



UNIVERSIDADE D
COIMBRA

Sónia Filipa Gomes dos Santos

ESTUDOS DE DINÂMICA MOLECULAR EM
ENDONUCLEASES DE RESTRIÇÃO DO TIPO 2
DE RARO CORTE

Dissertação do Mestrado em Bioquímica orientada pela Doutora Alexandra Pires
Carvalho e apresentada ao Departamento de Ciências da Vida na Faculdade de
Ciências e Tecnologia da Universidade de Coimbra

Setembro de 2019

1 2



9 0

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

Estudos de dinâmica molecular em endonucleases de restrição do tipo 2 de raro corte

ESTUDO DA ENZIMA SgrAI

Dissertação do Mestrado em Bioquímica orientada pela Doutora Alexandra Pires Carvalho e apresentada ao Departamento de Ciências da Vida na Faculdade de Ciências e Tecnologia da Universidade de Coimbra

Sónia Filipa Gomes dos Santos

2019

“It is always seems impossible until it is done”
Nelson Mandela

Quando me propus ao Mestrado em Bioquímica, e em concreto ao salto para fora da zona de conforto que representa a área da química computacional, logo se previa uma longa viagem, que incluía inúmeros desafios e incertezas, mas também muitas alegrias, superação e desafio. Este trabalho reúne contributos de várias pessoas, indispensáveis durante cada momento da caminhada.

Trilhar este caminho só foi possível com o apoio, energia e força de várias pessoas, a quem dedico especialmente este projeto.

Em primeiro lugar, à minha orientadora, Doutora Alexandra Carvalho, que sempre acreditou em mim, agradeço a orientação exemplar, pela visão crítica e oportuna. O seu empenho foi determinante a cada passo deste trabalho. Muito Obrigado!

À Beatriz, ao Pedro, ao Daniel e à Ana, agradeço o apoio, a motivação e a paciência. Ajudaram a tornar este trabalho uma agradável experiência de aprendizagem. Estou grata pela vossa ajuda, especialmente a ti Pedro pelas imagens fantásticas (you rock!).

Agradeço à Professora Paula Veríssimo, que ao saber da minha disponibilidade limitada me encorajou a seguir em frente e a lutar, ainda que o esforço fosse maior.

Agradeço aos meus pais, aos meus amigos e colegas de trabalho. Sem vocês, sem o apoio característico e os desafios que me propuseram este trabalho não teria o mesmo sabor.

Agradeço a todos os meus amigos, que sempre acreditaram em mim, aos coaches do Crossfit Mondego que lidavam com as minhas neuras após muitas horas de trabalho e ajudavam a manter o meu equilíbrio.

Ao meu irmão Alexandre, que é sempre a minha claque e aquele a que jamais tenciono desiludir, por ti faço sempre mais.

À minha querida Tânia que sempre me orienta quando me sinto perdida.

À minha amora, Ritinha que está sempre presente.

À Professora e amiga Fernanda Ferreira que sempre acreditou.

E por último ao meu namorado, Bruno Casimiro, que deu uma volta completa na minha vida, no entanto nessa volta manteve sempre o meu mundo seguro. Sou grata pelo amor, partilha, companheirismo e apoio incondicional, agradeço a enorme compreensão, generosidade e alegria constante, contribuindo para chegar ao fim deste percurso.

1. Resumo

A enzima *SgrAI* pertence às endonucleases de restrição do tipo IIF. É descrita como altamente específica devido não só à sua sequência de reconhecimento de 8 pares de bases, bem como tendo modulação alostérica. Porém, a reação de catálise da *SgrAI* ainda se encontra muito pouco caracterizada e, apesar de se esperar que esta ocorra da mesma forma que as enzimas que usam o mesmo cofator, dois iões de Mg^{2+} , ainda não há evidências publicadas.

Dessa forma, o objetivo deste trabalho passa por caracterizar em detalhe as ligações entre o complexo enzima/ADN e a especificidade das mesmas através de métodos baseados em dinâmica molecular clássica e métodos de QM/MM.

Para este estudo foram utilizados quatro modelos da enzima *SgrAI*, mais concretamente a forma de dímero, dímero com uma mutação num dos pares de bases, em forma de tetrâmero com iões de Ca^{2+} e na forma de tetrâmero com iões de Mg^{2+} . Tanto para os modelos diméricos como para o modelo tetramétrico com iões de Ca^{2+} apenas foram realizadas simulações de dinâmica molecular para a estrutura inicial. Para o modelo tetramétrico com iões de Mg^{2+} , descrito como sendo a forma mais eficiente da *SgrAI*, o estudo prosseguiu utilizando métodos QM/MM, com o resíduo Lys em diferentes estados de protonação, para caraterizar o mecanismo catalítico da enzima. Em todas as hipóteses testadas a posição do segundo ião magnésio revelou-se problemática e as barreiras de energia livre associadas foram sempre demasiado elevadas para sistemas biológicos. Este facto leva-nos a propor que o mecanismo para esta enzima poderá ter variações em relação ao mecanismo dos dois iões metálicos.

Palavras-chave: *SgrAI*, endonucleases tipo IIF, catálise, Dinâmica Molecular, QM/MM

2. Abstract

The *SgrAI* enzyme belongs to the type IIF restriction endonucleases. It is described as being highly specific due not only to its 8 base pair recognition sequence as well as having allosteric modulation. However, the catalysis reaction of *SgrAI* is still poorly characterized and although it is expected to occur in the same way as the enzymes that use the same cofactor, two Mg^{2+} ions, there is still no published evidence.

Thus, the purpose of the present work is to characterize in detail the interactions between the enzyme/ADN complex and the specificity of the same through Classical molecular dynamics and QM/MM methods.

In order to attain that objective four models were used, namely *SgrAI* in the form of dimer, dimer with a mutation in one of the bases, in the form of tetramer with Ca^{2+} ions and in the form of tetramer with Mg^{2+} ions. The initial structures for the dimeric models and the tetrameric model with the Ca^{2+} ions were subjected to MD simulations. The tetrameric model with Mg^{2+} ions, which is described as the most efficient form of *SgrAI*, was also studied using QM/MM methods with Lys in different protonation states. In all tested hypothesis, the position of the second metal ion was problematic and the associated free energy barriers were too high for a biological reaction. This led us to propose that, for this specific case, there could be deviations to the archetypal two-metal ions catalysis mechanism.

Key words: *SgrAI*, type IIF endonucleases, catalysis, Molecular Dynamics, QM / MM.

3. Índice

1. Resumo.....	v
2. Abstract.....	vii
3. Índice	ix
4. Índice de Figuras.....	xi
5. Índice de Equações.....	xvii
6. Índice de Tabelas.....	xix
7. Lista de Abreviaturas.....	xxi
8. Objetivo.....	xxiv
9. Introdução	1
10. Métodos Computacionais	15
10 Resultados e Discussão	25
11 Conclusão	49
12 Bibliografia.....	51
Anexo I.....	57
Anexo II.....	59
Anexo III.....	61
Anexo IV	63
Anexo V	65

4. Índice de Figuras

Figura 1 Representação das estruturas de diferentes tipos de endonucleases pertencentes ao tipo II das endonucleases de restrição. Adaptado de Pingoud, A., Fuxreiter, M., Pingoud, V., & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. <i>Cell Mol Life Sci</i> , 62(6), 685-707. doi:10.1007/s00018-004-4513-1.....	2
Figura 2 Representação esquemática dos métodos utilizados em química computacional	5
Figura 3 Representação dos parâmetros necessários para representação dos modelos moleculares. Ilustração adaptada de van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., . . . Yu, H. B. (2006). Biomolecular Modeling: Goals, Problems, Perspectives. <i>Angewandte Chemie International Edition</i> , 45(25), 4064-4092. doi:10.1002/anie.200502655.	6
Figura 4 Ilustração do princípio do método híbrido QM/MM com divisão do sistema em duas partes. Na figura temos representada a parte do sistema <i>SgrAI</i> &ADN onde foi o utilizado QM nos resíduos representados dentro da “nuvem” a <i>licorice</i> e <i>CPK</i> e o restante sistema tratado apenas com MM representado com <i>New Cartoon</i>	14
Figura 5 <i>Workflow</i> da metodologia aplicada no estudo da endonuclease <i>SgrAI</i>	16
Figura 6 Modelo do mecanismo da clivagem do ADN na presença de dois íons metálicos, identificados pelas posições M1 e M2 que devem ser ocupadas pelos íons Mg^{2+} . As setas a vermelho identificam as alterações nas ligações, representando de forma esquemática o mecanismo catalítico esperado. Imagem adaptada de (Dunten et al., 2008).....	26
Figura 7 Ilustração do dímero da enzima <i>SgrAI</i> representado com <i>New Cartoon</i> , o ADN cognato representado com <i>Tube</i> e os íons de Mg^{2+} como cofator representados como esferas amarelas.....	28
Figura 8 Figura representativa do RMSD dos átomos do <i>backbone</i> do ADN (P, O3', O5', C3', C4', C5').....	29
Figura 9 Figura representativa do RMSD dos resíduos de interesse da proteína <i>SgrAI</i> .	30
Figura 10 Representação visual dos nucleótidos DC 696 e DG 697 ao centro em <i>Licorice</i> , a proteína é representada em <i>New Cartoon</i> , a molécula de ADN em <i>Tube</i> e os íons Mg^{2+} representados como esferas de cor salmão.	31
Figura 11 Representação do centro ativo do dímero da enzima <i>SgrAI</i> com <i>Licorice</i> e <i>CPK</i> , o restante sistema está representado com <i>New Cartoon</i> e <i>Tube</i>	32
Figura 12 Representação do centro ativo da enzima e da posição incorreta dos resíduos representados em relação ao que seria esperado. Destaque para o distanciamento do Mg^{2+} MG707 ao nucleótido DA 695.	33
Figura 13 Representação gráfica da análise do RMSD considerando a variação média das distâncias dos carbonos α da proteína.....	34
Figura 14 Figura representativa do par de bases DA 696 - DT 679 e dos resíduos que interagem com o nucleótido DA 696.....	35
Figura 15 Representação gráfica da análise do RMSD considerando a variação média das distâncias dos resíduos de interesse.	36

Figura 16 Representação gráfica da análise do RMSD considerando a variação média das distâncias dos carbonos α da proteína.	38
Figura 17 Gráfico representativo do perfil da energia durante a aproximação da WAT1773 ao átomo de fósforo do nucleótido DC 1364.	40
Figura 18 Gráfico representativo do perfil da energia durante a aproximação do MG1426 ao <i>backbone</i> do ADN.	41
Figura 19 Gráfico representativo do perfil da energia durante a aproximação do da WAT1773 ao fosfato do nucleótido DC 1364 partindo da estrutura com os iões Mg^{2+} na posição correta para a catálise.	42
Figura 20 Imagem representativa do centro ativo no modelo tetramérico da enzima <i>SgrAI</i> com a Lys neutra.	42
Figura 21 Esquema das interações em torno do Mg^{2+} iões da estrutura do <i>SgrAI</i>	43
Figura 22 Representação gráfica da análise do RMSD considerando a variação média das distâncias dos resíduos de interesse.	44
Figura 23 Gráfico representativo do perfil da energia durante a aproximação do ião Mg^{2+} para a posição correta para a catálise.	46
Figura 24 Gráfico representativo do perfil da energia durante a aproximação do resíduo WAT 1763 ao fosfato da DC 1364.	46
Figura 25 Imagem representativa do centro ativo no modelo tetramérico da enzima <i>SgrAI</i> com a Lys protonada.	47
Figura 26 Gráfico representativo do perfil da energia durante o ataque nucleófilo do grupo hidroxilo ao fosfato do nucleótido DC 1364.	48
Figura 27 Gráfico representativo da variação da temperatura (K) durante a DM do dímero de <i>SgrAI</i>	57
Figura 28 Gráfico representativo da variação da pressão (atm) durante a DM do dímero de <i>SgrAI</i>	57
Figura 29 Figura representativa da energia do Sistema durante a DM do dímero de <i>SgrAI</i> : a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total.	57
Figura 30 Gráfico representativo da variação da temperatura (K) durante a DM do dímero de <i>SgrAI</i> com mutação no par de bases DC-DG.	59
Figura 31 Gráfico representativo da variação da pressão (atm) durante a DM do dímero de <i>SgrAI</i> com mutação no par de bases DC-DG.	59
Figura 32 Figura representativa da energia do sistema durante a DM do dímero de <i>SgrAI</i> com mutação no par de bases DC-DG: a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total.	59
Figura 33 Gráfico representativo da variação da temperatura (K) durante a DM do tetrâmero de <i>SgrAI</i> com iões de Ca^{2+}	61
Figura 34 Gráfico representativo da variação da pressão (atm) durante a DM do tetrâmero de <i>SgrAI</i> com iões de Ca^{2+}	61
Figura 35 Figura representativa da energia do sistema durante a DM do tetrâmero de <i>SgrAI</i> com iões de Ca^{2+} : a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total.	61
Figura 36 Gráfico representativo da variação da temperatura (K) durante a DM do tetrâmero de <i>SgrAI</i> com iões de Mg^{2+} e Lys neutra.	63

