybranych fragmentów w zwijanym łańcuchu białkowym, lokalnie modyfikując warunki fizykochemiczne, w wyniku czego powstają słabe wiązania stabilizujące topologię białka, które inaczej by nie powstały¹⁸. Po odłączeniu chaperonu wiązania te są zachowywane. Zidentyfikowano również białka, które po zwinięciu do postaci natywnej mają wyższą energię swobodną aniżeli przed nim, co podważa intuicję Anfinsena, że białka fałdują się zawsze do postaci o najniższej energii swobodnej (Sohl, Jaswal, Agard 1998). Również w tym procesie zasadniczą rolę odgrywają chaperony steryczne, pozwalające niejako „przeskoczyć” barierę energetyczną i zwinąć białko do stabilnej postaci o wyższej energii swobodnej. Białko opiekuńcze jest zatem dla zwijanego łańcucha źródłem informacji sterycznej (strukturalnej), co potraktować można jako naruszenie dogmatu Anfinsena głoszącego, że cała informacja niezbędna do poprawnego zwinięcia białka zawarta jest w jego strukturze I-rzędowej — a więc zależy wyłącznie od informacji dyskretnej (Pauwels i in. 2007). Być może zatem sztywną regułę „struktura I-rzędowa w sposób konieczny i wystarczający determinuje strukturę natywną” powinniśmy osłabić, usuwając z niej warunek wystarczalności, uwzględnivszy fakt, że działają tu również inne czynniki mające wpływ na topologię zwijanego łańcucha białkowego (własności ośrodka, w którym białko się zwija, a następnie funkcjonuje). Z naszego punktu widzenia zasadnicze znaczenie ma fakt, że dyskretna kodowana informacja zapisana w kwasach nukleinowych okazuje się niewystarczająca do stworzenia w pełni funkcjonalnego białka. Brakująca informacja dostarczana jest przez obiekty fizyczne (chaperony) ingerujące w termodynamikę procesu zwijania białek.

¹⁸ Dokładniej rzecz ujmując: mogłyby powstać, ale w bardzo dużych skalach czasowych. Czas potrzebny do osiągnięcia postaci natywnej dla proteazy α LP bez wspomagania wynosi ok. 1700 lat. Wobec tego spontaniczne zwijanie po prostu nie może mieć miejsca (Jaswal i in. 2005).

W związku z tym, jeśli zakłada się obliczeniowy punkt widzenia, zasadne wydaje się przyjęcie takiego kierunku stawiania dalszych hipotez co do przebiegu procesów warunkujących powstawanie białek, który uwzględniałby różne modele obliczeń typu AN-C. Modele tego typu mogłyby być testowane, ale także modyfikowane, ze względu na adekwatny opis różnych etapów omówionego procesu. Byłoby to dojrzałe metodologicznie odniesienie do informatycznej, a więc obliczeniowej, kategorii analogowości.

4.4. Po przyjęciu struktury natywnej i trafieniu do docelowego miejsca w komórce białko zaczyna pełnić swoją funkcję¹⁹. Jak widzieliśmy, aktywność chemiczna białka jest pochodną topologii zwiniętego łańcucha polipeptydowego. Ustaliliśmy także, że przyjęcie natywnej struktury III-rzędowej nie zależy wyłącznie od liniowej sekwencji aminokwasów (a przynajmniej, że stwierdzenie takie wzięte dosłownie jest nie do utrzymania). Dalej uzupełnimy ten obraz o zjawisko allosterii, czyli modyfikacji konformacji (topologii) zwiniętego białka w wyniku dołączenia bądź odłączenia substancji chemicznej zwanej efekтором allosterycznym. W celu objaśnienia posłużymy się dobrze zbadanym i opisanym przykładem funkcjonowania niedużego (148 reszt aminokwasowych) białka powszechnie występującego u organizmów eukariotycznych — kalmoduliny. Na istotę funkcjonowania tego białka naprowadza już sama nazwa będąca skrótowcem od *calcium-modulated protein*. Kalmodulina jest przedstawicielem dużej grupy białek wiążących wapń (CaBP — *Calcium Binding Proteins*) niezbędnych dla życia organizmów eukariotycznych i powszechnie spotykanym w płynie wewnątrzkomórkowym (cytozolu) komórek eukariotycznych. Część cząsteczki kalmoduliny ma kształt hantli o dwóch globularnych końcówkach połączonych prostą α -helisą. Na każdym z końców znajdują się dwa miejsca wiążące jony wapnia (Ca^{2+}). Po przyłączeniu czterech jonów kalmodulina podlega zmianie konformacyjnej umożliwiając jej wiązanie się z różnymi białkami docelowymi, a w rezultacie zmianę ich aktywności²⁰. Topologia kalmoduliny oraz zależna od niej funkcjonalność jest

