

Modelagem de Sistemas de Gerenciamento de *Workflows* em Redes de Petri para Sequenciamento de DNA

Geise Kelly da Silva Santos, Liliane do Nascimento Vale

¹Departamento de Computação – Universidade Federal de Goiás (UFG)
Campus Catalão (CaC) – Catalão – GO – Brazil

Abstract. *This paper presents the utilization for process of workflow applied in DNA sequencing, and modelled using Petri nets. Firstly, do are shows the concepts of Workflow and Petri Net, their functions too. The biochemistry activities are mentioned, then DNA computing is specified. Therefore, it is instituted an algorithm for to represent the process of sequencing congruous the settings of Petri net. Finally the model is validates through utilizing the software Process Maker, examining the main results obtained.*

Resumo. *Este artigo apresenta a utilização de processos de workflow aplicados no sequenciamento de DNA, e modelado usando Rede de Petri. Inicialmente, são expostos os conceitos de Workflow e Rede de Petri, assim como suas funcionalidades. Também são descritas as atividades bioquímicas do DNA, e então é detalhada a computação do DNA. Dessa forma, é estabelecido um algoritmo para representar o processo de sequenciamento de acordo com as configurações da Rede de Petri. Finalmente, o modelo é validado usando o software Process Maker, discutindo-se os principais resultados obtidos.*

1. Introdução

Bioinformática é a ciência aplicada à utilização de ferramentas e técnicas computacionais e fundamentos matemáticos para auxiliar na realização de atividades comuns à Biologia Molecular. Seus principais objetivos se resumem em criar, melhorar, desenvolver e manipular banco de dados para a Biologia Molecular, incluindo ferramentas computacionais para coletar, organizar, validar e analisar dados experimentais.

Um dos pontos imprescindíveis da Bioinformática é a representação do sequenciamento de DNA para detecção de inconsistências e anomalias, em que a montagem de fragmentos de DNA é um problema computacional expressivo nessa área [Meidanis 1996].

No contexto abordado, o DNA é uma fita composta por fragmentos e para o sequenciamento deste, atividades simultâneas e sequenciais devem ocorrer. Dessa forma, temos que as características assumidas pela molécula abrangem também propriedades inerentes à processos de *workflow*. Já que estes são fluxos de atividades que definem as tarefas a serem realizadas e a ordem em que ocorrem [Lemos 2004].

O mapeamento da sequência de DNA envolve alto custo computacional, medidas alternativas de análise da fita de DNA devem ser discutidas, quando por exemplo, que esta é submetida a representação por meio de processos de *workflow* usando Redes de Petri. Isto favoreceu, por exemplo, no acompanhamento visual do sequenciamento e a detecção de inconsistências que produzem resultados ambíguos quando empregadas técnicas de sequenciamento pouco eficientes.

Dessa forma, é preciso descrever uma modelagem específica para a representação do sequenciamento de DNA e a rota assumida para tal processo. O uso de Redes de Petri para modelagem e análise de sistemas de *Workflow* é amplamente discutido. Pois, desde as características formais que a Rede de Petri apresenta e aspectos dinâmicos e contínuos que a mesma assume, oferece condições satisfatórias para análise formal de processos de negócio aplicados ao sequenciamento de DNA.

Neste artigo, serão apresentados os conceitos de *Workflow* e Redes de Petri, assim como a importância de cada um para a Bioinformática, enfatizando os motivos da utilização destes. Logo após, será abordado o uso de *Workflow* para o sequenciamento de DNA usando a Rede de Petri. A validação é obtida usando o *software Process Maker*.

2. Trabalhos Relacionados

Em [Schuster 2008], é descrito aspectos inerentes à computação do DNA, no que tange ao processamento bioquímico para o sequenciamento do mesmo efetuados através da modelagem da Rede de Petri. Neste é apresentado um algoritmo para processos de DNA seguidos de importantes considerações, entretanto em [Schuster 2008] não é manipulado em específico o processo de sequenciamento de DNA e sim outros processos que comumente existem são apresentados.

[Lemos 2004] descreve o modelo de *workflow* para processos da Bioinformática. Neste trabalho foram apontados as atividades para o sequenciamento de DNA e as modelagens específicas para cada processo da Bioinformática. Contudo, a representação e a ordem de execução de tais atividades não foram questionados.

No trabalho de [Meidanis 1996] sucedeu-se ampla discussão sobre a montagem dos fragmentos de DNA usando processos de *workflow*, apresentando os desafios e as estratégias atualmente utilizadas. Através deste seguimento, foi proposto a utilização da arquitetura WASA para o processo.