Figura 37 Gráfico representativo da variação da pressão (atm) durante a DM do tetrâmero de <i>SgrAI</i> com iões de Mg^{2+} e Lys neutra.	63
Figura 38 Figura representativa da energia do sistema durante a DM do tetrâmero de <i>SgrAI</i> com iões de Mg^{2+} e Lys neutra: a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total.	63
Figura 39 Gráfico representativo da variação da temperatura (K) durante a DM do tetrâmero de <i>SgrAI</i> com iões de Mg^{2+} e Lys protonada.	65
Figura 40 Gráfico representativo da variação da pressão (atm) durante a DM do tetrâmero de <i>SgrAI</i> com iões de Mg^{2+} e Lys protonada.	65
Figura 41 Figura representativa da energia do sistema durante a DM do tetrâmero de <i>SgrAI</i> com iões de Mg^{2+} e Lys protonada: a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total.	65

5. Índice de Equações

- Equação 1** Equação representativa da energia potencial adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71. 7
- Equação 2** Termo que corresponde à variação de energia causada pela deformação da ligação entre dois átomos. Representação adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71. 7
- Equação 3** Parcela da equação representativa dos ângulos de flexão entre átomos. Representação adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71. 8
- Equação 4** Parcela da equação representativa dos ângulos diedros. Representação adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71. 8
- Equação 5** Parcela da equação representativa das interações intermoleculares correspondentes às interações electrostáticas descritas pela lei de Coulomb. Representação adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71. 9
- Equação 6** Equação representativa do modelo clássico utilizados na DM, equação retirada de Fuxreiter, M. (2014). *Computational Approaches to Protein Dynamics: From Quantum to Coarse-Grained Methods*. Boca Raton: Taylor & Francis Group. 10
- Equação 7** Representação da equação que descreve o método híbrido QM/MM descrito por Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of Lys ozyme. *J Mol Biol*, 103(2), 227-249. doi:10.1016/0022-2836(76)90311-9. 14

6. Índice de Tabelas

Tabela 1 Representação dos <i>inputs</i> utilizados para a minimização do sistema.	19
Tabela 2 Representação do <i>input</i> utilizado para o aquecimento do sistema para simulação da DM de 0 K a 300K.	20
Tabela 3 Representação do <i>input</i> utilizado para simulação da DM.	21
Tabela 4 Tabela referente à distância média das interações entre os resíduos da proteína <i>SgrAI</i> e o ADN durante a simulação.	31
Tabela 5 Tabela referente à variação da distância média das interações entre os resíduos da proteína <i>SgrAI</i> e o ADN.	37
Tabela 6 Tabela com distância médias das interações entre os resíduos da proteína <i>SgrAI</i> e o ADN.	39
Tabela 7 Tabela com distância médias das interações entre os resíduos da proteína <i>SgrAI</i> e o ADN.	45

7. Lista de Abreviaturas

A lista que se segue contém todas as abreviaturas usadas ao longo deste relatório. Unidades e símbolos do sistema métrico que seguem o Sistema Internacional de Unidades bem como os códigos para aminoácidos e bases de ADN definidos pela IUPAC não serão incluídos nesta lista, uma vez que publicações de autoridade reconhecida sobre este tópico podem ser consultados. Para distinguir o ADN do ARN a letra D foi colocada antes da respetiva base.

ADN	Ácido desoxirribonucleico
AMBER	<i>Assisted Model Building with Energy Refinement</i>
ATP	Trifosfato de adenosina
DFT	Teoria funcional de densidade
DM	Dinâmica molecular
GAFF	<i>General AMBER Force Field</i> (Campo forças geral do software AMBER)
HWMS	Espécies de alto peso molecular
MM	Mecânica molecular
PDB	Banco de dados de proteínas
PMEMD	<i>Particle Mesh Ewald Molecular Dynamics</i>
QM	Mecânica Quântica

RMSD	<i>Root-mean-square deviation of atomic positions</i> (Desvio médio quadrático da distância média entre os átomos)
SANDER	<i>Simulated Annealing with NMR-Derived Energy Restraints</i>
VMD	Visualização de dinâmica molecular

8. Objetivo

O objetivo deste trabalho é caracterizar detalhadamente o complexo endonuclease de restrição *SgrAI* com o ADN, nomeadamente as interações a nível molecular essenciais para que ocorra a reação de corte específica, recorrendo a métodos de dinâmica molecular e cálculos de mecânica quântica/mecânica molecular.

9. Introdução

9.1. Endonucleases de restrição

As endonucleases de restrição são enzimas capazes de se ligar ao ADN e reconhecer uma sequência palindrômica de nucleótidos (sequência de reconhecimento ou cognata). A sua principal função biológica é a proteção do ADN do hospedeiro face ao ADN exógeno, em particular o ADN de um bacteriófago (Pingoud & Jeltsch, 2001).

Por definição estas enzimas fazem parte de um sistema de restrição-modificação (RM), que inclui uma endonuclease e uma metiltransferase. O seu substrato é o ADN exógeno que em resposta ao seu reconhecimento pela sequência específica é clivado nesse local. No entanto, o ADN do hospedeiro está protegido do ataque da endonuclease uma vez que este se encontra metilado na sequência de reconhecimento (Smith & Nathans, 1973).

As endonucleases de restrição dividem-se em quatro tipos: tipos I, II, III e IV. Todos os tipos destas enzimas reconhecem sequências de ADN e realizam a clivagem do ADN, originando fragmentos específicos de cadeia dupla com grupos fosfatos nos terminais 5'. Os tipos de enzimas diferem na sequência de reconhecimento, composição da subunidade, posição de clivagem e tipo de cofator.

Assim, as enzimas do tipo I clivam em locais remotos do local de reconhecimento, são dependentes de ATP e de S-adenosil-L-metionina e são enzimas multifuncionais com atividades de restrição e de metilase. As enzimas do tipo II clivam no local de reconhecimento ou a curtas distâncias específicas; a maioria requer magnésio (Mg^{2+}) e a sua única função é a de restrição independentes da metilase (Bickle & Kruger, 1993). Sistemas RM (Restrição-Modificação) do tipo III consistem em duas subunidades (Mod e Res), que formam uma holoenzima funcional, com atividade de restrição e modificação. Esta classe enzimática reconhece sequências específicas assimétricas e corta o ADN a uma distância fixa destas regiões de 25-37bp relativamente a um dos lados do local de reconhecimento. (Dryden, Murray, & Rao, 2001). Por fim, as enzimas do tipo IV têm como alvo o ADN metilado (Bickle & Kruger, 1993).

As endonucleases de restrição do tipo II (figura 1) caracterizam-se pelo reconhecimento de uma curta sequência normalmente de 4-8pb, na presença do cofator Mg^{2+} , clivando o ADN na sequência de reconhecimento ou muito perto dela. De acordo com as características particulares de cada uma, estas enzimas subdividem-se em diferentes subtipos. No geral, todas possuem em comum a estrutura do *core*, possuindo quatro cadeias β e uma hélice α (Aggarwal, 1995; J.M., 2004; Niv et al., 2007; Venclovas, Timinskas, & Siksnys, 1994)

Tal como outras proteínas, estas enzimas são capazes de se ligar a locais não específicos de ADN, o que é essencial para o reconhecimento eficiente do sítio alvo. Nestas ligações (não específicas) não ocorre interação com as bases, mas apenas com o *backbone* do ADN. Já as ligações específicas caracterizam-se pela interação direta com as bases e indireta com o *backbone* do ADN (Dunten et al., 2008).

Dentro dos subtipos das enzimas de restrição do tipo II encontra-se o tipo IIF. As endonucleases do tipo IIF são caracterizadas por interagirem com duas cópias da sua sequência de reconhecimento, clivando ambas numa reação concertada (Halford et al., 1999)

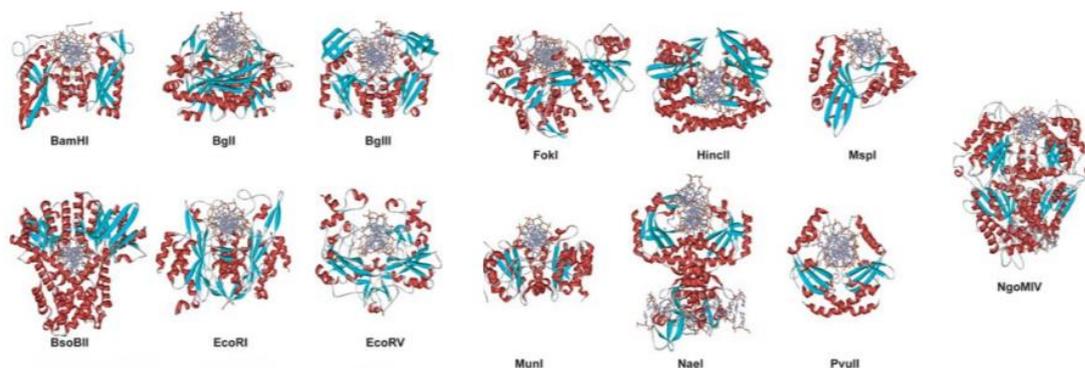


Figura 1 Representação das estruturas de diferentes tipos de endonucleases pertencentes ao tipo II das endonucleases de restrição. Adaptado de Pingoud, A., Fuxreiter, M., Pingoud, V., & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci*, 62(6), 685-707. doi:10.1007/s00018-004-4513-1.

9.2. *SgrAI*

SgrAI é uma endonucleases de restrição do tipo IIF que, corta uma sequência de reconhecimento invulgarmente longa (8 bp). Esta endonuclease é derivada do organismo *Streptomyces griseus*, fazendo parte do seu sistema imunitário contra bacteriófagos (Park

et al., 2018) Para além disso esta enzima exhibe auto modulação alostérica da atividade de clivagem do ADN e especificidade de reconhecimento da sequência (Hingorani-Varma & Bitinaite, 2003; Park, Joshi, et al., 2010). A *SgrAI* é uma endonuclease que cliva a sequência cognata, CR|CCGGYG (onde | representa o local de restrição). A sequência central (CCGG) é reconhecida por estabelecimento de interações diretas com as bases, tal como na enzima *NgoMIV* (Deibert, Grazulis, Sasnauskas, Siksnyš, & Huber, 2000). O par de bases degenerado (CR|CCGGYG) é reconhecido indiretamente por interações com os grupos fosfato do ADN, enquanto que as bases mais exteriores (CR|CCGGYG) são reconhecidas por um contato único de um resíduo arginina (Park, Joshi, et al., 2010).

Esta enzima apresenta propriedades bioquímicas pouco usuais, nomeadamente, uma clivagem mais rápida quando os plasmídeos possuem dois locais de restrição em vez de um, podendo ainda clivar locais secundários CR|CCGGY(A,C,T) e CR|CCGGGG, após clivar o local cognato (Park, Joshi, et al., 2010). Na ausência de ADN, a *SgrAI* existe sob forma de dímero. Após ligação da enzima ao ADN, e sob um conjunto específico de condições, a enzima pode formar espécies maiores que dímeros ou tetrâmeros. Para que estas HWMS se formem a concentração de *SgrAI* e do local primário (clivado ou não) tem que ser suficiente. Quando o ADN contém apenas o local secundário, não há formação de HWMS, no entanto, após a ligação a um local secundário podem-se formar HWMS na presença de um local primário (Park, Stiteler, et al., 2010).

