Desenvolvimento de Modelo Geométrico dos Cromossomos levando em conta Mapas Hi-C

Palavras-Chave: Modelo, Cromossômico, Hi-C

Autores/as:

Carolina Arnosti [UNICAMP]

Prof./ª Dr./ª Mario Antonio Bernal Rodríguez (orientador/a) [UNICAMP]

INTRODUÇÃO:

O presente projeto utiliza os mapas Hi-C, os quais fornecem informação sobre os pontos de contato da fibra do DNA com si próprio - sendo determinados experimentalmente, juntamente com o modelo da fibra de cromatina de 30 nm com resolução atômica para desenvoltura do modelo geométrico dos cromossomos. Assim, pode ser utilizado em qualquer aplicação na qual seja necessário um modelo geométrico de DNA, como por exemplo, em investigações relacionadas aos efeitos das radiações ionizantes no material genético humano. Essas investigações são de larga experiência no estudo a partir da utilização do código Geant4-DNA. Inicialmente, foi estudado a literatura do modelo geométrico de resolução atômica da configuração B-DNA, o modelo da fibra de cromatina de 30 nm. Uma vez que, para simulações de interação DNA-radiação, é necessário um formalismo e código associado para determinar o átomo mais próximo de um ponto arbitrário no espaço. Assim, esses pontos correspondem aos locais de deposição de energia e posição das espécies químicas reativas, produzidas pela radiólise da água principalmente. Posteriormente, foi estudado a literatura referente ao modelo Hi-C, método de ligação de proximidade que apreende a estrutura organizacional da cromatina em três dimensões, em que sequências genômicas distantes entre si linearmente podem estar próximas na conformação tridimensional. E, por fim, foram realizadas simulações no Geant4-DNA de feixes de prótons e partículas alfa de 2 e 10 MeV, a fim de obter a probabilidade de dano sobre o modelo de DNA em questão para três passos da supra-hélice da fibra de cromatina de 30 nm.

METODOLOGIA:

Para o desenvolvimento do modelo, foi estabelecido inicialmente a conformação de um par de base (pb), visto que é o elemento fundamental para construção de todo o modelo de estrutura do genoma desenvolvido. Esse par de base foi resultado do conhecimento da posição explícita dos átomos conformando um par de nucleotídeos, retirada da página web do projeto GLACTONE [5] [1]. Assim, obteve-se a posição, elemento químico e a base nitrogenada correspondente a cada um dos 63 átomos que compõem esse par de base. Em seguida, essas coordenadas foram transformadas referenciando-

as ao centro geométrico do par de bases, ponto que foi determinado pela semissoma das coordenadas extremas dos átomos e todo o par de base foi girado −50° em torno do eixo Z para alinhar o eixo X à dimensão alongada do par [1]. Com isso, foi realizada a combinação das coordenadas dadas e os raios de van der Waals correspondentes, o que resulta nas sobreposições de átomos. Por fim, a dupla hélice é formada por torção e deslocamento de cada par de base 36° e 0,33 nm, respectivamente [1].

A determinação das dimensões das estruturas ocorreu a partir de relações geométricas, de acordo com o diagrama de configuração do nucleossomo dentro da fibra de cromatina, considerando 6 nucleossomos para uma fibra de cromatina de 30 nm. Uma vez que, esse modelo cumpre a razão de empacotamento dos diferentes níveis de organização de um material genético humano típico [1]. Assim, esta configuração (Figura 1) foi usada para determinar a razão de raio de nucleossomo para cromatina.

- $sen\frac{\pi}{6} = \frac{R_{nucl}}{r} \rightarrow r = 2R_{nucl}$ [1]
- $2r + 2R_{nucl} = D_{chrom} \rightarrow R_{nucl} = D_{chrom}/6$ [1]

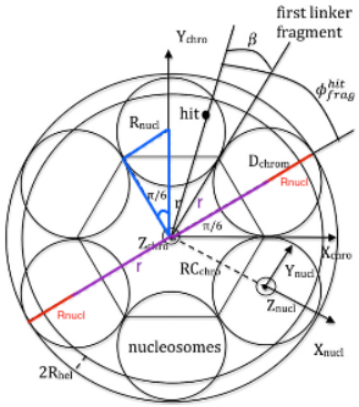

Figura 1- Diagrama da configuração do nucleossomo dentro da fibra de cromatina. O primeiro fragmento de ligação liga-se aos dois primeiros nucleossomos [1].