¹⁹ Zanim tak się stanie, białka są jeszcze modulowane chemicznie. Modulacje te opatrzuje się zbiorczą nazwą modyfikacji potranslacyjnych. Białka mogą być modyfikowane na wiele rozmaitych sposobów, które wpływają na ich własności fizyczne i chemiczne, aktywność oraz stabilność, a przez nie — na ich funkcjonowanie. Nie ma tu miejsca, aby szerzej potraktować tę obszerną klasę zjawisk, niemniej wspominamy o niej, ponieważ regulacja funkcjonowania białek w ten sposób jest pięknym przykładem analogowych procesów o charakterze informacyjnym.

²⁰ U człowieka najwięcej kalmoduliny i białek z nią oddziałujących zawiera tkanka nerwowa. Wśród białek zależnych od kalmoduliny znajduje się m.in. kinaza II odpowiedzialna za fosforylację (przyłączanie reszt fosforanowych do aminokwasów) rozmaitych białek istotnych dla poprawnego funkcjonowania mózgu (Kuźnicki, Filipek 1997).

więc ściśle sprzężona ze stężeniem Ca^{2+} w komórce. Również prawdopodobieństwo zmodyfikowania białek zależnych od kalmoduliny (oraz tego, czym owa modyfikacja dalej skutkuje) jest funkcją stężenia jonów wapnia. Tym samym prosty parametr analogowy (stężenie Ca^{2+} w cytozolu), którego wartość zmienia się w sposób ciągły, pośrednio „zarządza” wieloma procesami wewnątrzkomórkowymi (Chin, Means 2000).

PODSUMOWANIE

Przedstawione analizy pozwalają sformułować pewne wnioski ogólne. Po pierwsze, odnosząc się do opisanego procesu tworzenia, formowania i regulowania aktywności białek, należy stwierdzić, że dla funkcjonowania żywej komórki, a ostatecznie złożonego z komórek organizmu, niezbędna jest zarówno informacja dyskretna (zakodowana w łańcuchach DNA), jak i analogowa – rozproszona i oparta na ciągłych wielkościach fizycznych, takich jak konfiguracje przestrzenne, stężenia, pH czy temperatury²¹. Dlatego też w biologii molekularnej potrzebne jest teoretyczne zastosowanie jakiegoś pojęcia analogowości (związanego z ciągłością).

Po drugie, w informatyce teoretycznej istnieją dobrze określone i ufundowane matematycznie modele obliczeń analogowych ciągłych, takie jak GPAC, EAC czy BSS, przy czym określona teoretycznie moc obliczeniowa tych modeli, rozumiana tutaj jako klasa rozwiązywanych za ich pomocą problemów, jest wyższa niż moc obliczeń cyfrowych (turingowskich). Tak więc po stronie informatyki istnieje teoretyczny potencjał do ścisłego i użytecznego operacyjnie zastosowania pewnych modeli obliczeń do opisu analogowych zjawisk biologicznych. W tej pracy nie wskazano wprawdzie konkretnego modelu, który można by zastosować, uzasadniono jednak sensowność rozważania tego rodzaju modeli (w ramach pewnych postulatów badawczych, por. 4.3).

