

Metody bioinformatyki

Biokomputery - model splatania

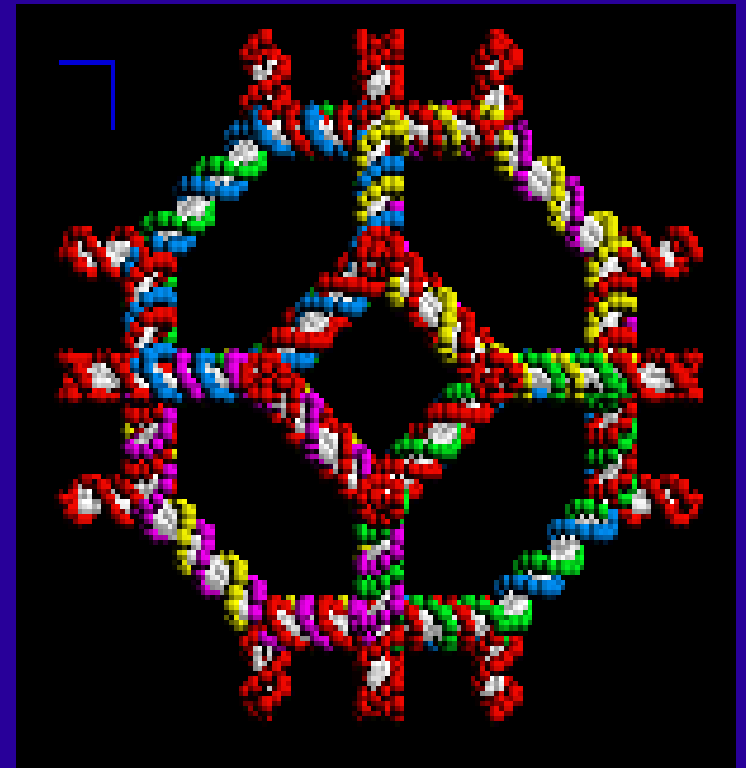
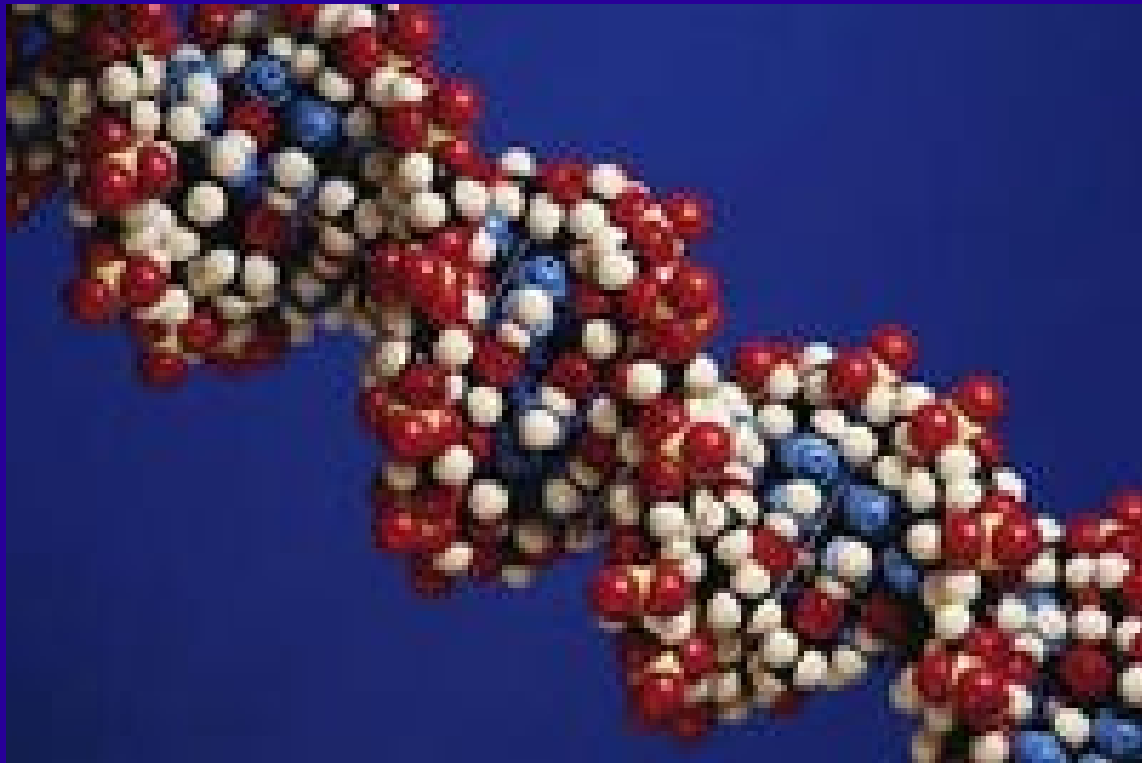
prof. dr hab. Jan Mulawka