Em [Barjis 1999] foi discutido sobre a formalização da produção de proteínas utilizando a Rede de Petri, no qual foi construído um modelo de Rede de Petri baseado nos processos da produção de proteína e após, foi expresso módulos de programa em pseudocódigo com esta formalização. Porém, neste trabalho não é utilizado fluxos de trabalho para o seu desenvolvimento, o que facilitaria a validação e a formalização.

A tese de [Digiampietri 2007] referiu-se ao emprego de *workflows* científicos em experimentos de Bioinformática, construindo, implementando e validando uma infraestrutura computacional para oferecer suporte a tais atividades.

No trabalho de [Barjis 2006] foi mostrado o processo de produção de proteína a partir do DNA, o qual foi modelado e analisado na Rede de Petri. Além disso, são retratados os eventos moleculares e reações químicas que ocorrem durante a execução.

3. Sistemas de *Workflows*

Workflows, também conhecidos como processos, definidos pela Coligação de Gerenciamento de *Workflow* (*Workflow Management Coalition* - WfMC) [WfMC] como a automação do processo de negócio, identificando as atividades do processo, regras de procedimento e controle de dados.

Documentos, informações ou tarefas são passados entre os participantes de acordo com um conjunto pré definido de regras, para se alcançar ou contribuir para um objetivo global de um negócio [Lemos 2004]. Apesar de um *workflow* poder ser manualmente organizado, na prática a maioria dos *workflows* são organizados dentro de um contexto de um sistema de informação para prover um apoio automatizado aos procedimentos [Lemos 2004].

Um sistema de *workflow* baseia-se na gerência do fluxo de atividades que existem por exemplo, em organizações. Devido a representação de rotas (paralelas, sequencias, iterativas e alternativas) e ao controle de recursos, a mesma pode ser aplicada para a análise dos fragmentos de DNA para detecção de inconsistências e anomalias que ocorrem durante o sequenciamento do mesmo.

4. Redes de Petri

O modelo da Rede de Petri foi proposto em 1962 por Carl Adam Petri em sua tese de doutorado para modelar a comunicação entre máquinas de estado. A Rede de Petri é uma ferramenta gráfica e matemática que se adapta bem a um grande número de aplicações em que noções de eventos e de evoluções simultâneas são importantes [Valette 1997]. A configuração da Rede de Petri é determinada da seguinte forma:

- Possui lugares (P), transições (t) e marcas;
- Um lugar nunca é diretamente conectado com outro, sempre há intermédio de transições, ocorrendo da mesma forma com transições;
- Um lugar está "ativo" quando possui uma ficha e uma transição dispara quando algum lugar anterior está "ativo", então este perde a ficha e o próximo lugar recebe a ficha;
- Este "acionamento" de transições é determinante no comportamento da Rede.

A Figura 1 ilustra o exemplo de Rede de Petri para o problema de sequenciamento de DNA proposto a seguir.

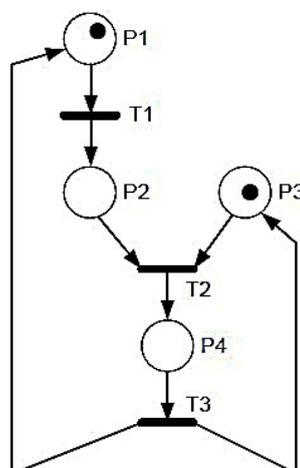


Figura 1. Exemplo de Rede de Petri - Sequenciamento de DNA

Podemos classificar as Redes de Petri (RdP) em: RdP de baixo nível (ordinária) e RdP de alto nível (Colorida, Objeto, etc.), tendo ambas a mesma configuração primária citada. As redes ordinárias se caracterizam pelo tipo de suas fichas que não carregam informações complexas, indicando basicamente o estado da rede. Enquanto que as de alto nível possuem marcas de tipos particulares, com valores e estados distintos podendo até mesmo representar verdadeiras estruturas de dados.

A modelagem de sistemas de *workflow* é significativa, pois facilita o desenvolvimento das etapas do projeto e sua posterior análise e/ou correção [Vale 2010]. Assim, a Rede de Petri torna-se uma linguagem totalmente viável, já que oferece representação gráfica, fácil compreensão, formalismo matemático e representação de aspectos comportamentais e estruturais sobre o sistema.