Po trzecie, na styku biologii i informatyki można wskazać pewne procesy (w artykule opisano proces zwijania białek), które mają uzasadnioną teoretycznie interpretację obliczeniową, to znaczy pozwalają – przy pewnej interpretacji określonych parametrów fizykochemicznych – rozwiązywać problemy *stricte* obliczeniowe. Można je zatem opisywać, a w pewnych warunkach być może nawet wykorzystywać, jako obliczenia analogowe empiryczne (por. 4.2). Istnienie takich procesów, po pierwsze, wzmacnia tezę, że interakcja in-

²¹ Dobrze znana i opisana w literaturze jest zależność aktywności wielu enzymów oraz katalizowanych przez nie reakcji chemicznych od parametrów fizykochemicznych środowiska, w którym funkcjonują, takich jak temperatura, pH bądź siła jonowa (Eed 2012).

formatyki i biologii jest poznawczo i aplikacyjnie pożądana, a po drugie prowadzi do domysłu (w kontekście badawczym do hipotez), że za efektywność obliczeniową pewnych procesów odpowiada ich ciągłość, a tę można opisać w kategoriach informatycznych.

BIBLIOGRAFIA

- Alberts B., Bray D., Hopkins K., Johnson A., Lewis J., Raff M., Roberts K., Walter P. (2005), *Podstawy biologii komórki*, przekład zbiorowy pod redakcją H. Kmity i P. Wojtaszka, Warszawa: Wydawnictwo Naukowe PWN.
- Anfinsen C. B. (1973), *Principles that Govern the Folding of Protein Chains*, „Science” 181(4096). <https://doi.org/10.1126/science.181.4096.223>
- Bekenstein J. D. (1981), *Universal Upper Bound on the Entropy-to-Energy Ratio for Bounded Systems*, „Physical Review” D 23(2), 287-298. <https://doi.org/10.1103/PhysRevD.23.287>
- Blum L., Shub M., Smale S. (1989), *On a Theory of Computation and Complexity over the Real Numbers: NP—Completeness, Recursive Functions and Universal Machines*, „Bulletin of the American Mathematical Society (New Series)” 21, 1-46. <https://doi.org/10.1090/S0273-0979-1989-15750-9>
- Bour P., Kubelka J., Keiderling T. A. (2002), *Ab Initio Quantum Mechanical Models of Peptide Helices and Their Vibrational Spectra*, „Biopolymers” 65(1), 45-59. <https://doi.org/10.1002/bip.10224>
- Chin D., Means A. R. (2000), *Calmodulin: A Prototypical Calcium Sensor*, „Trends in Cell Biology” 10, 322-328. [https://doi.org/10.1016/S0962-8924\(00\)01800-6](https://doi.org/10.1016/S0962-8924(00)01800-6)
- Cobb M. (2017), *Największa tajemnica życia. Jak rozszyfrowano kod genetyczny*, tłum. A. Tuz, Warszawa: Prószyński i S-ka.
- Copeland J. (2002), *Hypercomputation*, „Mind and Machines” 12(4), 461-502. <https://doi.org/10.1023/A:1021105915386>
- Costa J. F., Graça D. (2003), *Analog Computers and Recursive Functions over the Reals*, „Journal of Complexity” 5(19), 644-664. [https://doi.org/10.1016/S0885-064X\(03\)00034-7](https://doi.org/10.1016/S0885-064X(03)00034-7)
- Deutsch D. (1985), *Quantum Theory, the Church—Turing Principle and the Universal Quantum Computer*, „Proceedings of the Royal Society A” 400, 97-117. <https://doi.org/10.1098/rspa.1985.0070>
- Downes S. M. (2006), *Biological Information [w:] The Philosophy of Science: An Encyclopedia*, S. Sarkar, J. Pfeifer (eds.), New York: Routledge, 64-68.
- Eed J. (2012), *Factors Affecting Enzyme Activity*, „ESSAI” 10, 48-51.
- Ellis R. J. (1996), *Discovery of Molecular Chaperones*, „Cell Stress & Chaperones” 1(3), 155-160. [https://doi.org/10.1379/1466-1268\(1996\)001<0155:DOMC>2.3.CO;2](https://doi.org/10.1379/1466-1268(1996)001<0155:DOMC>2.3.CO;2)
- Fraenkel A. S. (1993), *Complexity of Protein Folding*, „Bulletin of Mathematical Biology” 55(6), 1199-1210. [https://doi.org/10.1016/S0092-8240\(05\)80170-3](https://doi.org/10.1016/S0092-8240(05)80170-3)
- Godfrey-Smith P., Sterelny K. (2016), *Biological Information [w:] The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), Edward N. Zalta (ed.), <https://stanford.io/365r6zg>.