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY





Podział biokomputerów

1. **autonomiczne**

Nie wymagają udziału laboranta w trakcie obliczeń.

Obliczenia w trzech krokach: przygotowanie molekuł wejściowych, właściwa reakcja, odczytanie wyniku.

Udział człowieka jest niezbędny tylko w pierwszym i ostatnim kroku, przetwarzanie danych w probówce odbywa się samoczynnie.

2. **nieautonomiczne**

Modele biokomputerów

1. model filtrowania

2. model splatania

Wykorzystuje naprzemienne działanie enzymu ligazy oraz enzymu restrykcyjnego.

3. model oparty na samoskładaniu

4. model oparty na membranach

Kodowanie informacji

Zdania logiczne w koniunkcyjnej postaci normalnej

$$A = B_1 \text{ and } B_2 \text{ and } \dots \text{ and } B_n$$

B_i - klauzula, jest alternatywą dowolnej liczby literałów

$$B_i = b_1 \text{ or } b_2 \text{ or } \dots \text{ or } b_m$$

Literał – pojedyncza zmienna zdaniowa lub negacja

Literały

Każdą z klauzul Bi kodujemy za pomocą pojedynczej molekuly

Literał reprezentujemy unikalną sekwencją czterech nukleotydów tak, że literały reprezentujące zmienną atomową i jej negację są do siebie komplementarne

p → 3' -AAAC- 5' to NOT **p** → 5' -TTTG- 3'

Odrzucamy czwórki komplementarne same ze sobą

3' -AATT- 5'

NOT **p** oznaczmy **p**

Klauzula faktu

Wszystkie molekuly faktów mają taką długość 50 nukleotydów i tę samą sekwencję początkową, różnią się lepkiem końcem reprezentującym literał.

5'-(50)CCGGCCGGGCAGCGCTG -3'

3'-(50)GGCCGGGCCGTCGCGAC(. . .) -5'

3' -GGCCGGGC -5' stała sekwencja

3' -CGTCG -5' sekwencja rozpoznawalna BseXI

Reprezentacja klauzul faktów **p** oraz **p**

Klauzula faktu **p**

5' -(50)CCGGCCGGGCAGCGCTG -3'

3' -(50)GGCCGGGCCGTCGCGAC(**p**) -5'

Klauzula faktu NOT **p**

5' -(50)CCGGCCGGGCAGCGCTG -3'

3' -(50)GGCCGGGCCGTCGCGAC(**p**) -5'

Klauzula zawierająca dwa literały

Schemat zawiera:

1. dwa lepkie końce, zawierające informacje o składowych literałach alternatywy

2. stałą sekwencję

3' -GTCGCGAC- 5'

rozdzielającą łańcuchy reprezentujące literały

Schemat dla dwuliterałowej klauzuli

5' - (. . .) CAGCGCTG - 3'
3' - GTCGCGAC (. . .) - 5'

Molekuła reprezentująca klauzulę (**p v q**)

5' - (**p**) CAGCGCTG
3' - GTCGCGAC (**q**)

Schemat dla dwuliterałowej klauzuli

5' - (. . .) CAGCGCTG - 3'
3' - GTCGCGAC (. . .) - 5'

Molekuła reprezentująca klauzulę (**p** v **q**)

5' - (**p**) CAGCGCTG
3' - GTCGCGAC (**q**)

Klauzule o większej liczbie literałów

(p1 v p2 v p3)

Zestawiamy obok siebie molekuly reprezentujace

(p1 v p2) oraz (p2 v p3)