Atualmente, as Redes de Petri geram importantes modelos para diversos sistemas, pois possuem elementos competentes para descrever partes de um sistema como controle, conflitos, compartilhamento, entre outros. Rede de Petri é uma técnica de modelagem que permite a representação de sistemas, utilizando como alicerce uma forte base matemática [Maciel P.R.M.; Lins 1996].

Sua utilização para moldar o processo de sequenciamento de DNA projetado em *Workflows* deve-se ao fato de possuir semântica formal além da sua natureza gráfica, elas podem representar explicitamente os estados do sistema, além da capacidade de execução e simulação das atividades.

5. Sequenciamento de DNA

O DNA é formado por duas fitas helicoidais antiparalelas (com mesmo sentido mas em direções opostas) com filamentos paralelos ligados por pontes de hidrogênio. O DNA é composto por pequenas moléculas chamadas nucleotídeos, sendo estes a principal fonte de informação genética. Os nucleotídeos constituem-se de uma base, um açúcar e um fosfato, em que são diferenciados através das quatro bases existentes: adenina (A), timina (T), guanina (G) e citosina (C) [Timmer 2009].

Os nucleotídeos também são responsáveis por combinar fitas de DNA, pois suas bases determinam as ligações possíveis. Assim, base (A) liga-se apenas com base (T); base (G) liga-se apenas com base (C) e vice-versa. Estas características juntamente com a formação assumida determina um modelo chamado de "Modelo de Watson-Crick". Aplicando estes conceitos, podemos representar sequências de DNA através de cadeias de letras representativas das bases, como por exemplo a fita ATAGCCTAGCCATATAGGCA é complemento da fita TATCGGATCGGTATATCCGA [Schuster 2008].

O sequenciamento de DNA é um processo que determina a ordem dos nucleotídeos para a constituição do código genético, o qual se encontra na forma de unidades chamadas genes. Dessa forma, surge o conceito de computação de DNA, já que este processo assume grande potencial devido a alta capacidade de paralelismo, de armazenamento e habilidade para resolver problemas combinatórios.

Nessa perspectiva, é almejado a construção de um projeto para computar DNA, definindo algoritmos computacionais para este fim. Em semelhança aos procedimentos bioquímicos em laboratório, a computação de DNA executa algoritmos para manipulação do DNA em uma série de processos e eventos, como redução, incisão, ligação e replicação

de DNA.

6. Modelagem do Processo de sequenciamento de DNA usando Redes de Petri

Nesta seção será abordado o comportamento do sistema de sequenciamento de DNA modelado em Redes de Petri, apontando cada processo e atividade realizados durante a execução deste. Já foi mencionado que o comportamento da Rede é determinado pelo disparo das transições, em que estas disparam somente quando lugares próximos estão ativos fazendo com que novos lugares sejam ativados.

Dessa forma, é possível comparar a ativação de lugares com fitas de DNA, em que cada lugar ativo pode representar um sequência. Além disso, o algoritmo que descreve o comportamento da Rede de Petri é análogo aos processos necessários para o sequenciamento de DNA. Assim, são descritos alguns desses processos por meio de funções:

- Amplifica: Dada uma fita N , $amplifica(N)$ produz duas cópias de si;
- Detecta: Dada uma fita N , $detecta(N)$ retorna verdadeiro se N contém parcela da sequência de DNA, caso contrário retorna falso;
- Funde: Dada as fitas $N1$ e $N2$, $funde(N1; N2)$ produz uma nova fita $N3$ formada da união das duas fitas;
- Separa: Dado uma fita N e uma sequência de DNA w composta dos nucleotídeos conhecidos, $separa(+N; w)$ produz uma nova fita $N1$ que consiste de todas as fitas em N que contém w como sub fita. Da mesma forma, $(-N; w)$ produz uma nova fita $N1$ que consiste de todas fitas em N que não contém w como sub fita.