Quando são adicionados locais primários de pré-clivados, verifica-se um aumento da taxa de clivagem do ADN pela *SgrAI*, o que sugere capacidade de modulação alostérica. A adição de ADN pré-clivado aumenta a velocidade de catálise, uma vez que induz a enzima a adotar conformações mais eficientes, tetrâmeros e HWMS (Park, Stiteler, et al., 2010). A incomum atividade de clivagem do ADN da *SgrAI* pode ser uma consequência evolutiva do grande genoma do *S. griseus*, do qual é derivado. (Dunten et al., 2008; Park, Stiteler, et al., 2010; Tautz et al., 1990).

As endonucleases de restrição são coexpressas com uma metiltransferase com a mesma especificidade de sequência, que protege o genoma do hospedeiro da atividade de clivagem da endonuclease. Assim, a metiltransferase deve metilar todas as sequências de reconhecimento *SgrAI* dentro do genoma antes de ocorrer a clivagem pela endonuclease, e este requisito pode ser difícil devido ao grande tamanho do genoma (8 milhões de pb). Consequentemente, a sequência relativamente longa reconhecida pela *SgrAI* de 8pb ao invés dos 4-6 pb usuais pode ter evoluído devido à pressão seletiva, uma vez que uma sequência mais longa reduz grandemente o número de locais que têm que ser metilados no ADN do hospedeiro. Além disso, a inerente baixa atividade da *SgrAI* na

ausência de ADN clivado no local primário também reduz a pressão seletiva no ADN do hospedeiro, bem como da metiltransferase (Bitinaite & Schildkraut, 2002; Park, Joshi, et al., 2010; Woodbury, Downey, & von Hippel, 1980). No entanto, uma sequência de reconhecimento tão longa também irá ocorrer muito menos frequentemente no ADN do fago e, portanto, colocará pressão seletiva na enzima para uma maior atividade de modo a proteger convenientemente o hospedeiro da infeção do fago. Uma das formas pela qual a atividade da enzima é aumentada é através da estimulação da sua atividade de clivagem no local secundário que induzirá mais clivagens no ADN do fago do que nos primários por si só e, portanto, poderia proteger melhor o hospedeiro. Estes locais secundários do hospedeiro são normalmente protegidos da atividade de clivagem da *SgrAI*, uma vez que são necessários locais primários clivados para induzir a clivagem dos secundários e os primários são protegidos por metilação do hospedeiro (Bitinaite & Schildkraut, 2002; Park, Joshi, et al., 2010; Woodbury et al., 1980). No entanto, a ativação da *SgrAI* pelo fago invasor pode resultar da clivagem do ADN do hospedeiro, para além do ADN do fago, nos locais secundários. Por conseguinte, a formação de agregados, ou oligómeros, de dímeros ligados aos ADN pode ter um papel importante na fixação da *SgrAI* ativado no ADN do fago e longe do ADN genómico do hospedeiro protegendo-o.

Um mecanismo que envolve trocas de domínios é proposto de forma a explicar as propriedades alostéricas pouco usuais da *SgrAI*. Os tetrâmeros que trocam de domínios quando ligados a ADN vão associar-se em oligómeros de ordem elevada. A troca de domínio implica a troca de domínios idênticos entre duas cópias da mesma cadeia polipeptídica, resultando daí a formação de um péptido mais denso (Park, Joshi, et al., 2010; Park, Stiteler, et al., 2010).

9.3. Química computacional: aplicação de métodos de mecânica molecular e mecânica quântica

A investigação científica tem sido enriquecida pelo crescente desenvolvimento de métodos computacionais, facto reconhecido pela atribuição do Nobel da química em 2013, distinguindo o desenvolvimento de modelos de multi-escala para sistemas químicos complexos (NobelPrize.org, 2013).

A crescente importância desta área tem vindo a ser suportada não só pelo desenvolvimento de computadores mais potentes e financeiramente acessíveis com

maior poder processamento computacional, bem como pelo aperfeiçoamento das ferramentas que permitem descrever os sistemas.

Os métodos utilizados na simulação de sistemas dividem-se em duas classes: os métodos empíricos (representados pela MM e baseado na Mecânica Clássica) e os métodos quânticos (representado pelos métodos baseados na função de onda - semi-empíricos e *ab initio* e DFT) (figura 2).

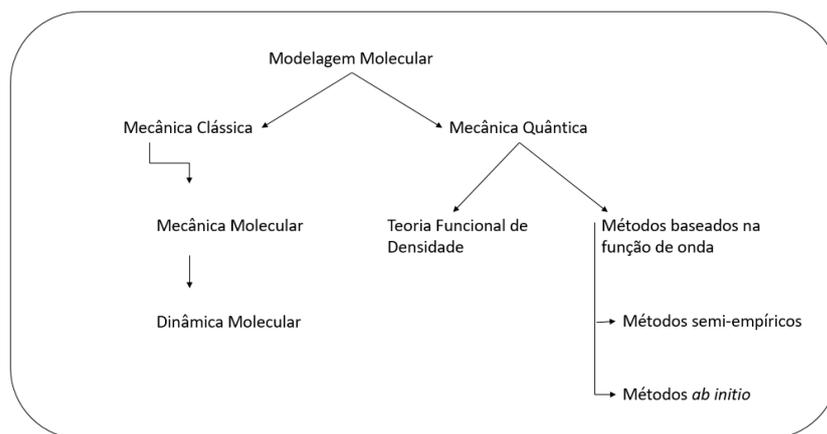


Figura 2 Representação esquemática dos métodos utilizados em química computacional

Os métodos híbridos ou de QM/MM são utilizados quando é necessário estudar reações, pois o sistema descrito pela mecânica molecular não permite a quebra e formação de ligações.

Neste caso, usufruímos da rapidez da mecânica molecular e da precisão da mecânica quântica (Vreven, Morokuma, Farkas, Schlegel, & Frisch, 2003).

9.3.1. Mecânica Molecular (MM)

MM refere-se ao uso da mecânica clássica para modelação de sistemas moleculares. Nestes métodos a superfície de energia potencial da nossa molécula é aproximada recorrendo a modelos clássicos em que a molécula é tratada como um conjunto de esferas ligadas entre si por molas.

Este método permite calcular a energia de sistemas moleculares de uma forma relativamente rápida e simples. No entanto, existem parâmetros a serem previamente

definidos, a saber: i) graus de liberdade atômica ou molecular explícitos no sistema, ii) campos de força que descrevem de forma mais consistente a energia do sistema em função dos graus de liberdade escolhidos, iii) amostragem dos graus de liberdade e iv) modelação de fronteiras do sistema e de forças externas (van Gunsteren et al., 2006) .

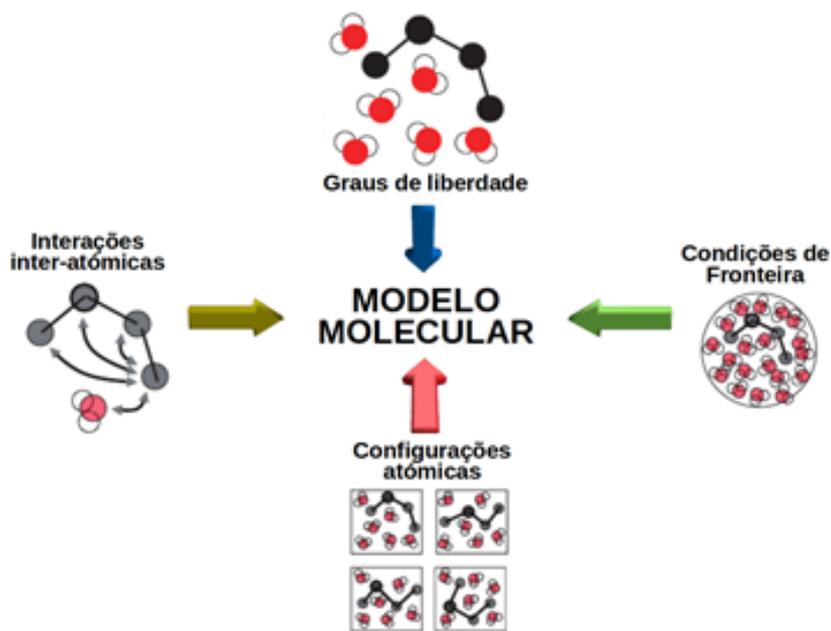


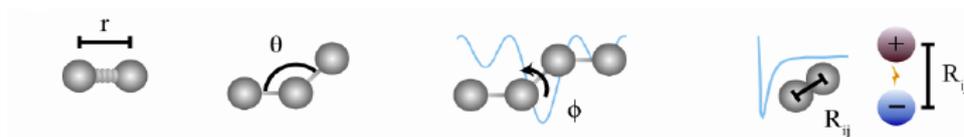
Figura 3 Representação dos parâmetros necessários para representação dos modelos moleculares. Ilustração adaptada de van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., . . . Yu, H. B. (2006). Biomolecular Modeling: Goals, Problems, Perspectives. *Angewandte Chemie International Edition*, 45(25), 4064-4092. doi:10.1002/anie.200502655.

Como referido acima, o modelo MM (figura 3) considera as moléculas como uma coleção de átomos ligados entre si por forças elásticas ou harmônicas. Estas forças são definidas por um conjunto de equações potenciais e pelos parâmetros que as integram, estando representadas numa única função de energia potencial (equação 1) e cujos componentes se caracterizam pelas diferentes variáveis, de natureza intra e intermolecular (González, 2011). O alongamento da ligação entre dois átomos, o ângulo da ligação, as torções dos ângulos impróprios ou diedros são parâmetros descritos nas interações covalentes. Já, nas interações não covalentes os termos da função definem as

interações de van der Waals, recorrendo ao potencial de Lennard-Jonnes, e a interações eletrostáticas pela lei de Coulomb (Durrant & McCammon, 2011).

Equação 1 Equação representativa da energia potencial adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71.

$$E_{tot} = \sum_{\text{ligações}} K_r (r - r_{eq})^2 + \sum_{\text{ângulos}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{diédros}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]^2 + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}^{12}} \right]$$



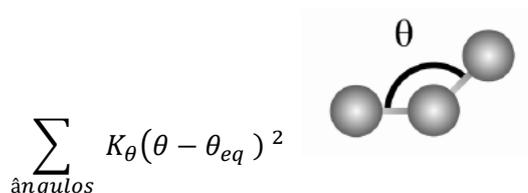
O primeiro componente da função de energia potencial corresponde à variação de energia causada pela deformação da ligação entre dois átomos (equação 2). Esta função harmónica simples controla o comprimento de ligações covalentes. Da análise deste termo verifica-se que quanto maior for a diferença entre a distância de equilíbrio e a distância num determinado momento maior será a energia (Antunes, 2016).

Equação 2 Termo que corresponde à variação de energia causada pela deformação da ligação entre dois átomos. Representação adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71.

$$E_{lig} = \sum K_r (r - r_{eq})^2$$

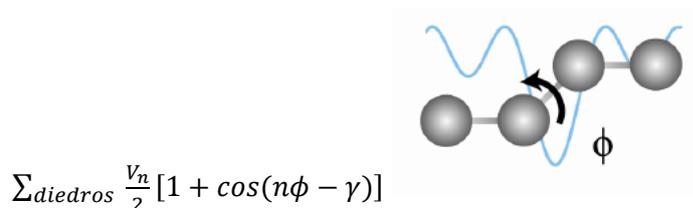
O segundo componente das relações intramoleculares corresponde aos ângulos de flexão entre átomos, considerando o ângulo instantâneo entre os átomos e o ângulo de equilíbrio (equação 3).

Equação 3 Parcela da equação representativa dos ângulos de flexão entre átomos. Representação adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71.



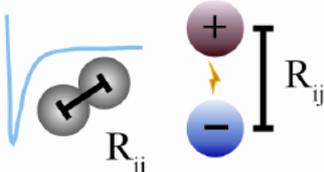
Para descrever quatro átomos em sequência devemos incluir um termo para a torção/ ângulos diedros (equação 4). Este desempenha um papel importante na determinação da estrutura da macromolécula e da sua estabilidade.

Equação 4 Parcela da equação representativa dos ângulos diedros. Representação adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71.