5 - (p1)CAGCGCTG (p2)CAGCGCTG -3'
3'- GTCGCGAC(p2) GTCGCGAC(p3)-5'

Zlepiamy je dopisujac brakujace fragmenty

5 - (p1)CAGCGCTG (p2) (p2)CAGCGCTG -3'
3'- GTCGCGAC(p2) (p2) GTCGCGAC(p3)-5'

Molekuła terminalna i wnioskowanie

Aby wnioskowanie przebiegało poprawnie, oprócz odpowiednio zakodowanych klauzul, niezbędne jest umieszczenie w próbówce także molekuły terminalnej, która uczestniczy w każdej reakcji wnioskowania

5' - (50) CCGGCCCGG - 3'

3' - (50) GGCCGGCCGTCG - 5'

Wnioskowanie - reguła odrywania

Niech będą molekuly dla klauzul p



oraz $(p \vee q)$



Reakcja wnioskowania

Ligacja molekuł dla klauzul p oraz $(p \vee q)$

(50)CCGGCCGGCAGCGCTG(p)CAGCGCTG

(50)GGCCGGCCGTCGCGAC(p)GTCGCGAC(q)

Rozcięcie enzymem BseXI



oraz wniosek częściowy



Wynik wnioskowania

Ligacja molekulej terminalnej i wniosku częściowego daje:

5' - (50)CCGGCCGGGCAGCGCTG - 3'

3' - (50)GGCCGGGCCGTCGCGAC (**q**) - 5'

czyli klauzulę faktu **q**

Takie cząsteczki mogą brać udział w dalszych reakcjach (bo ma lepki koniec).

Zapytania

Molekuła zapytania składa się z trzech elementów:

1. stałej sekwencji 4 nukleotydów w formie lepkiego końca, reprezentujących literał zanegowany względem tego, o który pytamy
2. unikalnej czwórki nukleotydów komplementarnych z samym sobą
3. łańcucha dowolnych nukleotydów o unikalnej długości identyfikującej dane zapytanie.

Schemat molekuly reprezentujacej zapytanie

5' - () **NNNN** (...) - 3'

3' - **NNNN** (...) - 5'

Zapytanie o **p**

5' - (**p**) **AATT** (250) - 3'

3' - **TTAA** (250) - 5'

Zapytanie o **NOT q**

5' - (**q**) **ATAT** (300) - 3'

3' - **TATA** (300) - 5'

Reakcja zapytania

Niech w probówce będą molekuly reprezentujące klauzulę **p** oraz zapytanie o **p**?

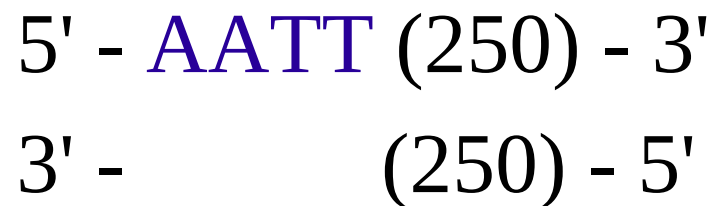
Ligacja tych dwóch molekuł daje:

5'-(50)CCGGCCGG**GCAGC**GCTG(**p**)**AATT**(250)-3'
3'-(50)GGCCGGGC**CGTCG**CGAC(**p**)**TTAA**(250)-5'

Rozcięcie enzymem BseXI



oraz



Wynik zapytania

Ligacja powstałych tych samych molekuł ze sobą

5' – (250)AATT(250) – 3'

3' – (250)TTAA(250) – 5'

Odpowiedzi rozróżniamy po długości powstałych molekuł za pomocą elektroforezy w żelu.

Otrzymaliśmy molekułę o długości

504 nukleotydów.

Oznacza to pozytywną odpowiedź na zapytanie **p?**

Wykrywanie sprzeczności

W próbówce występują molekuly literałów **p** i **p**
dodatkowo jest molekula terminalna wnioskowania.

Ligacja molekuł **p** i **p** daje:

5'-(50)CCGGCCGG**G**C**A**G**C**GCTG (**p**) . . .

