Zadanie 1.

Ponieważ analiza i zrozumienie systemu przetwarzania informacji i ewolucji realnych organizmów jest trudne rozważmy uproszczony model komputerowy.

Jako DNA użyjemy dowolnych ciągów liter `A` i `B`. RNA używa identycznego alfabetu. Kod genetyczny jest następujący:

$$''AA'' \to F, \quad ''AB'' \to G, \quad ''BA'' \to H, \quad ''BB'' \to STOP \tag{1}$$

Aminokwasami `F, G, H` będą nastepujące proste funkcje jednej zmiennej:

$$F[x] = x + 1, \quad G[x] = 1, \quad H[x] = \frac{1}{x} \tag{2}$$

Odpowiednikiem procesy syntezy polipeptydu będzie składanie funkcji, np:

$$FGH \to F[G[H[x]]] \tag{3}$$

Po wstawieniu odpowiednich definicji funkcji F, G, H dostaniemy pewną funkcję jednej zmiennej, będącą wynikiem całego procesu syntezy „białka" i jego „zwinięcia".

Dla przykładu rozważmy DNA w postaci sekwencji `AABABBBA`. Składa się ona z kodonów: { AA, BA, BB, BA}, czyli po translacji {F, H, STOP, H}. Odrzucamy wszystko co znajduje za kodonem `STOP`, i otrzymujemy „strukturę pierwszorzędową białka" `FH`. Zgodnie z definicją `FH` zamieniamy na $F[H[x]]$, czyli

$$F[H[x]] = \frac{1}{x} + 1.$$

Aby zapoczątkować ewolucję i wzrost złożoności potrzebujemy trzech czynników:

1. populacji, czyli zbioru sekwencji DNA
2. mutacji, czyli procesu zmieniającego DNA w sposób losowy
3. rozmnażania

4. kryteriów przeżywalności

Potrzebne są minimum 3 typy mutacji:

1. substytucje, czyli zamiany litery DNA na inną
2. delecje, czyli skasowanie fragmentu DNA
3. insercje, czyli wydłużenie DNA poprzez wstawienie losowych liter, lub duplikacje pewnego fragmentu

Prawdopodobieństwo mutacji **musi** być proporcjonalne do długości DNA. W przeciwnym wypadku sekwencje (na ogół śmieciowe) będą rozrastały się w nieskończoność. Dla przykładu można przyjąć następujace wagi:

- kopiowanie bezbłędne: L_0^2/L
- substytucja L
- delecja $L/10$
- insercja $L/100$

gdzie L jest długością DNA, a L_0 długością, przy której prawdopodobieństwa substytucji i czystej duplikacji zrównują się.

Kluczowe jest kryterium na podstawie dokonamy selekcji. Dla przykładu będziemy próbować „wyhodować" najlesze możliwe przybliżenia wymierne liczb π, e oraz $\sqrt{2}$. Liczby te są niewymierne, więc ewolucja teoretycznie może wyprodukować dowolnie złożone sekwencje. Funkcją dopasowania niech będzie:

$$d(DNA) = \frac{1}{Abs[\pi - f(1)]} + \frac{1}{Abs[e - f(1)]} + \frac{1}{Abs[\sqrt{2} - f(1)]},$$

gdzie $f(1)$ jest wartością „zwiniętej" funkcji $f(x)$ obliczonej dla $x = 1$. Fakt, że $f(1)$ jest dowolną liczbą wymierną nie jest całkiem oczywisty, ale wytworzenie tej „wiedzy" zostawiamy procesowi ewolucyjnemu.

Na początek wystarczy użyć populacji około 200 4-literowych sekwencji DNA. W każdym pokoleniu należy:

1. dokonać translacji każdego DNA
2. obliczyć funkcję d dla każdego z nich
3. posortować wartości d od największej do najmniejszej
4. skasować „najsłabszych”, np. połowę
5. pozostałych zduplikować i poddać mutacjom.

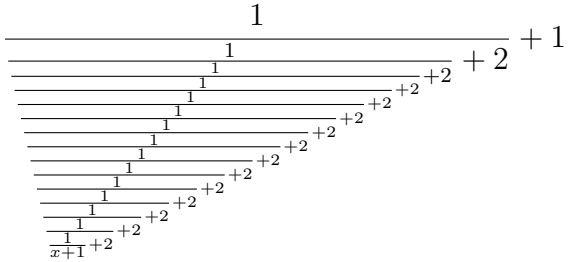
Proces należy powtarzać wielokrotnie (przynajmniej 100-1000 razy). Przechowywać należy tylko i wyłącznie bieżące DNA. Pozostałe elementy (sekwencje „aminokwasowe”, wytworzone funkcje) należy traktować jako elementy tymczasowe.

Przebieg eksperymentu jest losowy. U mnie wyglądało to tak:

- w pierwszych pokoleniach dominowały liczby 3 i 2 o prostych sekwencjach `AAAA, AAAAAA`
- na dalszym etapie pojawiły się proste liczby wymierne które z grubsza pasowały do jednej lub kilku z liczb $\pi, e\sqrt{2}$ jak $\frac{7}{5}$ o różnych sekwencjach jak `ABAABAAAAABAAA` czy `BABAAABAAAAABAAAB`
- potem DNA wyspecjalizowało się w przybliżaniu $\sqrt{2}$, pojawiła się np. liczba $\frac{17}{12} \simeq 1.41667$, `AABAAAAABAAAAABAAA`
- po kilkuset tysiącach pokoleń, DNA „wymyśliło” ułamki łańcuchowe, od tego momentu ewolucja polegała już tylko na wydłużaniu tego poprawnego, ostatni zanotowany wynik to:

```
BABAABABABABABABABAABAABABABABABABAAABABABABAABABABABABAABABABABAAAAABA
ABABABABABABABABAAAABAABABABABAAABABABABAABABABAAABABABABAABAAAAABAABABAAAB
ABAABAABABABAAAABAAAABAABAAAAABAAAABABAABAAAABAABAAAABABAABAABAAABAABAAA
BBABAAABAABABABBBABBAAABABAABA
```

odpowiadający funkcji:



Po wstawieniu $x = 1$ dostajemy $\frac{665857}{470832}$, co zgadza się z $\sqrt{2}$ z dokładnością do 12 miejsc po przecinku. Fakt, że powyższy (ale nieskończony) ułamek łańcuchowy jest rozwinięciem liczby $\sqrt{2}$ jest dobrze znany, proszę np. wpisać `sqrt(2)` w WolframAlpha.

Aby sprecyzować cel zadania, proszę spróbować „wyhodować" własne najlepsze przybliżenie np. $\sqrt{2}$ czy π i porównać między sobą podobieństwo (dopasowanie) uzyskanych sekwencji DNA. Zastanowić się, jak można użyć opisanego systemu do przybliżonego rozwiązywania równań algebraicznych lub jeszcze trudniejszych zadań jak np. całkowanie. Wyszukać standardowe motywy pojawiające się w DNA i „białkach", np. ułamek łańcuchowy wymaga naprzemiennego dodawania pewnej liczby (funkcji F) oraz odwracania (funkcji H), a wytworzenie dużej liczby całkowitej wielokrotnego powtarzania się F. Czy są podsekwencje „konserwatyne" czyli niezmienne w toku ewolucji? Na jakim etapie się pojawiły i czy mogą zaniknąć? Zadanie ma charakter „otwarty", więc podobne pytania można mnożyć, jak również wyszukiwać podobieństwa i różnice w stosunku do procesów biologicznych.