As interações intermoleculares são muito importantes uma vez que ocorrem entre componentes do sistema, nomeadamente interações com as moléculas do ambiente envolvente. Os primeiros termos das interações intermoleculares correspondem ao potencial de Lennard- Jones (LJ), onde se mostra a energia potencial em função da distância entre o centro de massa de duas partículas (balanço entre as forças de atração e de repulsão); o último termo das interações intermoleculares corresponde às interações electrostáticas descritas pela lei de Coulomb (equação 5).

Equação 5 Parcela da equação representativa das interações intermoleculares correspondentes às interações electrostáticas descritas pela lei de Coulomb. Representação adaptada de Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71.

$$\sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}^{12}} \right]$$


No âmbito deste trabalho para a representação do campo de forças e simulação das DM, foi utilizado o campo de forças GAFF em conjunto com o campo de forças parm99SB presente no software AMBER18. O campo de forças GAFF possui parâmetros para praticamente todas as moléculas orgânicas, complementando o AMBER, já conhecido pela sua capacidade de simular sistemas com proteínas e ácidos nucleicos (D.A. Case, 2018).

9.3.1.1. Minimização energética

A minimização energética permite otimizar o sistema usando algoritmos de forma a encontrar um estado de energia mínima. Numa minimização de energia, estes algoritmos procuram mínimos de energia locais, em relação ao ponto de partida. Uma minimização de energia não toma em conta efeitos como a temperatura logo, a trajetória dos átomos perde o sentido do ponto de vista físico, devido a isto, apenas obtemos a configuração final que corresponde a um mínimo local da energia potencial do sistema.

Para ter acesso às diferentes conformações associadas a pontos mínimos da superfície de energia, é possível recorrer a alguns programas de minimização. Um programa capaz de realizar minimização de energia, é o SANDER presente no software AMBER. Este é um programa básico de minimização de energia e dinâmica molecular, que suporta vários tipos de algoritmos de minimização entre eles o método *steepest descent* e o gradiente conjugado (D.A. Case, 2018).

9.3.1.2. Dinâmica Molecular (DM)

A DM clássica é um método de simulação computacional que estuda os movimentos físico dos átomos e moléculas usualmente em solvente, das quais se aproxima o potencial de interação entre estas partículas e calcula as equações que regem o seu movimento por métodos clássicos (Phillips et al., 2005). Nestas simulações existe interação entre os átomos e moléculas num intervalo de tempo, permitindo observar a evolução dinâmica do sistema. As trajetórias são descritas por equações de movimento de Newton, enquanto que as forças entre as partículas e as suas energias potenciais são calculadas utilizando potenciais interatômicos ou campos de força descritos através de métodos de mecânica molecular (Alder & Wainwright, 1959) .

De forma geral, este modelo clássico pode ser representado pela equação 6 onde k_B é a constante de Boltzmann, m_i e v_i representam a massa e velocidade do átomo i e N é o número de átomos envolvidos no sistema (Fuxreiter, 2014).

Equação 6 Equação representativa do modelo clássico utilizados na DM, equação retirada de Fuxreiter, M. (2014). *Computational Approaches to Protein Dynamics: From Quantum to Coarse-Grained Methods*. Boca Raton: Taylor & Francis Group.

$$3k_B T = \sum_{i=1}^N m_i v_i \cdot v_i / N$$

Estes métodos permitiram a percepção de que a dinâmica molecular das proteínas é essencial à sua função biológica e metabolismo, abrindo portas para o avanço nos estudos da Enzimologia (Carvalho et al., 2014).

9.3.1.2.1. *Ensembles*

O estado termodinâmico de um sistema pode ser caracterizado pela especificação de funções como o volume, a temperatura e a pressão.

O termo ensemble corresponde a um conjunto de microestados, gerados nas simulações de DM que estão sujeitos às mesmas restrições termodinâmicas. Existem vários ensembles: (i) *ensemble* canónico (NVT) mantém-se constante, o número de partículas (N), o volume (V) e a temperatura (T); (ii) o *ensemble* microcanónico (NVE), são mantidas constantes o número de partículas (N), o volume (V) e a energia total (E); (iii) o isotérmico-isobárico (NTP), são constantes o número de partículas (N) a temperatura (T) e a pressão (P); (iv) e o grande canónico (μVT), em que se mantém constante o potencial químico (μ) o volume (V) e a temperatura (T). Com as funções de estado constantes, as restantes propriedades do sistema podem ser calculadas através das relações fundamentais da termodinâmica (Andersen, 1980; D.A. Case, 2018; Ramos, 2012).

9.3.1.2.2. Tempo de integração

Para o sucesso da simulação de DM é crucial a seleção do tempo de integração. A escolha de um tempo de integração curto implica um aumento do tempo computacional necessário, um tempo de integração longo pode levar a uma incorreta integração das equações de movimento. Assim, é muito importante encontrar o equilíbrio entre a exatidão e a eficiência. O equilíbrio estabelecido para as biomoléculas é 1 fentossegundo (fs), que corresponde a um valor dez vezes inferior à escala temporal do movimento associado ao *stretching* de ligações que contêm átomos de hidrogénio. Não obstante, é possível utilizar tempos de integração superiores. Para isso é utilizada uma restrição das ligações que envolvam átomos de hidrogénio utilizando o algoritmo SHAKE. Com esta restrição é possível utilizar um tempo de integração de 2 fs (D.A. Case, 2018; Jean-Paul Ryckaert, 1977; Ramos, 2012).

9.3.1.2.3. Condições Periódicas de Fronteira

As simulações da DM podem ser executadas em vácuo, em solvente implícito ou em solvente explícito. Uma vez que pretendemos que a simulação seja o mais real possível, isto é, que simule o que ocorre nos sistemas biológicos, o mais indicado será a simulação em solvente explícito, neste caso em água. Quando a simulação é realizada com solvente explícito é necessário definir condições periódicas de fronteira em que as moléculas de interesse (soluto) são colocadas numa caixa de água. Esta caixa de simulação é cercada por um número infinito de réplicas de si mesma (figura 3), sendo que o sistema real simulado consiste na caixa primitiva e todas as suas réplicas.

Apenas os N átomos dentro da célula principal são considerados explicitamente, mas assim que um dos átomos sai da célula, uma partícula de imagem entra lado oposto para substituí-lo, o que minimiza o possível efeito de superfície (van Gunsteren et al., 2006).

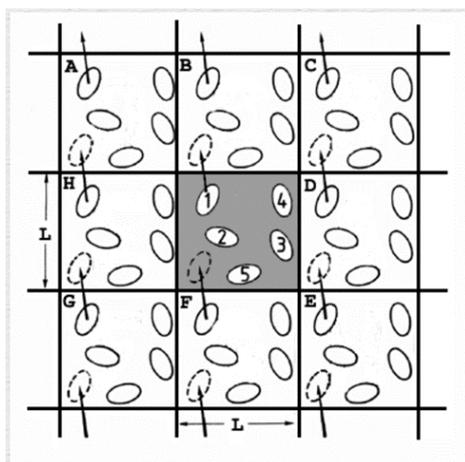


Figura 3 Esquema representativo das caixas de água. Os átomos reais estão numerados de 1 a 5. Figura retirada de Benjamin D. Madej , R. W., Sifath Mannan, Michael Barton, Tyler Luchko. (2015). *An Introduction to Molecular Dynamics Simulations using AMBER*.

Ao implementar as condições periódicas de fronteira, originam-se outros problemas, como por exemplo, o número de interações intermoleculares aumenta infinitamente. A convenção mínima da imagem estipula também que nenhum átomo deve interagir com o soluto (por exemplo uma macromolécula) e a imagem periódica desta.

Uma solução para ultrapassar esses problemas é considerar uma esfera imaginária com centro em cada átomo e com um raio pré-determinado (designado por *cut-off*), onde as energias das interações intermoleculares, com átomos fora desse *cut-off*, são consideradas zero. O *cut-off* utilizado neste trabalho foi de 10 Å, valor recomendado para as simulações de DM em solvente explícito água. No caso das interações de longo alcance, como as interações eletrostáticas, o *cut-off* não é eficiente. Nesse caso utiliza-se o método Particle Mesh Ewald (PME). O método PME considera todas as interações entre cargas dentro do sistema e as interações entre as cargas e as respectivas imagens. (Cheatham, T.E., I. et al., 1995)

9.3.2. Métodos híbridos (QM/MM)

Apesar de cada vez mais os métodos computacionais serem utilizados no estudo de grandes sistemas biomoleculares, principalmente com recurso a simulações de DM pela mecânica molecular, estes tornam-se limitados no que toca a descrever alguns processos que ocorrem no sistema tais como a quebra/formação de ligações e transferência de cargas. Assim, torna-se necessário recorrer à QM, embora se torne computacionalmente mais exigente para sistemas mais complexos (Queirós, 2016). De forma a usufruir da precisão da QM e da velocidade de computação da MM, Warshel & Levitt (1976) apresentaram o método híbrido QM/MM (Warshel & Levitt, 1976).

Os métodos QM/MM consistem na divisão do sistema em duas partes, sendo cada uma das partes descrita com diferentes níveis de teoria. A região de interesse no sistema em estudo, também designada de camada alta, é descrita usando QM o que permite o estudo pormenorizado de todas as reações químicas. O restante sistema é descrito com mecânica molecular (MM), que permite obter algum conhecimento sobre a contribuição deste, na região de interesse (figura 4) (Durrant & McCammon, 2011).

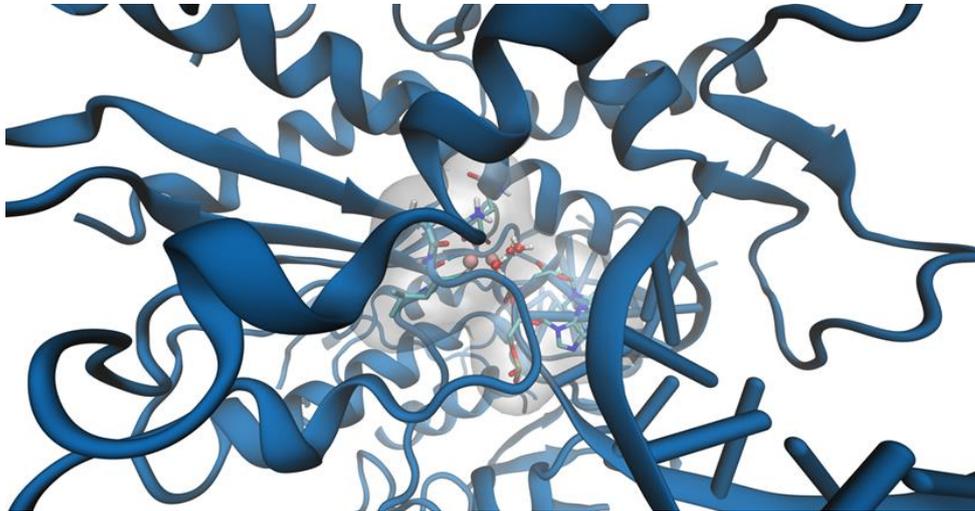


Figura 4 Ilustração do princípio do método híbrido QM/MM com divisão do sistema em duas partes. Na figura temos representada a parte do sistema *SgrAI&ADN* onde foi o utilizado QM nos resíduos representados dentro da “nuvem” a *licorice* e *CPK* e o restante sistema tratado apenas com MM representado com *New Cartoon*.

Este método híbrido é representado através de uma simples equação (equação 7) onde H_{QM} e H_{MM} representam a fórmula Hamiltoniana para as regiões de Mecânica Quântica e Molecular, respetivamente, e $H_{QM/MM}$ representa a interação entre as duas regiões (Fuxreiter, 2014).

Equação 7 Representação da equação que descreve o método híbrido QM/MM descrito por Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of Lys ozyme. *J Mol Biol*, 103(2), 227-249. doi:10.1016/0022-2836(76)90311-9.

$$H = H_{QM} + H_{QM/MM} + H_{MM}$$

10. Métodos Computacionais

Recursos computacionais e software

Todos os cálculos para as simulações bem como as dinâmicas e análises gráficas foram realizados em workstations com placas gráficas e com o sistema operativo Linux.

Para as simulações de DM da enzima *SgrAI* foi usado o software Amber Molecular Dynamics Package versão 18. AMBER. O software AMBER é constituído por um conjunto de programas que permite simulações de DM (D.A. Case, 2018).