3'-(50)GGCCGGG**C**G**T****C**GCGAC (**p**)

. . . CAGC**G**CT**G**CCGGCCGG(50)-3'

GTCG**C**G**A**C**G**GCCGGCC(50)-5'

Rozcięcie enzymem BseXI

5'-(50)CCGGCCGGGCAGCGCTG (p) -3'
3'-(50)GGCCGGGCCGTCGCGAC (p) GTCG -5'

oraz

5'-CAGCGCTGCCGGCCGG(50)-3'
3'- CGACGGCCGGCC(50)-5'

Ligacja powstałej cząsteczki i molekuły terminalnej

5'-(50)CCGGCCGGGCAGCGCTGCGGGCCGG(50)-3'
3'-(50)GGCCGGCCGTCGCGACGGCCGGCC(50)-5'

Rozcięcie enzymem BseXI

5'-(50)CCGGCCGGGCAGCGCTGCCGG -3'
3'-(50)GGCCGGGCCGTCGCGACGGCCGGGCC-5'

oraz

5'-CCGG(50) -3'
3'- (50)-5'

Ligacja powstałych molekuł ze sobą

5'-(50) CCGG (50)-3'

3'-(50) GGCC (50)-5'

Cząsteczka niezawierająca lepkich końców o unikalnej długości 104 nukleotydów oznacza wystąpienie sprzeczności w systemie.

Przykład złożonego wnioskowania

Niech będą dane aksjomaty początkowe:

$a, d, u, (((a \& b) \vee d) \rightarrow c), ((c \& u) \rightarrow x)$

oraz zaytanie **x** ?

Po sprowadzeniu do koniunkcyjnej postaci normalnej dostajemy aksjomaty:

$(a \& d \& u \& (\underline{a} \vee \underline{b} \vee c) \& (\underline{d} \vee c) \& (\underline{c} \vee \underline{u} \vee x))$

Kodowanie klauzul

fakt **a**

5'-(50)CCGGCCGG**GCAGC**GCTG -3'

3'-(50)GGCCGGG**CGTCG**CGAC(**a**)-5'

fakt **d**

5'-(50)CCGGCCGG**GCAGC**GCTG -3'

3'-(50)GGCCGGG**CGTCG**CGAC(**d**)-5'

fakt **u**

5'-(50)CCGGCCGG**GCAGC**GCTG -3'

3'-(50)GGCCGGG**CGTCG**CGAC(**u**)-5'

Kodowanie klauzul

klauzula (a v b v c)

5'-(a)CAGCGCTG(**b**) (b)CAGCGCTG -3'

3'- GTCGCGAC(b) (**b**)GTCGCGAC(**c**)-5'

klauzula (d v c)

5'- (d)CAGCGCTG -3'

3'- GTCGCGAC(**c**)-5'

klauzula (c v u v x)

5'-(c)CAGCGCTG(**u**) (u)CAGCGCTG -3'

3'- GTCGCGAC(u) (**u**)GTCGCGAC(**x**)-5'

Kodowanie molekuł

Molekuła terminalna wnioskowania

5' - (50) CCGGCCCGG - 3'

3' - (50) GGCCGGCCGTCG - 5'

Molekuła reprezentująca zapytanie **x**?

5' - (x) AATT (250) - 3'

3' - TTAA(250) - 5'

Ligacja klauzul **a** oraz (**a** v **b** v **c**)

5' – (50) CCGGCCCG**G**C**A**G**C**GCTG (**a**) CAGCG
3' – (50) GGCCGGC**C**G**T**C**G**CGAC (**a**) GTCGC

... CTG (**b**)(**b**) CAGCGCTG - 3'
GAC (**b**)(**b**) GTCGCGAC (**c**) -5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCCG**G**C**A**G**C**GCTG (**a**) -3'
3' – (50) GGCCGGC**C**G**T**C**G**CGAC (**a**) GTCG-5'

5'-CAGCGCTG (**b**)(**b**) CAGCGCTG - 3'
3'- CGAC (**b**)(**b**) GTCGCGAC (**c**) -5'

Ligacja molekuly terminalnej i wniosku częściowego

5' – (50) CCGGCCG **G C A G C** GCTG

3' – (50) GGCCGGC **C G T C G** CGAC

... (**b**)(**b**) CAGCGCTG - 3'