O resultado foi de que, para uma fibra de cromatina de 30 nm, o nucleossomo possui um raio de 5 nm, de acordo com o esperado. Posteriormente, foram calculados o número de pares de bases por nucleossomo e seu raio, bem como o diâmetro da fibra de cromatina. Considerando que cada pb ocupa 0,33 nm ao longo do eixo da dupla hélice e que cada nucleossomo deve conter 200 pb, o sistema de equações foi:

- $200\,pb = \frac{l_{nucl} + l_{bind}}{0{,}33nm/pb}$ [1]

com $l_{nucl} = tamanho\ do\ eixo\ da\ dupla\ hélice$ e $l_{bind} = tamanho\ do\ fragmento\ de\ ligação.$

- $R_{nucl} = D_{chrom}/6$ [1]
- $RC_{nucl} = R_{nucl} - R_{hel}$ [1]
- $l_{nucl} = 4\pi RC_{nucl}$ [1]
- $l_{bind} = \frac{\pi}{3}\left(\frac{D_{chrom}}{2} - R_{hel}\right)$ [1]

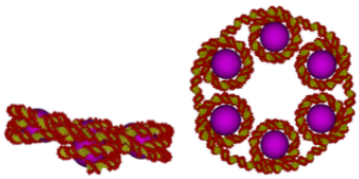

Figura 2 - -Fragmento da fibra de cromatina de 30 nm contendo 6 nucleossomos (um passo de hélice). [1]

Com a resolução desse sistema de equações, obteve-se um diâmetro de 31,380 nm e descobriu-se que as duas voltas helicoidais do DNA no nucleossomo contêm 154 pb e o fragmento de ligação, 46 pb, totalizando até 200 pb por nucleossomo, novamente dentro do esperado. Por fim, o nucleossomo é construído dobrando duas alças helicoidais em torno de uma esfera que simula uma histona [1]. Portanto, a fibra de cromatina de 30 nm é formada por uma hélice com 6 nucleossomos/volta e dois nucleossomos adjacentes estão ligados por um fragmento de DNA contendo 46 pb, conforme demonstrado na Figura 2 [1]. Em seguida, o modelo foi estabelecido ao algoritmo que busca encontrar o átomo alvo mais próximo no nucleossomo em relação ao local de acerto (hit). São utilizadas algumas mudanças de coordenadas para a realização

deste processo. Inicialmente, é definido o nucleossomo mais próximo ao hit, com isso o hit é determinado por coordenadas em relação ao sistema do nucleossomo. Após, é estabelecido o par de base mais próximo ao hit e assim, o hit sofre uma mudança de coordenadas para o sistema de par de bases. Por fim, o átomo mais próximo do hit é encontrado com a varredura dos 63 átomos que compõem o par de base. Posteriormente, foi estudado o método Hi-C. A literatura estudada está nas referências [2], [3], [4] e [6]. Então, entende-se que como o modelo de conformação tridimensional dos cromossomos está envolvido na compartimentação do núcleo e na aproximação espacial tridimensional de certos elementos linearmente separados, torna-se necessário compreender como o genoma está dobrado no interior do núcleo da célula. Uma vez que, essa compreensão fornece informações sobre as relações entre a cromatina, a atividade do gene e o estado funcional da célula [2].