A construção dos complexos (proteína e ligante) a serem estudados, foi feita usando o software VMD. Este programa foi projetado para ser capaz de modelar, visualizar e analisar sistemas biológicos. Neste estudo, foi usado para analisar a trajetória das simulações de DM e, posteriormente construir as esferas que seriam estudadas pelo método híbrido QM/MM.

Métodos

Neste capítulo serão explicadas passo a passo as metodologias e protocolos da simulação computacional dos diferentes sistemas em estudo, nomeadamente, a enzima *SgrAI* em forma de dímero, mutada e em forma de tetrâmero.

Com o objetivo de analisar o comportamento dinâmico da proteína e a sua interação com a molécula de ADN ao longo do tempo, foram realizadas simulações de DM através do software AMBER18 (D.A. Case, 2018) no sistema operativo Linux. Recorrendo ao mesmo software e após as DM foi aplicado o QM/MM com o objetivo de estudar a reação de catálise.

A metodologia aplicada no estudo da endonuclease *SgrAI* é descrito sumariamente na figura 5.

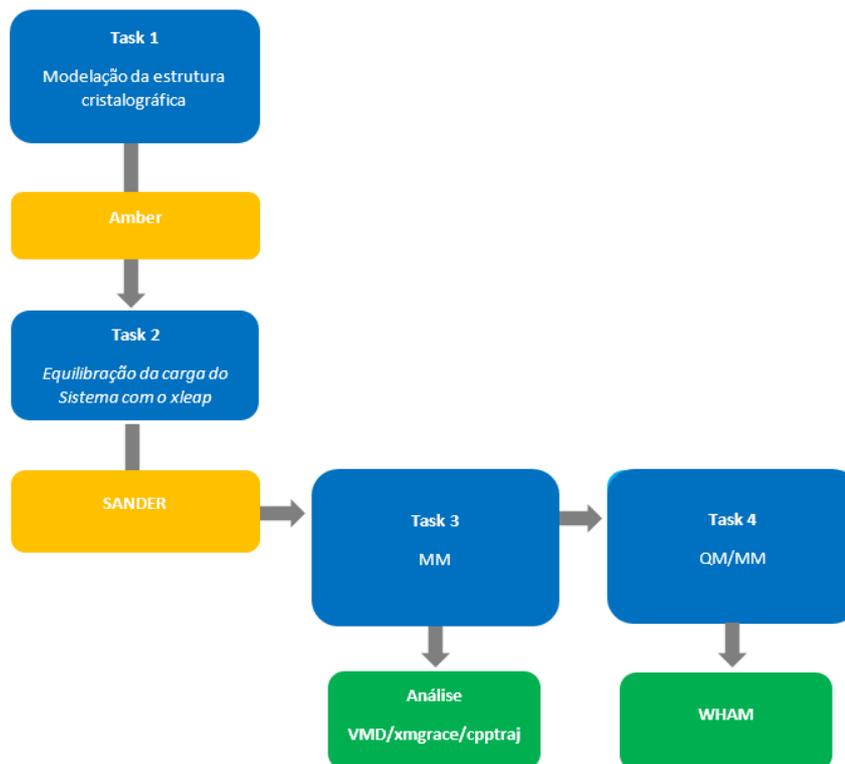


Figura 5 Workflow da metodologia aplicada no estudo da endonuclease *SgrAI*.

9.1 Preparação e solvatação

O primeiro passo foi o download da estrutura *SgrAI* tanto em forma de dímero - 3DW9 (Cristal de *SgrAI* com ADN cognato e Manganês) (Dunten et al., 2008), como em formato tetrâmero – 6MQ6 (Cristal em forma de tetrâmero com ADN e iões de cálcio) (Park, Joshi, et al., 2010) da plataforma Protein Data Bank.

Os requisitos para a utilização de ficheiros, com formato PDB, no AMBER são: nome dos átomos, nome dos resíduos, identificação das cadeias – caso haja mais que uma- e, as coordenadas dos átomos. Deste modo, é crucial que se analise o ficheiro PDB e se adeque o mesmo. Para além disso foram feitas algumas alterações no PDB uma vez que, segundo a literatura, o cofator eficaz na catálise utilizado pela endonucleases *SgrAI* é o Mg^{2+} e não o cálcio (Ca^{2+}) ou manganês (Mn^{2+}). Estes iões são utilizados nos cristais para

evitar que haja catálise aquando da formação do cristal. Assim, no PDB os iões de Mn^{2+} no dímero e de Ca^{2+} no tetrâmero foram substituídos por Mg^{2+} .

LEaP é a denominação dada aos programas *tleap* e *xleap*, presentes no software AMBER. Através destes, é possível ler inputs contento o conjunto de parâmetros, ler e editar ficheiros com formato PDB, interligar resíduos e criar complexos, modificar as coordenadas internas de uma molécula e criar ficheiros de topologia, coordenadas e parâmetros compatíveis com o software AMBER. Assim, este programa consegue combinar e executar quatro módulos (*prep*, *link*, *edit* e *parm*) necessários para a preparação do sistema (D.A. Case, 2018).

Como responsável pela preparação do input - *LEaP*- o programa irá analisar o ficheiro PDB, verificando os nomes dos átomos e resíduos. Se as denominações não forem reconhecidas pelo *LEaP*, este irá criar uns novos, de modo a que não haja perda de nenhum átomo na proteína. No entanto, se esses novos resíduos não constarem na base de dados do *LEaP*, as cargas e tipologia dos mesmos serão tratados como desconhecidos.

É neste processo que são indicados os campos de forças a serem usados de forma a descrever a energia potencial da molécula. Após vários testes, nomeadamente com o campo de forças *ff14SB*, os que provaram ser mais eficientes para este trabalho foram o *leaprc.gaff*, preparado para as moléculas em geral e o *leaprc.ff99SB* (D.A. Case, 2018) indicado para proteínas, nucleótidos e iões.

Para além dos campos de forças foi dada indicação ao programa para adicionar iões sódio (Na^+) até neutralizar a carga de todo o sistema visto que o ADN tem carga negativa.

Usando o comando *solvateoct* criou-se uma caixa de solvatação. Com isto cria-se uma caixa de água octaédrica truncada de solvente à volta da estrutura, tratada como sendo o soluto. Assim, usou-se o modelo de água *TIP3P* e, especificou-se também que a distância entre o solvente e o limite da caixa seria de 10 Å.

A preparação dá-se como finalizada gerando os ficheiros com a sua topologia (*prm*) e as suas coordenadas (*rst*), recorrendo ao comando *SaveAmberParm* (D.A. Case, 2018).

9.2 SANDER

SANDER é um módulo do software AMBER de minimização energética do sistema e de DM. Este programa otimiza a energia da estrutura por iterações do movimento dos átomos, na procura de um mínimo de energia. As estruturas, por norma, são sujeitas a minimizações antes de simulações DM. Assim sendo, este programa é uma excelente opção, uma vez que é capaz de realizar uma minimização energética antes de iniciar as simulações DM.

Uma versão otimizada do SANDER que é usada normalmente, é a PMEMD que foi criado com o objetivo de melhorar o desempenho dos métodos utilizados no SANDER (D.A. Case, 2018). O método pmemd é realmente vantajoso no tratamento de forças em sistemas macromoleculares. É um método que confere alta precisão, pode ser implementado de forma eficiente em algoritmos DM convencionais como AMBER, as funções são contínuas (evitando problemas com integração de funções descontínuas) e é eficiente uma vez que se trata de um método rápido para grandes sistemas macromoleculares.

No pmemd é possível ainda, adicionar um acelerador de GPU, criando a versão pmemd.cuda (Darden, York, & Pedersen, 1993).

9.2.1 Minimização

Como dito anteriormente, a minimização tem como objetivo eliminar pontos de contacto indesejáveis, tais como interações que ocorrem entre átomos da mesma molécula, havendo sobreposição das nuvens eletrónicas, tendo como consequência afetar a conformação final da molécula. Na tabela 1 estão os inputs utilizados nas duas fases de minimização. Min1.in com o soluto fixo e min2.in em que é minimizado todo o sistema.

Tabela 1 Representação dos *inputs* utilizados para a minimização do sistema.

min1.in	min2.in
<pre> protein 10-mer: initial minimisation solvent + ions &cntrl imin = 1, maxcyc = 1000, ncyc = 500, ntb = 1, ntr = 1, cut = 10.0 / Hold the DNA fixed 500.0 RES 1 20 END END </pre>	<pre> protein 10-mer: initial minimisation whole system &cntrl imin = 1, maxcyc = 2500, ncyc = 1000, ntb = 1, ntr = 0, cut = 10.0 / </pre>

Na primeira fase de minimização usamos como ficheiros *input* o ficheiro prm com parâmetros e topologia do sistema, o ficheiro rst com as coordenadas e o min1.in em que damos indicação para realizar 1000 etapas de minimização, 500 passos de *steepest descent* e mais 500 passos de gradiente conjugado, limites de volume constante, restrição no soluto e o *cut-off* de 10 Å. Daqui é gerado o ficheiro output (ncrst) com as velocidades e coordenadas do sistema.

Na segunda fase para além dos ficheiros prm e rst é usado o ficheiro ncrst da execução anterior, que contém a informação do primeiro estágio de minimização, e o input min2.in dá ordem para executar 2500 etapas de minimização sem as restrições. Sendo assim minimizada a energia de todo o sistema.

9.2.2 Aquecimento

Após a fase da minimização de energia ter finalizado, a próxima etapa do protocolo de equilíbrio é fazer o aquecimento do sistema dos 0K aos 300K inicialmente com restrição no soluto.

Para dar todas as indicações utilizou-se o input md1.in representado na tabela 2.

Tabela 2 Representação do *input* utilizado para o aquecimento do sistema para simulação da DM de 0 K a 300K.

```
md1.in

protein 10-mer: 20ps MD with res on DNA
&cntrl
  imin = 0,
  irest = 0,
  ntx = 1,
  ntb = 1,
  cut = 10.0,
  ntr = 1,
  ntc = 2,
  ntf = 2,
  tempi = 0.0,
  temp0 = 300.0,
  ntt = 3,
  gamma_ln = 1.0,
  nstlim = 10000, dt = 0.002,
  ntpr = 100, ntwx = 100, ntwr = 1000
/
Keep DNA fixed with weak restraints
10.0
RES 1 20
END
END
```

A esta altura a minimização é desativada ($imin=0$) e geramos velocidades iniciais aleatórias a partir de uma distribuição de Boltzmann e apenas lemos nas coordenadas do rst IREST = 0, NTX =1. Estes valores podem ser alterados para indicar que queremos recomeçar uma dinâmica molecular a partir de onde paramos. NTB = 1 indica que são usados limites periódicos de volume constante (o PME está sempre "on" quando o NTB > 0). É usado um limite de 10 Å e NTR = 1 indica restrições de posição com base na entrada GROUP fornecida no arquivo de entrada. Neste caso, vamos restringir o ADN com uma força constante de 10 Å.

NTC = 2, NTF = 2 dá indicação para usar o algoritmo SHAKE de forma a restringir as ligações que envolvem hidrogênio.

A simulação inicia-se com uma temperatura de 0 K e irá subir até aos 300 K. De forma a simular um sistema biológico real. A dinâmica de Langevin deve ser usada para controlar a temperatura usando uma frequência de colisão de 1,0 ps⁻¹ (NTT = 3, GAMMA_LN = 1.0).

No total são executados 10.000 passos de dinâmica molecular com um intervalo de tempo de 2 fs por etapa, para obter um tempo total de simulação de 20 ps. A cada 100 passos (200fs) é gravado um ficheiro de saída (NTPR), a cada 100 passos é gravado no arquivo da trajetória (NTWX) e no caso deste trabalho parar será reiniciado a cada 1.000 passos.

9.2.3 Equilíbrio

Feito o aquecimento com volume constante a próxima etapa é usar pressão constante e remover as restrições. Assim o *input* utilizado está descrito na tabela 3.