(**b**)(**b**) GTCGCGAC (**c**) -5'

Ligacja molekuł reprezentujących klauzule **d** oraz (**d** v **c**)

5' – (50) CCGGCCCG**G C A G C**GCTG

3' – (50) GGCCGGC**C G T C G**CGAC

... (**d**) CAGCGCTG - 3'

(**d**) GTCGCGAC (**c**) -5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCCG**G**C**A**G**C**GCTG (**b**) -3'

3' – (50) GGCCGGC**C**G**T**C**G**CGAC (**b**)(**b**) -5'

5'-(**b**) CAGCGCTG - 3'

3'- GTCGCGAC (**c**) -5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCCG**G**C**A**G**C**GCTG (**b**) -3'

3' – (50) GGCCGGC**C**G**T**C**G**CGAC (**b**)(**b**) -5'

5'-(**b**) CAGCGCTG - 3'

3'- GTCGCGAC (**c**) -5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCCGCAGCGCTG (d) -3'
3' – (50) GGCCGGCCGTCGCGAC (**d**)GTCG -5'

5'- CAGCGCTG - 3'
3'- GTCGCGAC (**c**) -5'

Ligacja molekuly terminalnej i wniosku częściowego

5' – (50) CCGGCCG **G C A G C** GCTG -3'
3' – (50) GGCCGGC **C G T C G** CGAC (c) -5'

powstaje molekula reprezentujaca
klauzule faktu (c)

Ligacja klauzul **c** oraz (**c** **v** **u** **v** **x**)

5' – (50) CCGGCCCG**G****C****A****G****C**GCTG (**c**) CAGCG
3' – (50) GGCCGGC**C****G****T****C****G**CGAC (**c**) GTCGC

... CTG (**u**)(**u**) CAGCGCTG - 3'
GAC (**u**)(**u**) GTCGCGAC (**x**) -5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCCG **G C A G C** GCTG (c) -3'
3' – (50) GGCCGGC **C G T C G** CGAC (**c**) GTCG-5'

5'-CAGCGCTG (**u**)(u) CAGCGCTG - 3'
3'- CGAC (u)(**u**) GTCGCGAC (**x**) -5'

Ligacja molekuly terminalnej i wniosku częściowego

5' – (50) CCGGCCCGGCAGCGCTG

3' – (50) GGCCGGCCGTCGCGAC

... (u)(u) CAGCGCTG - 3'

(u)(u) GTCGCGAC (x) -5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCCGCAGCGCTG (**u**) -3'

3' – (50) GGCCGGCCGTCGCGAC (**u**)(u) -5'

5'-(**u**) CAGCGCTG - 3'

3'- GTCGCGAC (**x**) -5'

powstaje molekula reprezentujaca klauzule (**u v x**)

Ligacja klauzul u oraz (u v x)

5' – (50) CCGGCCCGGCAGCGCTG (u)

3' – (50) GGCCGGCCGTCGCGAC (u)

... CAGCGCTG - 3'

GTCGCGAC (x) -5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCG **G C A G C** GCTG (u) -
3'

3' – (50) GGCCGGC **C G T C G** CGAC (u)GTTCG -5'

5'- CAGCGCTG - 3'

3'- CGAC (**x**) -5'

Ligacja molekuly terminalnej i wniosku częściowego

5' – (50) CCGGCCCGGCAGCGCTG –3'
3' – (50) GGCCGGCCGTCGCGAC (**x**) –5'

powstaje molekula reprezentujaca
klauzule faktu (**x**)

Ligacja molekuly reprezentujacej **x**
oraz molekuly zapytania **x**?

5'-(50)CCGGCCGG**G C A G C**GCTG(**x**)AATT(250)-3'
3'-(50)GGCCGGC**C G T C G**CGAC(**x**)TTAA(250)-5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCG **G C A G C** GCTG (x) -3'

3' – (50) GGCCGGC **C G T C G** CGAC (**x**)TTAA -5'

5'- AATT(250) - 3'

3'- (250) -5'

Ligacja powstałych molekuł ze sobą

5'- (250)AATT(250) - 3'

3'- (250)TTAA(250) - 5'

Otrzymaliśmy molekułę o długości 504 nukleotydów, która oznacza odpowiedź twierdzącą na zapytanie **x**?