Logo, o método de mapas Hi-C busca identificar quais partes da cromatina estão em contato/ interação consigo própria em todo o genoma, e para isto, as células são reticuladas com formaldeído, e o DNA é recortado com uma enzima de restrição que deixa uma proeminência 5'. Em seguida, essa proeminência é preenchida com a inclusão de um resíduo biotinilado e as extremidades desses fragmentos são ligadas sob condições diluídas que favorecem eventos de ligação entre os fragmentos de DNA reticulados. A amostra de DNA resultante desse processo contém produtos de ligação que são originalmente fragmentos próximos espacialmente no núcleo, que estão, após o processo, marcados com biotina na junção. Por fim, um mapa Hi-C é criado buscando as junções com a marcação de biotina em cortes do DNA com esferas de estreptavidina e são identificadas a partir do sequenciamento de DNA massivamente paralelo, produzindo um catálogo de fragmentos de interação [2]. Esses mapas, portanto, fornecem uma representação bidimensional da organização tridimensional do genoma.

Por conseguinte, foi descoberto que os genomas são divididos em domínios de contato a partir de loops. Esses domínios estão associados a padrões distintos de marcas de histonas e, portanto, se segregam em subcompartimentos. Esses loops ligam promotores e intensificadores, correlacionados à ativação do gene, e mostram conservação em todos os tipos de células e espécies [2]. As bases desses loops que os ancoram são proteínas CTCF (proteína de dedo de zinco altamente conservada - mais conhecida como fator de transcrição, pode atuar também como ativador transcricional, um repressor ou isolante [4]) em orientação convergente.

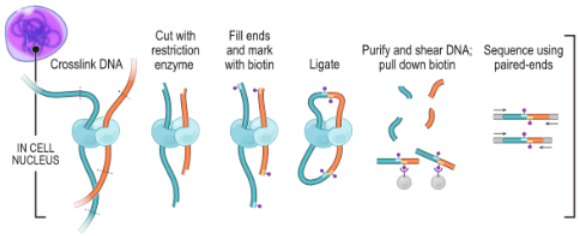

Figura 3 - Exemplificação durante in situ Hi-C, como a ligação de proximidade DNA-DNA é realizada em núcleos intactos. [2]

Dessa forma, foi utilizado Hi-C para demonstrar que o genoma é particionado em vários domínios que se dividem em compartimentos distintos. Análises subsequentes sugeriram a presença de domínios menores e levaram à importante proposta de que os compartimentos são particionados em estruturas condensadas de 1 Mb de tamanho, apelidadas de "domínios topologicamente associados" (TADs) (Dixon et al., 2012; Nora e outros, 2012) [2]. Os resultados da literatura referenciada em [2] mostram um mapeamento de forma abrangente dos contatos de cromatina em todo o genoma, usando in situ Hi-C, em que a ligação de proximidade DNA-DNA é

realizada em núcleos intactos. A importância desses mapas é a capacidade de discernir a estrutura do domínio, a compartimentalização e os milhares de loops de cromatina. Os mapas fornecem uma imagem da arquitetura genômica com resolução de até 1 kb, conforme demonstrado abaixo [2].

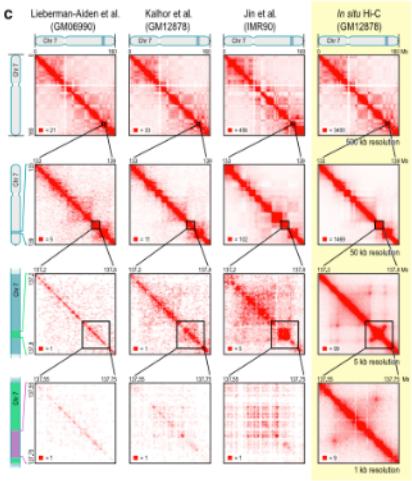

Figura 5 - Comparação do mapa do cromossomo 7 em GM 12878 (última coluna) com mapas Hi-C anteriores: Lieberman-Aiden et al. (2009), Kalhor et al. (2012), e Jin et al. (2013) [2]

RESULTADOS E DISCUSSÃO:

O modelo, então, foi testado com dados de entrada provenientes de simulações de Monte Carlo. As simulações realizadas no Geant4-DNA de feixes de prótons e partículas alfa de 2 10 MeV, foram usadas para determinar as probabilidades de dano físico sobre o modelo de DNA em questão para três passos da supra-hélice da fibra de cromatina de 30nm, passos de 7.11 nm, 10.77 nm e 14.44 nm. Assim, após rodar a simulação de irradiação com um feixe de partículas pesadas (prótons ou partículas alfas) em processos paralelos, é necessário rodar um código específico para gerar um único arquivo com deposições de energia, agrupando todos esses processos que foram rodados em paralelo. E, em seguida, rodar o código HiCModel para cada passo da supra-hélice, o qual superpõe as partículas sobre o modelo do DNA proposto a fim de calcular as probabilidades de dano. As probabilidades de danos apresentados para cada feixe com sua respectiva energia correspondem, primeiramente à dose total média (Mean Total Dose), à probabilidade de ruptura total da fita (TSB yield) e para quebras de fita dupla (DSB yield).

Para a obtenção desses dados buscou-se uma incerteza em torno de 20% a 40%. Os resultados adquiridos estão apresentados na tabela 1 abaixo. Cabe ressaltar que os tempos de simulação foram longos, o que impediu conseguir incertezas menores. Assim, todos os rendimentos de danos foram determinados por dose unitária absorvida e número de par de base (1/Gy Gbp).

Os resultados nos mostram que, como as partículas alfas apresentam maior LET, o rendimento de dano TSB e DSB são maiores para tais.

PRÓTON - 2 MeV	7.11nm	10.77nm	14.44nm
Mean Total Dose (Dosis per batch)	$(16.13 \pm 0.34)\times10^{-1}$	$(15.66 \pm 0.80)\times10^{-1}$	$(150.52 \pm 0.10)\times10^{-2}$
TSB yield physics (1/GyGbp)	$(6.30 \pm 0.70)\times10^{1}$	$(6.90 \pm 0.80)\times10^{1}$	$(7.70 \pm 0.70)\times10^{1}$
DSB yield physics (1/GyGbp)	(2.02 ± 0.54)	(3.61 ± 0.97)	(3.50 ± 0.90)
NTSB physics	$(2.53 \pm 0.30)\times10^{1}$	$(2.70 \pm 0.31)\times10^{1}$	$(2.88 \pm 0.25)\times10^{1}$
NDSB physics	(0.80 ± 0.20)	(1.40 ± 0.37)	(1.30 ± 0.33)
PRÓTON - 10 MeV	7.11nm	10.77nm	14.44nm
Mean Total Dose (Dosis per batch)	$(28.22 \pm 0.20)\times10^{-1}$	$(160.10 \pm 0.20)\times10^{-2}$	$(667.60 \pm 0.83)\times10^{-3}$
TSB yield physics (1/GyGbp)	$(7.40 \pm 0.50)\times10^{1}$	$(1.13 \pm 0.09)\times10^{2}$	$(1.80 \pm 0.13)\times10^{2}$
DSB yield physics (1/GyGbp)	(1.85 ± 0.53)	(2.51 ± 0.92)	(5.93 ± 2.34)
NTSB physics	$(5.20 \pm 0.33)\times10^{1}$	$(4.50 \pm 0.34)\times10^{1}$	$(3.01 \pm 0.22)\times10^{1}$
NDSB physics	(1.30 ± 0.37)	(1.00 ± 0.37)	(1.00 ± 0.40)
ALFA - 2 MeV	7.11nm	10.77nm	14.44nm
Mean Total Dose (Dosis per batch)	(6.52 ± 0.40)	(3.60 ± 0.20)	$(15.34 \pm 0.80)\times10^{-1}$
TSB yield physics (1/GyGbp)	$(5.00 \pm 0.70)\times10^{1}$	$(6.54 \pm 0.80)\times10^{1}$	$(1.28 \pm 0.17)\times10^{2}$
DSB yield physics (1/GyGbp)	(3.61 ± 0.70)	(4.61 ± 1.20)	$(1.34 \pm 0.30)\times10^{1}$
NTSB physics	$(7.90 \pm 1.15)\times10^{1}$	$(5.65 \pm 0.55)\times10^{1}$	$(5.00 \pm 0.80)\times10^{1}$
NDSB physics	(6.90 ± 1.30)	(3.90 ± 0.84)	(5.30 ± 1.15)
ALFA - 10 MeV	7.11nm	10.77nm	14.44nm
Mean Total Dose (Dosis per batch)	(6.63 ± 0.41)	(3.65 ± 0.21)	$(15.57 \pm 0.91)\times10^{-1}$
TSB yield physics (1/GyGbp)	$(4.95 \pm 0.55)\times10^{1}$	$(7.21 \pm 0.80)\times10^{1}$	$(1.52 \pm 0.18)\times10^{2}$
DSB yield physics (1/GyGbp)	(3.15 ± 0.70)	(4.37 ± 0.86)	$(1.53 \pm 0.20)\times10^{1}$
NTSB physics	$(8.20 \pm 1.14)\times10^{1}$	$(6.40 \pm 0.65)\times10^{1}$	$(5.78 \pm 0.63)\times10^{1}$
NDSB physics	(6.50 ± 1.60)	(4.00 ± 0.91)	(5.80 ± 0.70)