Tabela 3 Representação do *input* utilizado para simulação da DM.

md2.in
<pre> protein 10-mer: 100ps MD &cntrl imin = 0, irest = 1, ntx = 7, ntb = 2, pres0 = 1.0, ntp = 1, taup = 2.0, cut = 10.0, ntr = 0, ntc = 2, ntf = 2, tempi = 300.0, temp0 = 300.0, ntt = 3, gamma_ln = 1.0, nstlim = 25000000, T = 0.002, ntpr = 500, ntwx = 500, ntwr = 1000 / </pre>

A simulação DM será reiniciada de onde paramos, IREST informa que queremos reiniciar nos 20ps e, as coordenadas, velocidades e informações da caixa serão lidas a partir de um arquivo rst (NTX). É utilizado um limite periódico de pressão constante com uma pressão média de 1 atm. A pressão é mantida (NTP = 1) e usa-se um tempo de relaxamento de 2 ps(TAUP = 2,0). Não são usadas restrições, apenas se utiliza o SHAKE

para os átomos de hidrogénio. Nesta fase a temperatura é mantida nos 300K e são executados 2500000 passos de DM com intervalos de 2fs para um total de simulação de 5000ps ou 5 ns. Como forma de apresentar resultados fidedignos esta simulação é feita 10 vezes em cada sistema bem como são utilizadas réplicas por forma a confirmar e comprovar todos os resultados com simulações de 50 ns cada.

9.3 Análise

A primeira análise é feita pelo *xmgrace*, uma ferramenta que permite fazer gráficos bidimensionais de dados numéricos. Esta ferramenta possui uma interface gráfica que permite obter-se gráficos representativos das variações das principais propriedades do sistema: Energia Potencial, Cinética e Total, Temperatura, Pressão, Volume, Densidade, RMSD utilizando os ficheiros output e os dados da DM.

O cálculo do desvio médio quadrático (RMSD) da distância média entre os átomos das diferentes conformações da proteína resultantes da DM, e a obtenção de imagens instantâneas do sistema obtidas em intervalos de tempo regulares foi auferida pelo *cpptraj*. O *cpptraj* é um programa versátil, incluído no AMBER, utilizado no pós-processamento da trajetória obtida com DM e para análise de dados. Para além de ser capaz de processar ficheiros de trajetórias com diferentes topologias, ao mesmo tempo, é também capaz de extrair coordenadas, calcular valores de ligações, ângulos e diedros, variação de posições atómicas, etc. (D.A. Case, 2018).

A cada etapa recorreu-se ao VMD para visualizar a dinâmica e as transformações e verificar se tudo se apresentava como esperado.

9.4 Mutações

De forma a sustentar a especificidade de uma ligação e a sua importância é criada uma mutação num par de bases (com maior interação com os resíduos da proteína) e voltar a fazer DM. No caso, através da análise pelo *cpptraj* das distâncias das interações entre a proteína e as bases do ADN, verificou-se que havia uma grande interação da proteína, especificamente o aminoácido Arg com o nucleótido DC. Através do programa *Chimera* mudou-se o resíduo pretendido (DC) para uma DA. Após DM da molécula

mutada procedeu-se à análise da interação entre resíduos de forma a perceber que implicações resultaram da mutação.

9.5 QM/MM

Por forma a estudar a reação de catálise utilizamos o método QM/MM. Assim, delimitou-se a zona onde ocorre a reação através da identificação dos resíduos que fazem parte e contribuem para a mesma.

Esta parte do sistema, também designada como camada alta, é tratada com QM utilizando o método semiempírico *PM6*, o restante sistema é tratado com MM. Os métodos semiempíricos são versões simplificadas da teoria de Hartree-Fock usando correções empíricas que possibilitam a redução do custo computacional sem comprometer os resultados permitindo que moléculas complexas possam ser modeladas de maneira realista (Martins, L *et al.*, 2016).

De modo a estudar a reação foi definida uma coordenada reacional entre água a nucleófila e o grupo fosfato do nucleótido. Foram testadas várias hipóteses com diferentes constantes de força, tal como é descrito nos resultados.

10 Resultados e Discussão

Os modelos utilizados no estudo da reação catalítica da endonucleases *SgrAI* foram os seguintes: a enzima em forma de dímero proveniente do PDB com o ID 3DW9 com ADN cognato e utilizando o Mg^{2+} como cofator; enzima em forma de dímero proveniente do mesmo PDB com o ID 3DW9 com ADN cognato com mutação no resíduo 696 e utilizando o Mg^{2+} como cofator; enzima na forma de tetrâmero com ADN cognato e íões de Ca^{2+} originalmente proveniente do PDB com o ID 3MQ6; e enzima na forma de tetrâmero com o ID 3MQ6 com ADN cognato e tendo Mg^{2+} como cofator subdividido em 2 modelos: com a Lys neutra e com a Lys protonada.

O estudo feito com as estruturas do dímero, quer na conformação original quer na forma mutada e no modelo contendo a estrutura em tetrâmero tendo os íões de Ca^{2+} como cofator, apenas foram tratados por métodos de MM. Os resultados obtidos não foram favoráveis, o que impediu a sua progressão para ensaios que incluíssem métodos QM.

De acordo com a literatura (Park, Joshi, et al., 2010), a enzima *SgrAI* sob forma de tetrâmero e com o cofator Mg^{2+} (ID 3MQ6) apresenta uma maior atividade catalítica, eficiência e funcionalidade. Por este motivo, o estudo prosseguiu incidindo especialmente sobre a conformação tetramérica da *SgrAI*, que se apresentava como mais promissora.

A *SgrAI* é uma enzima pouco abordada e caracterizada pela literatura e, portanto, o seu mecanismo ainda está muito pouco caracterizado. Contudo, podemos ter em conta que esta enzima se insere no grupo II e família F das endonucleases de restrição para a descrever um pouco melhor. Segundo a literatura (Bitinaite & Schildkraut, 2002; Dunten et al., 2008; Park, Joshi, et al., 2010), a conformação e o comportamento da enzima *SgrAI* tem muita similaridade com enzimas da mesma família, como a *NgoMIV* e *Cfr10I*. Assim, tomamos como ponto de partida o mecanismo destas enzimas e a forma como os resíduos envolvidos direta e indiretamente influenciam os centros ativos, de forma a desencadear as mudanças conformacionais necessárias para a formação do complexo pré-Michaelis ou para a permitir a sua preorganização para a estabilização dos estados de transição, de modo que possa ocorrer clivagem da ligação fosfodiéster.

O mecanismo usado pelas endonucleases do tipo II é caracterizado pela inversão da configuração do fosfato (Connolly, Eckstein, & Pingoud, 1984; Pingoud, Fuxreiter, Pingoud, & Wende, 2005). De forma geral, o mecanismo de hidrólise do fosfodiéster passa por três etapas: a) a preparação do nucleófilo atacante por desprotonação (é necessária uma base para desprotonar a molécula de água que faz o ataque catalítico), b) o ataque nucleófilo ao fosfato levando ao estado de transição (é necessário um ácido de forma a

estabilizar o estado de transição com duas cargas negativas) e, c) a saída do grupo 3'-hidroxil (é necessário um ácido para protonar o grupo de saída) (Pingoud et al., 2005).

O Mg^{2+} é o metal divalente utilizado como cofator iónico e geralmente a Lys é sugerida como base, no entanto, pode ser substituída por outro resíduo. (Pingoud et al., 2005; Pingoud & Jeltsch, 2001)

Os modelos do mecanismo de clivagem de ADN por endonucleases de restrição são baseados no número variável de iões metálicos envolvidos na reação. Assim, com base na literatura e na filogenia e semelhança estrutural entre enzimas, o estudo foi baseado no mecanismo com dois iões metálicos (Deibert et al., 2000; Dunten et al., 2008).

Os dois iões metálicos são necessários para estabilizar o estado de transição pentavalente e, por isso, estes devem estar dispostos de forma a que um dos iões Mg^{2+} interaja com o oxigénio da água nucleófila, sendo responsável por baixar o pKa da molécula de água de forma a facilitar a sua desprotonação, enquanto o outro interage com o óxido do grupo de saída (Yang, Lee, & Nowotny, 2006). Idealmente estes iões deveriam distar aproximadamente 4 Å um do outro, com o intuito de reduzir a repulsão eletrostática do estado de transição (figura 6) (Deibert et al., 2000; Dunten et al., 2008; Pingoud et al., 2005).

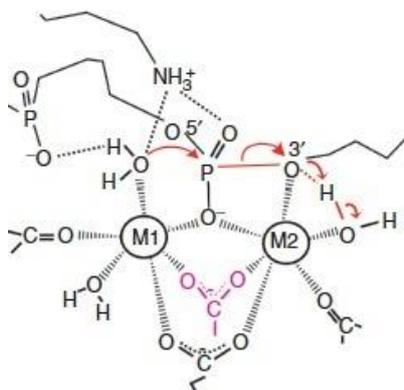


Figura 6 Modelo do mecanismo da clivagem do ADN na presença de dois iões metálicos, identificados pelas posições M1 e M2 que devem ser ocupadas pelos iões Mg^{2+} . As setas a vermelho identificam as alterações nas ligações, representando de forma esquemática o mecanismo catalítico esperado. Imagem adaptada de (Dunten et al., 2008).

Nenhuma das estruturas cristalográficas presentes na base de dados do Protein Data Bank (Berman et al., 2000) (Banco de Dados de Proteína, PDB) apresentava esta disposição dos iões metálicos no centro ativo (Dunten et al., 2008; Park, Joshi, et al., 2010). Tal deve-se ao facto de todos os cristais terem sido obtidos com iões divalentes

Ca^{2+} ou Mn^{2+} em vez de Mg^{2+} de modo a prevenir a clivagem do ADN (Dunten et al., 2008; Park, Joshi, et al., 2010). Consequentemente, foi necessário ajustar os sistemas para esta conformação, que passaremos a designar por mecanismo catalítico dos 2-iões-metálicos (em inglês two-metal-ion catalytic mechanism), uma vez que as estruturas cristalográficas não retratam a realidade funcional das enzimas. Assim, procurou encontrar-se a conformação ideal, alterando não só o cofator bem como o seu posicionamento.

A partir das estruturas com diferentes posições dos iões magnésio foram também testadas várias hipóteses para mecanismo catalítico através da variação não só da camada alta bem como da constante de força e diferentes bases para abstração do protão da água catalítica.

Nos próximos capítulos iremos analisar, de forma detalhada, todas as hipóteses testadas para os diferentes sistemas, considerando também as diferentes estruturas quaternárias e uma mutação (estruturas dimérica, dimérica mutada e estrutura tetramérica).

10.1 Dímero

Para obter o PDB com as coordenadas da enzima de interesse *SgrAI* na forma de dímero recorreu-se à base de dados PDB (Berman et al., 2000). Para esta enzima, na forma de dímero, existem os seguintes ficheiros na base de dados PDB: *SgrAI* com ADN não cognato e iões de Ca^{2+} (ID 3DPG), *SgrAI* com ADN cognato e iões de Ca^{2+} (ID 3DVO), e *SgrAI* com ADN cognato e iões de Mn^{2+} (ID 3DW9) (Dunten et al., 2008).

Após análise, o PDB com ID 3DW9 (figura 7), referente a *SgrAI* com ADN cognato e Mn^{2+} , foi selecionado uma vez que esta estrutura é a mais próxima da enzima funcional, isto é, com o ADN cognato e na presença dos iões Mg^{2+} . Tal como referido anteriormente, não é possível obter a estrutura cristalográfica da enzima funcional porque nas condições experimentais que permitem a formação deste complexo para posterior cristalização o ADN é clivado de imediato. Assim, são utilizados os iões metálicos divalentes Ca^{2+} ou Mn^{2+} como substituição dos Mg^{2+} , de modo manter a integridade do complexo *SgrAI*//ADN pois permite que as interações específicas se formem mas não há clivagem de ADN (Dunten et al., 2008).

A estrutura cristalográfica foi modificada por adição dos átomos de hidrogénio de acordo com estados de protonação e por modificação do Mn^{2+} para Mg^{2+} , já que o principal objetivo deste trabalho é o estudo da reação de catálise a qual depende das condições

ideais, ou seja, a presença da sequência de ADN cognata e de Mg^{2+} como cofator, figura 13 (Dunten et al., 2008).