Więc zdanie **x jest możliwe do wyprowadzenia.**

Przechodniość implikacji

$$((p \rightarrow q) \& (q \rightarrow r)) \rightarrow (p \rightarrow r)$$

lub rozpisując implikacje

$$((p \vee q) \& (q \vee r)) \rightarrow (p \vee r)$$

Przykład wnioskowania

fakt **p**

5'-(50)CCGGCCGG**GCAGC**GCTG -3'

3'-(50)GGCCGGC**CGTCG**CGAC(**p**)-5'

fakt (**p v r**) po uwzględnieniu przechodności

5'- (**p**)CAGCGCTG(**q**)CAGCGCTG -3'

3'- GTCGCGAC(**q**)GTCGCGAC(**r**)-5'

molekuła terminalna wnioskowania

5'-(50)CCGGCCGG -3'

3'-(50)GGCCGGC**CGTCG**-5'

Ligacja klauzul p oraz (p v r)

5' – (50) CCGGCCCGGCAGCGCTG (p)

3' – (50) GGCCGGCCGTCGCGAC (p)

... CAGCGCTG (q)CAGCGCTG - 3'
GTCGCGAC (q) GTCGCGAC(r)-5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCG **G C A G C** GCTG (**p**) -
3'

3' – (50) GGCCGGC **C G T C G** CGAC (**p**)GTTCG -5'

5'- CAGCGCTG (**q**)CAGCGCTG - 3'

3'- CGAC (**x**)GTTCGCGAC (**r**) -5'

Ligacja molekuly terminalnej i powstałej cząsteczki

5'-(50)CCGGCCGGGCAGCGCTG(**q**)CAGCGCTG -3'
3'-(50)GGCCGGGCCGTCGCGAC(**q**)GTCGCGAC(**r**)-5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCG **G C A G C** GCTG (**q**) -
3'

3' – (50) GGCCGGC **C G T C G** CGAC (**q**)GTTCG -5'

5'- CAGCGCTG - 3'

3'- CGAC (**r**) -5'

Wynik wnioskowania

Ligacja molekuly terminalnej i powstałej cząsteczki daje:



Powstała cząsteczka reprezentuje klauzulę faktu **r**

Rozwiązywanie problemu SAT

Boolean Satisfiability Problem (SAT)

Problem **spełnialności** – czy dla danego zdania złożonego istnieje takie wartościowanie zmiennych (0 lub 1), aby całe zdanie było prawdziwe.

Przykładowo formuła $(a \ \& \ (\underline{a} \ \vee \ \underline{b}))$ - spełnialna
dla $w(a) = 1$ oraz $w(b) = 0$

$(a \ \& \ (\underline{a} \ \vee \ \underline{b}) \ \& \ b)$ - niespełnialna

Rozwiązywanie problemu SAT

- 1) równoległe wygenerowanie wszystkich możliwych przypisań wartości zmiennych
- 2) poszukiwanie formuł niespełnialnych, poprzez upraszczanie danego zdania złożonego.

Jeśli ostatnie zdanie powstałe podczas reakcji nie jest spełnialne, to wejściowe zdanie złożone jest również niespełnialne.

W przeciwnym razie system nie jest w stanie określić, czy formuła jest spełnialna.

Jest to tzw. częściowe rozwiązanie problemu **SAT**.

Badanie spełnialności formuły (**a** & (**a** v **b**) & **b**)

fakt **a**

5'-(50)CCGGCCGG**GCAGC**GCTG -3'
3'-(50)GGCCGGC**CGTCG**CGAC(**a**)-5'

fakt **b**

5'-(50)CCGGCCGG**GCAGC**GCTG -3'
3'-(50)GGCCGGC**CGTCG**CGAC(**b**)-5'

molekuła terminalna wnioskowania

5'-(50)CCGGCCGG -3'
3'-(50)GGCCGGC**CGTCG**-5'

Reakcje dla badania spełnialności

Molekuła reprezentująca klauzulę (a v b)

5' - (a) CAGCGCTG - 3'

3' - GTCGCGAC (b) - 5'