Tabela 1 - Probabilidade de dano para 4 feixes de partículas carregadas (prótons e partículas alfas) com energias de 2 e 10 MeV para os respectivos passos da fibra de 7.11nm, 10.77nm e 14.44nm.

No entanto, foi necessário rever os resultados obtidos acima, uma vez que a dose total média (Mean Total Dose) deve ser aproximada para os três passos de fibra considerando um feixe de energia específica. Essa dose média deve ser a mesma, uma vez que conceitualmente tem-se que a dose é o produto da fluência e do “stopping power” de colisão em massa, considerando que esta dose independe da espessura do material (desde que as partículas se desloquem e não percam energia suficiente para fazer com que o “stopping power” mude significativamente). Portanto, como não há alteração na fluência nas simulações acima, a dose total média deveria ser a mesma para os três passos de fibra considerando um feixe de energia específico. Logo, foi revisado as dimensões da caixa que fornece a região de interesse nas simulações Monte Carlo e novos cálculos e resultados estão em estudo.

CONCLUSÕES:

Foi necessário rever os resultados obtidos acima, uma vez que a dose total média deve ser aproximada para os três passos de fibra considerando um feixe de energia específica. Além do mais, os resultados acima apresentam somente o dano físico da radiação nesse modelo cromossômico desenvolvido, está também em progresso as simulações para os danos químicos.

BIBLIOGRAFIA

[1] "An atomistic geometrical model of the B-DNA configuration for DNA–radiation interaction simulations", de M.A. Bernal, D. Sikansi, F. Cavalcante, S. Incerti, C. Champion, V. Ivanchenko, Z. Francis, https://doi.org/10.1016/j.cpc.2013.07.015.

[2] "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping", Suhas S.P. Rao, Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, Ido Machol, Arina D. Omer, Eric S. Lander, and Erez Lieberman Aiden, http://dx.doi.org/10.1016/j.cell.2014.11.021

[3] "Comprehensive mapping of long range interactions reveals folding principles of the human genome", Erez Lieberman-Aiden, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R. Lajoie, Peter J. Sabo, Michael O. Dorschner8, Richard Sandstrom, Bradley Bernstein, M. A. Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A. Mirny, Eric S. Lander, and Job Dekker, doi:10.1126/science.1181369.

[4] "CTCF as a multifunctional protein in genome regulation and gene expression", Somi Kim, Nam-Kyung Yu and Bong-Kiun Kaang, https://doi.org/10.1038/emm.2015.33.

[5] GLACTONE, http://chemistry.gsu.edu/Glactone/PDB/DNA_RNA/dna.html.

[6] A Cell Paperflick of "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping", directed by Suhas Rao, Sarah Nyquist, Miriam Huntley.