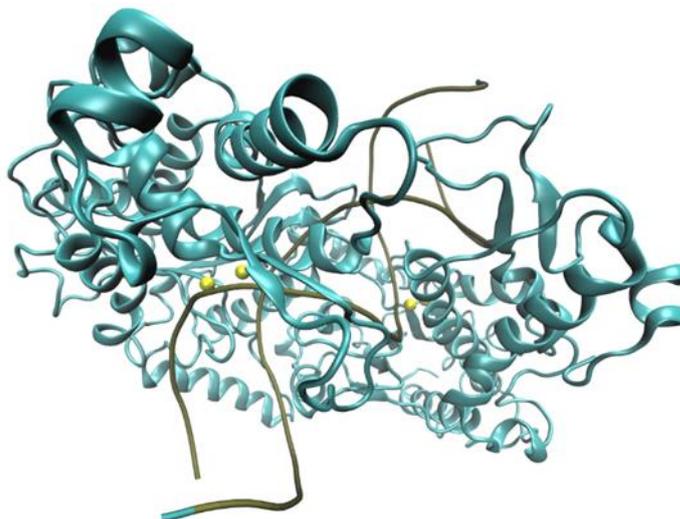


Figura 7 Ilustração do dímero da enzima *SgrAI* representado com *New Cartoon*, o ADN cognato representado com *Tube* e os iões de Mg^{2+} como cofator representados como esferas amarelas.

Após o tratamento do PDB com o *xleap* e minimização, equilíbrio e simulações de DM, foi feita a análise dos ficheiros de output com a informação das variáveis (temperatura, pressão e energia) de forma a garantir que o processo decorria em condições de equilíbrio (Anexo I).

O tempo total de cada simulação foi de 50 ns. De modo a obter uma estrutura de referência para cada trajetória foram realizadas análises de clusters (Ahmad, Keskin, Sarai, & Nussinov, 2008). *Clustering* ou agrupamento em dinâmica molecular é um modo de análise da amostragem conformacional de uma trajetória pelo agrupamento de conformações similares. As conformações foram classificadas de acordo com o RMSD (root-mean-square deviation), uma métrica de distância comumente usada. Este é capaz de monitorizar quantitativamente o comportamento da proteína, ao longo da simulação, em relação à estrutura inicial. É calculado através da seguinte equação:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n di^2}$$

correspondendo à média de n pares de átomos, em que di é a distância entre os dois átomos no par i .

Numa primeira abordagem foi utilizado o RMSD dos átomos do *backbone* do ADN (P, O3',O5',C3',C4',C5'). Como se pode verificar na figura 8, o mesmo não apresenta grandes variações durante a simulação (figura 8). Å

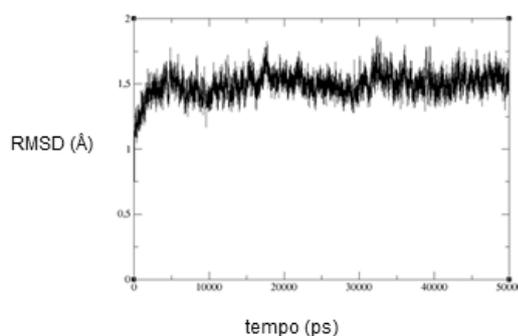


Figura 8 Figura representativa do RMSD dos átomos do *backbone* do ADN (P, O3',O5',C3',C4',C5').

Atentando à figura 8 e considerando a variação do RMSD, verifica-se que o comportamento da proteína vai estabilizando com o decorrer da dinâmica. No final da simulação é possível verificar que há uma menor flutuação do RMSD, indicando que a estrutura está equilibrada. Assim, foram selecionados os últimos 20 ns de simulação para prosseguir a análise.

Devido à baixa variação do *backbone* do ADN durante a simulação, a similaridade entre conformações foi também avaliada pelo cálculo do RMSD dos resíduos da proteína que interatuam especificamente com o ADN (figura 9), mais precisamente o Glu 102, Asp 187, a Leu 149 e Asn 148. Neste caso a variação é maior. Optámos por considerar esta medida mais representativa do comportamento deste complexo ADN-proteína ao longo do tempo de simulação e escolhe-la para o cálculo das estruturas de referência.

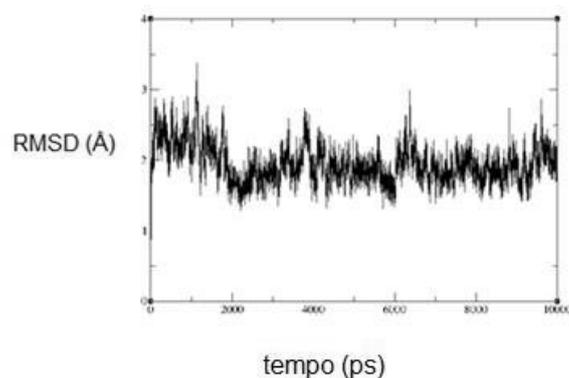


Figura 9 Figura representativa do RMSD dos resíduos de interesse da proteína *SgrAI*.

As estruturas de referência calculadas a partir do *clustering* foram utilizadas para análise das interações específicas entre os resíduos da proteína e o ADN (tabela 4). Analisando a tabela 4, os nucleótidos DC 696 (nucleótido +1) e DG 697 (+2) foram os que estabeleceram maior número de interação intermoleculares com a enzima. Fazendo a análise ao tipo de interações vemos que a Arg 248 interage com a DC 696 (+1) através de duas pontes de hidrogénio entre o segundo oxigénio do grupo fosfato do nucleótido e azoto NH1 e azoto NH2 do aminoácido. A Asn 246 forma uma ponte de hidrogénio entre o oxigénio do grupo carboxilo e o azoto N2 da base azotada da DG 698 (+3). O azoto da Leu 36 interage por ponte de hidrogénio com o oxigénio do grupo fosfato da DG 682 (-13). O azoto NE da Arg 245 forma uma ponte de hidrogénio com o N4 da DC 677 (-18). A Arg 365 interage com o nucleótido DG 700 (+5) por duas pontes de hidrogénio compreendidas entre o azoto N2 do aminoácido e o oxigénio da base do nucleótido e entre o azoto NH1 do aminoácido e o azoto N7 da ribose do nucleótido. O segundo oxigénio do grupo fosfato da DT 699 (+4) interage com a Thr 367 por pontes de hidrogénio com o grupo amina e com o oxigénio do radical do aminoácido. Por fim, os carbonos da cadeia lateral da Lys 95 encontram-se bastante próximos do azoto N2 da base azotada do nucleótido DG 680 (-15) formando interação por dipolo permanente e dipolo induzido.

Verifica-se, portanto, que uma das principais interações específicas é entre o ADN e o resíduo Arg. Este resíduo encontra-se descrito na literatura como responsável pelo reconhecimento por contacto dos nucleótidos mais exteriores da sequência cognata (CRCCGGYG) (Park, Stiteler, et al., 2010).

Tabela 4 Tabela referente à distância média das interações entre os resíduos da proteína *SgrAI* e o ADN durante a simulação.

Aminoácido	Nucleótido	Distância média (Å)
Arg 248 -NH1	DC 696 -O2P	2.75
Arg 248 -NH2	DC 696 -O2P	2.88
Asn 246 -O	DG 698-N2	2.96
Leu 36 -N	DG 682 -O1P	3.01
Arg 245 -NE	DC 677 -N4	3.05
Arg 365 -NH2	DG 700 -O	2.75
Arg 365 -NH1	DG 700 -N7	3.07
Thr 367 -OG1	DT 699 -O2P	2.83
Thr 367 -N	DT 699 -O2P	3.24

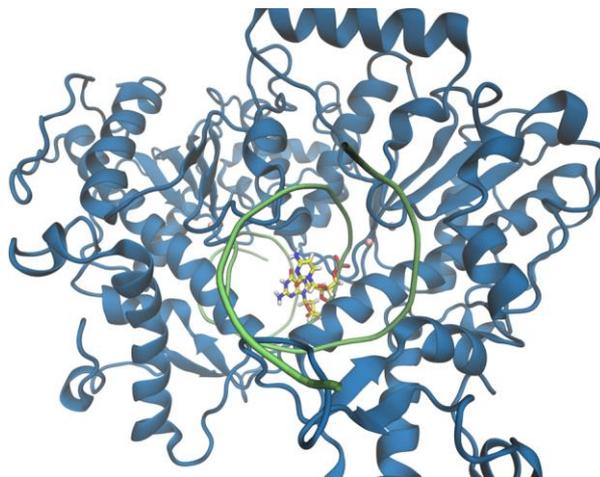


Figura 10 Representação visual dos nucleótidos DC 696 e DG 697 ao centro em *Licorice*, a proteína é representada em *New Cartoon*, a molécula de ADN em *Tube* e os íons Mg^{2+} representados como esferas de cor salmão.

Sendo o objetivo primordial deste trabalho a detalhada caracterização da atividade catalítica da enzima *SgrAI*, a atenção foca-se no centro ativo da enzima. De acordo com a literatura, e como referido anteriormente, a conformação geral da enzima e o posicionamento dos dois íons de Mg^{2+} são fulcrais no momento da catálise (Yang et al.,

a literatura, e como referido anteriormente, a conformação geral da enzima e o posicionamento dos dois íões de Mg^{2+} são fulcrais no momento da catálise (Yang et al., 2006). Tendo isto em conta, o posicionamento dos íões de Mg^{2+} e dos resíduos envolventes foram alvos de análise, de forma a verificar se a conformação da enzima era a apropriada. Tal como referido anteriormente, o mecanismo mais provável para esta enzima é a hipótese dos 2-íões-metálicos, em que um íão Mg^{2+} deverá ocupar a posição M1 sendo responsável pela ativação de uma molécula de água para o ataque ao grupo fosfato presente no carbono 5' da pentose, estabilizando o seu estado desprotonado que é mais nucleofílico do que a água. O segundo Mg^{2+} deve ocupar a posição M2 suportando o grupo de saída estabilizando a carga negativa após a clivagem (Dunten et al., 2008). Tendo a literatura em consideração, no centro ativo do sistema em estudo (figura 11) observou-se que, apesar da substituição dos íões de manganês (Mn^{2+}) por magnésio (Mg^{2+}), subsistia um mau posicionamento, não só dos magnésios, especialmente o Mg^{2+} 707, que corresponde ao M2, e que está a cerca de 7 Å do átomo O3' da DA 695 (grupo de saída, nucleótido -1), bem como como uma má conformação em geral da molécula. O Glu 102, que devia suportar o Mg^{2+} (M2), está neste caso orientado numa posição afastada do ADN.

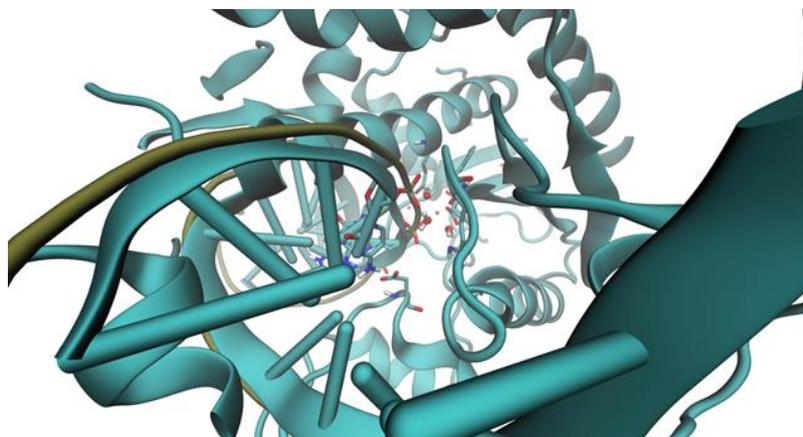


Figura 11 Representação do centro ativo do dímero da enzima *SgrAI* com *Licorice* e *CPK*, o restante sistema está representado com *New Cartoon* e *Tube*.

Os dois íões não se encontram nas posições teoricamente esperadas, posições essas que deviam permitir ao Mg^{2+} 706 (M1) estabilizar o ataque da água nucleofílica ao grupo fosfato, bem como ao Mg^{2+} 707 (M2) estabilizar o grupo de saída (figura 12).

Perante esta análise concluímos que esta estrutura não é a ideal para o estudo catalítico.