Ligacja molekuł reprezentujących a oraz (a v b)

5'-(50)CCGGCCGGGCAGCGCTG(a)CAGCGCTG -3'

3'-(50)GGCCGGCCGTCGCGAC(a)GTCGCGAC(b)-5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCG **G C A G C** GCTG (**a**) -
3'

3' – (50) GGCCGGC **C G T C G** CGAC (**a**)GTTCG -5'

5'- CAGCGCTG - 3'

3'- CGAC (**b**) -5'

Ligacja molekuly terminalnej i wniosku częściowego

5' – (50) CCGGCCG **G C A G C** GCTG -3'
3' – (50) GGCCGGC **C G T C G** CGAC (**b**) -5'

powstaje molekula reprezentujaca
klauzule faktu (**b**)

Reakcje dla badania spełnialności

Ligacja molekuł reprezentujących klauzule **b** i **b**
(powstałej w poprzednim kroku)

5'-(50)CCGGCCGG**G**C**A**G**C**GCTG (**b**) ...

3'-(50)GGCCGGGC**C**G**T**C**G**CGAC (**b**)

... CAGC**G**C**T**G**C**CGGGCCGG(50)-3'

GTCG**C**G**A**C**G**GCCGGGCC(50)-5'

Rozcięcie molekuly enzymem BseXI

5'-(50)CCGGCCGG**G C A G C**GCTG (**b**) -3'
3'-(50)GGCCGGC**C G T C G**CGAC (**b**)GTCG-5'

5'- CAGC**GCTGC**CCGGCCGG(50)-3'
3'- **CGACG**GGCCGGCC(50)-5'

Ligacja molekuly terminalnej i wniosku częściowego

5' – (50) CCGGCCG **G C A G C G C T G C** ...

3' – (50) GGCCGGC **C G T C G C G A C G**

... CGGCCGG (50) - 3'

GCCGGCC (50) - 5'

Rozcięcie molekuly enzymem BseXI

5'-(50)CCGGCCGGGCAGCGCTGCCGG -3'
3'-(50)GGCCGGCCGTCGCGACGGCCGGCC-5'

5' - CCGG (50) - 3'

3' - (50) - 5'

Ligacja powstałych molekuł ze sobą

5' – (50) CCGG (50) – 3'

3' – (50) GGCC (50) – 5'

Powstała cząsteczka o unikalnej długości 104 nukleotydów, **oznaczająca wystąpienie sprzeczności w systemie.**

Oznacza to, że analizowana formuła nie jest spełnialna dla żadnych wartości literałów.

Badanie spełnialności formuły (**a** & (**a** v **b**))

molekuła reprezentująca fakt **a**

5'-(50)CCGGCCGG**GCAGC**GCTG -3'

3'-(50)GGCCGGC**CGTCG**CGAC(**a**)-5'

molekuła reprezentująca klauzulę (**a** v **b**)

5' - (**a**) CAGCGCTG - 3'

3' - GTCGCGAC (**b**) - 5'

molekuła terminalna wnioskowania

5'-(50)CCGGCCGG -3'

3'-(50)GGCCGGC**CGTCG**-5'

Ligacja molekuł reprezentujących **a** oraz (**a** v **b**)

5' – (50) CCGGCCCGGCAGCGCTG (**a**) ...

3' – (50) GGCCGGCCGTCGCGAC (**a**)

... CAGCGCTG - 3'
GTCGCGAC (**b**) - 5'

Rozcięcie molekuly enzymem BseXI

5' – (50) CCGGCCG **G C A G C** GCTG (**a**) -
3'

3' – (50) GGCCGGC **C G T C G** CGAC (**a**) GTCG -5'

5'- CAGCGCTG - 3'

3'- CGAC (**b**) -5'

Ligacja molekulej terminalnej i wniosku częściowego

5'-(50)CCGGCCGGGCAGCGCTG -3'
3'-(50)GGCCGGCCGTCGCGAC(b)-5'

powstała molekulej reprezentuje fakt b,
nie uzyskano sprzeczności

Oznacza to, że system nie jest w stanie określić czy dana formuła jest spełnialna.

Podobną koncepcję można wykorzystać do sprawdzania czy dana formuła jest tautologią.