- Harel D. (2000), *Rzecz o istocie informatyki. Algorytmika*, Warszawa: Wydawnictwa Naukowo-Techniczne.
- Hogarth M. (1994), *Non-Turing Computers and Non-Turing Computability*, „PSA” 1, 126-138. <https://doi.org/10.1086/psaprobienmeetp.1994.1.193018>
- Ifrah G. (2006), *Historia powszechna cyfr*, t. 2, tłum. K. Marczevska, K. Szeżyńska-Mačkowiak, Warszawa: W.A.B.
- Jaswal S. S., Truhlar S. M. E., Dill K. A., Agard D. A. (2005), *Comprehensive Analysis of Protein Folding Activation Thermodynamics Reveals a Universal Behavior Violated by Kinetically Stable Proteases*, „Journal of Molecular Biology” 347, 355-366. <https://doi.org/10.1016/j.jmb.2005.01.032>
- Kari L., Rozenberg G. (2008), *The Many Facets of Natural Computing*, „Communications of the ACM” 10(51), 72-83. <https://doi.org/10.1145/1400181.1400200>
- Kozłowski J. (2011), *Czy teorię ewolucji można zmatematyzować?* [w:] *Ewolucja życia i ewolucja Wszechświata*, J. Mączka, P. Polak (red.), Kraków: Copernicus Center Press, 75-84.
- Kubelka J., Hofrichter J., Eaton W. A. (2004), *The Protein Folding “Speed Limit”*, „Current Opinion in Structural Biology” 14, 76-88. <https://doi.org/10.1016/j.sbi.2004.01.013>
- Kulka Z., Nadachowski N. (1979), *Analogowe układy scalone*, Warszawa: Wydawnictwa Komunikacji i Łączności.
- Kulka Z., Nadachowski M. (1982), *Wzmacniacze operacyjne i ich zastosowanie*, cz. 2, Warszawa: Wydawnictwa Komunikacji i Łączności.
- Kuźnicki J., Filipek A. (1997), *Różnorodność i wielofunkcyjność białek wiążących wapń (CaBP)*, „Kosmos. Problemy nauk biologicznych” 4(237), 603-608.
- Levinthal C. (1968), *Are There Pathways for Protein Folding?*, „Journal de Chimie Physique et de Physico-Chimie Biologique” 65, 44-45. <https://doi.org/10.1051/jcp/1968650044>
- Levinthal C. (1969), *How to Fold Graciously*, „University of Illinois Bulletin” 67(41), 22-24.
- Luscombe N. M., Greenbaum D., Gerstein M. (2001), *What is Bioinformatics? A Proposed Definition and Overview of the Field*, „Methods of Information in Medicine” 40, 346-358. <https://doi.org/10.1055/s-0038-1634431>
- Łuków P., Żekanowski C. (2005), *Pojęcie genu i genocentryczny paradygmat biologii*, „Przegląd Filozoficzny – Nowa Seria” 14(53), 85-105.
- Maynard Smith J. (2000), *The Concept of Information in Biology*, „Philosophy of Science” 67, 177-194. <https://doi.org/10.1086/392768>
- Mironov A. A., Fickett J. W., Gelfand M. S. (1999), *Frequent Alternative Splicing of Human Genes*, „Genome Research” 9, 1288-1293. <https://doi.org/10.1101/gr.9.12.1288>
- Moffat B. (2009), *A Reexamination of Biological Information from the Perspective of Practice*, <http://philsci—archive.pitt.edu/4803/>
- Moore C. (1996), *Recursion Theory on the Reals and Continuous-Time Computation*, „Theoretical Computer Science” 162, 23-44. [https://doi.org/10.1016/0304-3975\(95\)00248-0](https://doi.org/10.1016/0304-3975(95)00248-0)
- Mycka J. (2010), *Obliczenia dyskretne i ciągle jako realizacje antropomorficznej i fizycznej koncepcji efektywnej obliczalności* [w:] *Światy matematyki. Tworzenie czy odkrywanie*, I. Bondecka-Krzykowska, J. Pogonowski (red.), Poznań: Wydawnictwo UAM, 247-260.
- Mycka J. M., Piekarz M. (2004), *Przegląd zagadnień obliczalności analogowej* [w:] *Algorytmy, metody i programy naukowe*, S. Grzegórski, M. Miłosz, P. Muryjas (red.), Lublin: Polskie Towarzystwo Informatyczne, 125-132.