Então, é produzido um algoritmo para ilustrar o sequenciamento de DNA comparado ao comportamento de uma Rede de Petri da Figura 1. O algoritmo inicia com uma fita $N0$, completamente "vazia", sendo equivalente a uma Rede de Petri sem ficha. Ele também usa variáveis Booleanas (variáveis que assumem apenas dois valores: verdadeiro ou falso) ($b1$, $b2$, $b3$ ou $b4$) para representar estados na Rede de Petri e presença de sequências ($s1$ a $s4$) na fita de DNA. Se a variável Booleana possui valor verdadeiro, então a sequência em questão está presente na fita, e sua ausência é indicada pelo valor falso.

```
(1) introduz(N0) =  
(2) b1 = b2 = b3 = b4 = falso,  
(3) Marca a Rede de Petri início //função para iniciar a marcação na Rede  
(4) //Por exemplo  
(5) adiciona(s1;N0)  
(6) adiciona(s3;N0)  
(7) b1 = verdade  
(8) b3 = verdade  
(9) Executa alguma tarefa  
(10) fim  
(11) n = 0  
(12) repete  
(13) introduz(N1) =  
(14) se detecta(N0(s1)) então inicia
```

```

(15)   b1 = falso,
(16)   b2 = verdade, //adiciona(s2;N1) mais tarde
(17)   Executa alguma tarefa.
(18)   fim
(19)   se detecta((N0(s2) e N0(s3)) então inicia
(20)     b2 = b3 = falso
(21)     b4 = verdade, //adiciona(s4;N1) mais tarde
(22)     Executa alguma tarefa.
(23)   fim
(24)   se detecta(N0(s4)) então inicia
(25)     b4 = falso
(26)     b1 = verdade, //adiciona(s1;N1) mais tarde
(27)     b3 = verdade, //adiciona(s3;N1) mais tarde
(28)     Executa alguma tarefa.
(29)   fim
(30)   se b1 = verdade então adiciona(s1;N1)
(31)   se b2 = verdade então adiciona(s2;N1)
(32)   se b3 = verdade então adiciona(s3;N1)
(33)   se b4 = verdade então adiciona(s4;N1)
(34)   N0 = N1
(35)   n = n + 1
(36)   até (alguma condição relativa a n é colocada).

```

Como, inicialmente a fita N0 é "esvaziada", então temos que $b1 = b2 = b3 = b4 = \text{falso}$. Na linha três inicia a marcação na Rede de Petri, através da chamada da função "Marca a Rede de Petri". Na linha cinco é adicionada a sequência s1 para a fita N0, o que é equivalente ao adcionamento de uma ficha no lugar p1 de uma Rede de Petri. Na linha seis é adicionada s3 na fita N0, o que é equivalente ao adcionamento de uma ficha no lugar p3 da Rede.

Na linha sete, temos $b1 = \text{verdade}$ e na linha oito, $b3 = \text{verdade}$, indicando a presença das sequências s1 e s3 na fita N0. Como a presença de alguma ficha na Rede de Petri dispara um evento no sistema, então temos a linha nove, referenciando alguma tarefa externa que pode ser executada.

A linha 11 introduz a variável n . O algoritmo usa esta variável no laço repete - até, onde ela define um critério de saída da aplicação (linha 36). O laço repete - até estende da linha 12 até linha 36, e contém três condicionais se - então. Cada condicional se - então faz duas coisas: primeiro, verifica o estado de uma transição específica e segundo, inclui instruções para que ocorram se uma transição "acionar".

A linha 13 introduz uma nova fita - N1. Esta fita é sempre vazia ($N1 =$) até que o laço repete - até inicie uma nova iteração. Entre as linhas 30 e 33, são alterados os valores de b1 a b4, em que as sequências (s1, s2, s3, ou s4) são adicionadas para a fita N1. Dessa forma, na linha 35 a Rede de Petri foi até um estado de transição completa. Na linha 36 é decidido se o laço entra em uma nova iteração, dependendo do novo valor de n , o qual foi incrementado na linha 35.

7. Validação da Rede de Petri usando o *software Process Maker*

Process Maker é um *software* para gerenciamento de *Workflows*, [ProcessMaker]. Ele permite a construção de processos e a execução destes, produzindo fluxos de tarefas com alocação dinâmica.

Nesta abordagem, foi possível definir um fluxo de atividades para o sequenciamento de DNA a partir da especificação formal da Rede de Petri. Estas atividades foram baseadas no algoritmo apresentado anteriormente, baseando-se nas mesmas configurações. Observa-se a grande capacidade de adequação do *workflow* desenvolvido ao algoritmo e a conformidade deste com a modelagem da Rede de Petri.

A Figura 2 e a Figura 3 ilustram o funcionamento do *Workflow* em execução.