- Pauwels K., Van Molle I., Tommassen J., Van Gelder P. (2007), *Chaperoning Anfinsen: The Steric Foldases*, „Molecular Microbiology” 64(4), 917-922. <https://doi.org/10.1111/j.1365-2958.2007.05718.x>
- Pour-El M. B., Richards J. I. (1989), *Computability in Analysis and Physics*, Berlin: Springer. <https://doi.org/10.1007/978-3-662-21717-7>
- Roterman I., Konieczny L., Spólnik P. (2017), *Biologia systemów. Strategia działania organizmu żywego*, Warszawa: Wydawnictwo Naukowe PWN.
- Rozenberg G., Back T., Kok J. N. (2012), *Handbook of Natural Computing*, Berlin–Heidelberg: Springer. <https://doi.org/10.1007/978-3-540-92910-9>
- Rubel L. (1993), *The Extended Analog Computer*, „Advances in Applied Mathematics” 14, 39-50. <https://doi.org/10.1006/aama.1993.1003>
- Sarkar S. (1996), *Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology* [w:] *The Philosophy and History of Molecular Biology: New Perspectives*, S. Sarkar (ed.), Dordrecht: Kluwer. https://doi.org/10.1007/978-0-585-31233-0_11
- Schmucker D., Clemens J. C., Shu H., Worby C. A., Xiao J., Muda M., Dixon J. E., Zipursky L. (2000), *Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity*, „Cell” 101, 671-684. [https://doi.org/10.1016/S0092-8674\(00\)80878-8](https://doi.org/10.1016/S0092-8674(00)80878-8)
- Shagrir O. (2004), *Super-Tasks, Accelerating Turing Machines and Uncomputability*, „Theoretical Computer Science” 317, 105-114. <https://doi.org/10.1016/j.tcs.2003.12.007>
- Shannon C. (1941), *Mathematical Theory of the Differential Analyzer*, „Journal of Mathematics and Physics” 20, 337-354. <https://doi.org/10.1002/sapm1941201337>
- Siedliński R. (2016), *Turing Machines and Evolution. A Critique of Gregory Chaitin's Metabiology*, „Studies in Logic, Grammar and Rhetoric” 48(61), 133-150. <https://doi.org/10.1515/slgr-2016-0059>
- Siegelmann H. T. (1998), *Neural Networks and Analog Computation: Beyond the Turing Limit*, Boston: Springer. <https://doi.org/10.1007/978-1-4612-0707-8>
- Sohl J. L., Jaswal S. S., Agard D. A. (1998), *Unfolded Conformations of α -Lytic Protease are More Stable than its Native State*, „Nature” 395, 817-819. <https://doi.org/10.1038/27470>
- Stacewicz P. (2017a), *O redukcji informacji do danych* [w:] *Różne oblicza informacji*, P. Stacewicz (red.), Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 11-21.
- Stacewicz P. (2017b), *O różnych sposobach rozumienia analogowości w informatyce*, „Semina Scientiarum” 16, 121-137. <https://doi.org/10.15633/ss.2486>
- Stacewicz P. (2019), *Uncomputable Numbers and the Limits of Coding in Computer Science*, „Studia Semiotyczne” 30, 107-127.
- Stacewicz P. (2020), *Analogicity in Computer Science: Methodological Analysis*, „Studies in Logic, Grammar and Rhetoric” 63(76), w druku.
- Szathmáry E., Maynard Smith J. (1995), *The Major Evolutionary Transitions*, „Nature” 374, 227-232. <https://doi.org/10.1038/374227a0>
- Szcześniak M., Szweykowska-Kulińska Z. (2009), *Regulacja alternatywnego splicingu*, „Postępy Biologii Komórki” 36(1), 23-35.
- Turing A. M. (1936), *On Computable Numbers, with an Application to the Entscheidungsproblem*, „Proceedings of the London Mathematical Society” 42, 230-265. <https://doi.org/10.1112/plms/s2-42.1.230>
- Zwanzig R., Szabo A., Bagchi B. (1992), *Levinthal's Paradox*, „PNAS” 89, 20-22. <https://doi.org/10.1073/pnas.89.1.20>