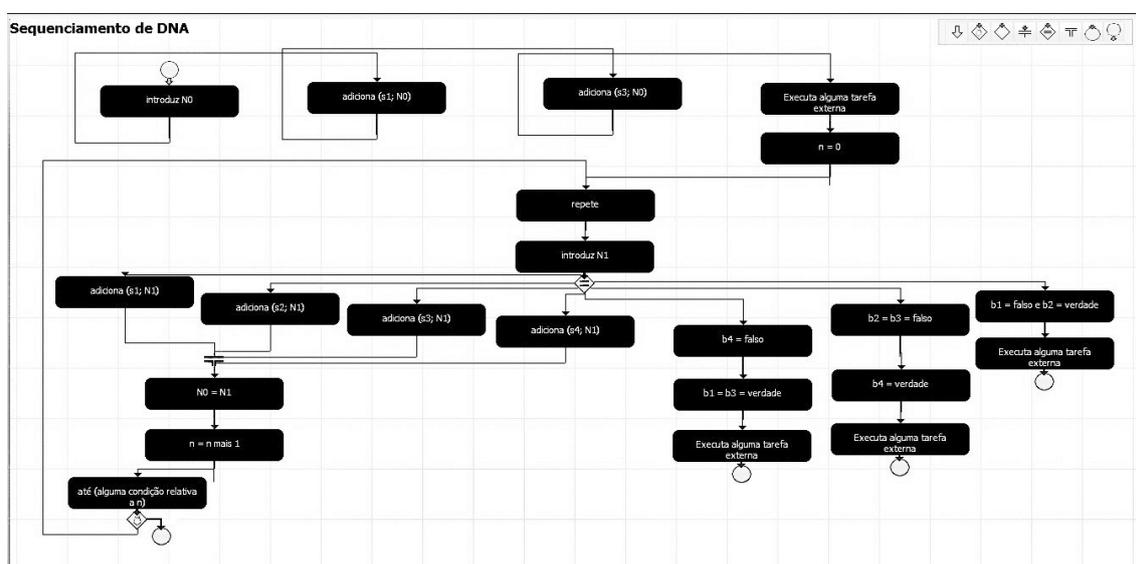


Figura 2. *Workflow* para o sequenciamento de DNA validado no *ProcessMaker*

Figura 3. Caso do sequenciamento de DNA sendo executado, em que é optado pela continuação do laço *repete* - até ou pelo término do processo

Percebe-se, então, que as Redes de Petri modelam as características e o comportamento descrito no algoritmo abordado. A consequente automatização usando *ProcessMaker* nos mostra a execução de casos do sequenciamento de DNA e a manipulação de estados provocados pelo acionamento e execução de atividades.

8. Conclusão

Neste trabalho foi demonstrado que o processo de sequenciamento de DNA pode ser representado como processos de *workflow* e a partir disso, a modelagem deste por meio da Rede de Petri. Já que esta possui capacidade de representar e preservar de forma efetiva características particulares da fita de DNA, como por exemplo, os ciclos (rotas) existentes durante o sequenciamento.

Os processos dinâmicos de sistemas biológicos e moleculares em geral podem ser modelados como um sistema dinâmico discreto [Barjis 2006]. Com isso, utilizando o conceito de *Workflows*, é realizada a validação através de um *software* característico para Processos, o *ProcessMaker*, conduzindo à formalização, modelagem e simulação da computação do sequenciamento de DNA.

Como trabalhos futuros, será construído um motor de *workflow* para a gerência de vários processos biológicos. Pois, em muitos casos o processamento é complexo, assim como o controle de tais processos.

Referências

- Barjis, J. B. I. (1999). Formalization of the protein production by means of petri nets.
- Barjis, I.; Samarrai, W. . A. I. B. J. (2006). Modeling of dna transcription and gene regulation using petri nets.
- Digiampietri, L. A. (2007). *Gerenciamento de Workflows Científicos em Bioinformática*. PhD thesis, Universidade Estadual de Campinas- Instituto de Computação.
- Lemos, M. (2004). *Workflow para Bioinformática*. PhD thesis, Pontifícia Universidade Católica do Rio de Janeiro.
- Maciel P.R.M.; Lins, R.D.; Cunha, P. (1996). *Introdução às Redes de Petri e Aplicações*.
- Meidanis, G. V. M. W. J. (1996). Using workflow management in dna sequencing.
- ProcessMaker. www.processmaker.com. Acessado em: 25 de agosto de 2011.
- Schuster, A. (2008). Dna algorithms for petri net modeling.
- Timmer, J. (2009). A brief guide to dna sequencing. <http://arstechnica.com/science/guides/2009/09/a-brief-guide-to-dna-sequencing.ars>. Acessado em: 30 de agosto de 2011.
- Vale, L. N.; Silva, N. R. D. M. S. J. S. (2010). Simulation of industrial metabolism using predicate transition petri nets combined with differential equations.
- Valette, J. C. R. (1997). *Redes de Petri*.
- WfMC. Workflow management coalition. www.wfmc.org. Acessado em: 30 de agosto de 2011.