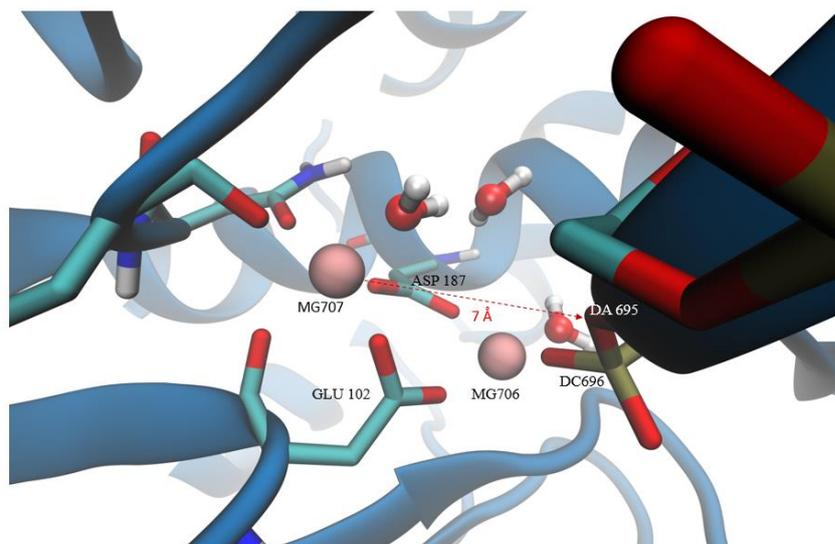


Figura 12 Representação do centro ativo da enzima e da posição incorreta dos resíduos representados em relação ao que seria esperado. Destaque para o distanciamento do Mg^{2+} MG707 ao nucleótido DA 695.

Para além disso, e tendo em conta que a enzima é mais eficiente sob forma de aglomerados, tetrâmero ou HWMS e não como dímero, avançámos com o estudo do tetrâmero da endonuclease *SgrAI*.

10.2 Dímero Mutado

A especificidade das interações entre os resíduos da proteína e o ADN foi testada ao criar uma mutação num dos pares de bases que mais interações estabelece com a proteína utilizando o programa *Chimera* (Pettersen et al., 2004). Os nucleótidos que estabelecem maior numero de interações com a proteína foram o par de bases DC 696 (+1) e DG 697 (+2). Deste modo a DC foi alterada para DT e conseqüentemente a DG para DA.

O PDB foi tratado e preparado para a DM e, antes de iniciar a mesma, verificou-se que a mutação foi introduzida corretamente. No lugar do resíduo DC 696 (+1) está agora DA 696. Esta mutação, localizada junto ao centro ativo faz com que a sequência de reconhecimento se altere de CR|CCGGYG para CR|CAGGYG.

Após verificação das variáveis (Anexo II), a análise do RMSD dos carbonos α da proteína (figura13), permite observar que existe alguma oscilação, principalmente no final

da DM, em que o valor do RMSD tende a aumentar. Uma vez que a mutação pode ter feito com que as ligações tivessem um comportamento diferente, é perfeitamente normal que a variação média dos resíduos seja maior uma vez que estes se estão a posicionar/ajustar devido às alterações.

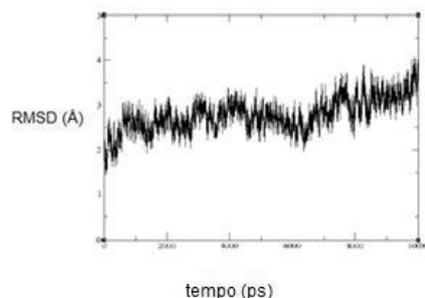


Figura 13 Representação gráfica da análise do RMSD considerando a variação média das distâncias dos carbonos α da proteína.

Na figura 14 estão representadas o par de bases mutadas DA 696 (+1) DT 679 (+2), o que confirma que a mutação ocorreu de forma eficaz. Ainda na mesma figura podemos observar os resíduos que interagem com a DA696 (+1). Ao analisar estes resíduos podemos observar que se verificam poucas alterações em relação às interações que aconteciam com a DC 696 (+1). A Arg 248 mantém a ponte de hidrogénio entre o azoto NH1 e NH2 com o segundo oxigénio do grupo fosfato do nucleótido DA 696 (+1) tal como acontecia quando DC 696 (+1). Para além disso neste modelo os carbonos da cadeia lateral deste aminoácido estão muito próximo do azoto da ribose do nucleótido DA 696 (+1). Já a Lys 241 mantém exatamente as mesmas interações após a mutação do nucleótido, fazendo uma ponte de hidrogénio entre o grupo amina (NZ) e o segundo oxigénio do grupo fosfato da DA 696 (+1).

Podemos concluir que não houve alterações significativas comparativamente com o modelo dimérico sem mutação. Uma vez que a conformação geral do complexo *SgrAI* /ADN não era a correcta a mutação não surtiu efeitos .

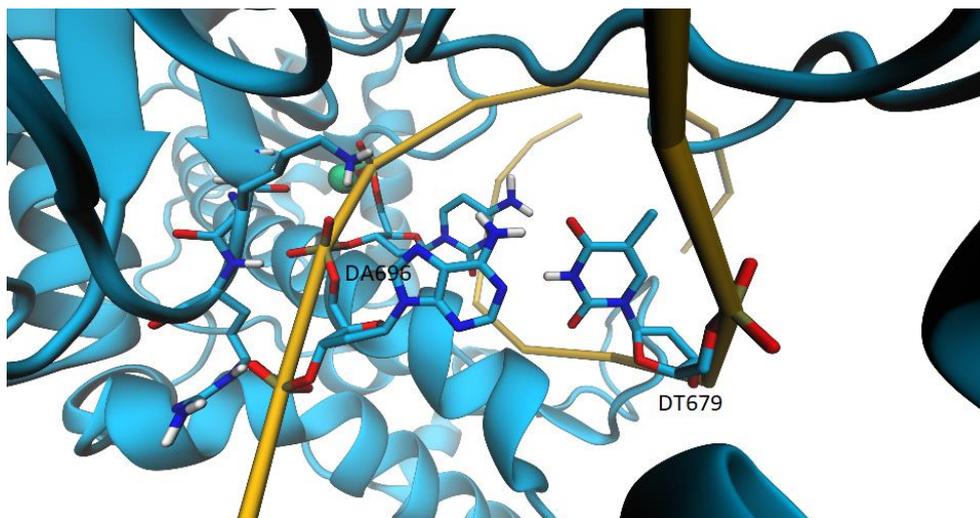


Figura 14 Figura representativa do par de bases DA 696 - DT 679 e dos resíduos que interagem com o nucleótido DA 696.

Assim, a reação de catálise não ocorre em consequência do mau posicionamento dos resíduos intervenientes, tornando a enzima não funcional cataliticamente o que não está diretamente relacionado com a mutação.

10.3 Tetrâmero da enzima *SgrAI* com iões de Ca^{2+} (TetraCa) e iões de Mg^{2+} (TetraMg)

Para modelar a estrutura tetramérica foi utilizado o PDB com ID 3MQ6 (Park, Joshi, et al., 2010). Este PDB contém dois dímeros que formam um tetrâmero ligado a duas cadeias de ADN, e três iões de Ca^{2+} ligados a cada sítio ativo. A estrutura cristalográfica foi modificada por adição dos átomos de hidrogénio de acordo com estados de protonação. Para além disso, no caso do TetraMg os iões de Ca^{2+} foram substituídos por iões de Mg^{2+} .

Com estes dois sistemas pretende-se avaliar a diferença da conformação da enzima no seu estado inativo (TetraCa) versus estado ativo (TetraMg).

10.4 Tetrâmero da enzima *SgrAI* com iões de Ca^{2+}

Numa primeira fase foi analisado o modelo com os iões Ca^{2+} , que são os metais divalentes que estão presentes originalmente na estrutura cristalográfica (Park, Joshi, et al., 2010). Tanto os iões Mg^{2+} como os iões Ca^{2+} são usados pelas endonucleases de forma a promover a ligação entre a enzima e o ADN. O Mg^{2+} confere atividade de clivagem à enzima enquanto que o Ca^{2+} inibe a clivagem, mas mantém a ligação à sequência específica (Ma et al., 2013; Vipond & Halford, 1995).

Com o objetivo de comparar as interações na presença dos iões de Ca^{2+} , a estrutura cristalográfica foi modificada adicionando os átomos de hidrogénio de acordo com estados de protonação. O PDB foi então tratado e preparado para a DM, que neste caso foi apenas de 30 ns.

Após a simulação DM verificaram-se as variáveis (Anexo III) e, segundo a análise do RMSD considerando a variação do RMSD dos carbonos α da proteína (figura 15), podemos observar que, aproximadamente aos 11000ps, existe uma forte oscilação. Contudo, no final da DM o valor RMSD tende a estabilizar.

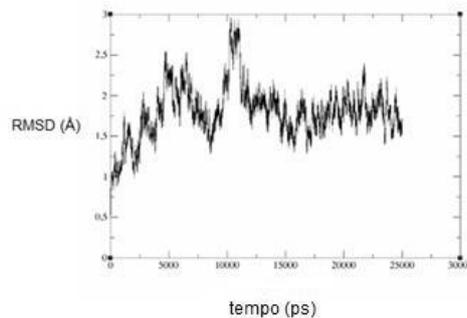


Figura 15 Representação gráfica da análise do RMSD considerando a variação média das distâncias dos resíduos de interesse.

Procedendo à análise das interações entre os resíduos proteicos e o ADN (tabela 5), este modelo volta a confirmar que, de facto, a Arg se destaca na interação com o ADN (Park, Stiteler, et al., 2010).

Tabela 5 Tabela referente à variação da distância média das interações entre os resíduos da proteína *SgrAI* e o ADN.

Aminoácido	Nucleótido	Distância média (Å)
Asn 761 -ND2	DG 1386 -O4'	3.37
Asn 761 -O	DG 1386 -N2	3.47
Arg 242 -NE	DC 1349 -N4	3.10
Arg 336 -NH2	DG 1368 -O2P	2.77
Arg 336 - NE	DG 1368 -O1P	2.91
Lys 89 - CH	DC 1366 - O2	2.85
Ser 575 -OG	DG 1351 - O2P	2.84

Atentando à tabela 5 observa-se que a Asn 761 interage com o ADN através da formação de duas pontes de hidrogénio, uma entre o azoto ND2 do aminoácido e o oxigénio 4' do nucleótido DG 1386 (+43), e outra entre o oxigénio do aminoácido e o azoto N2 do mesmo nucleótido. A Arg 242 forma uma ponte de hidrogénio com o azoto N4 da DC 1349 (0). Já a Arg 336 forma duas pontes de hidrogénio com a DG 1368. A primeira interação é entre o azoto NH2 e o segundo oxigénio do grupo fosfato do nucleótido, a segunda é entre o azoto NE do aminoácido e o primeiro oxigénio do grupo fosfato do nucleótido. Os carbonos da cadeia lateral da Lys 89 estão bastante próximos do oxigénio da base azotada do nucleótido 1366 (+17) exibindo interação dipolo permanente e dipolo induzido. Por fim, o oxigénio da Ser 575 faz interação por ponte de hidrogénio com o segundo oxigénio do grupo fosfato da DG 1351 (+2).

O estudo deste modelo tinha como objetivo a comparação da conformação do complexo *SgrAI*//ADN quando na presença dos diferentes iões (Mg^{2+} e Ca^{2+}). No próximo capítulo (complexo *SgrAI* /ADN com iões de Mg^{2+}) faremos essa análise.

10.4.1 Tetrâmero da enzima *SgrAI* com iões de Mg^{2+}

Nesta última fase foi considerado o modelo tetramérico com iões de Mg^{2+} . Neste ponto, sabendo que este é o modelo mais eficiente (Dunten et al., 2008), o trabalho focou-se no mecanismo da reação de catálise.

10.4.2 *SgrAI* tetramérica com Lys neutra

Com base na literatura o resíduo Lys é geralmente utilizado como base na reação catalítica uma vez que pode receber o protão proveniente da molécula de água nucleófila (Pingoud et al., 2005).

Portanto, a Lys do centro ativo, Lys 241, foi modificada para uma Lys neutra, por deleção de um hidrogénio do azoto da cadeia lateral e alteração das cargas pontuais dos átomos passando a designar-se Lyn241.

A simulação DM decorreu dentro do esperado, como podemos observar nas variáveis temperatura, pressão e energia do sistema (Anexo IV) e no gráfico representativo do RMSD (figura 16), onde se observa alguma oscilação sem ultrapassar um intervalo curto.

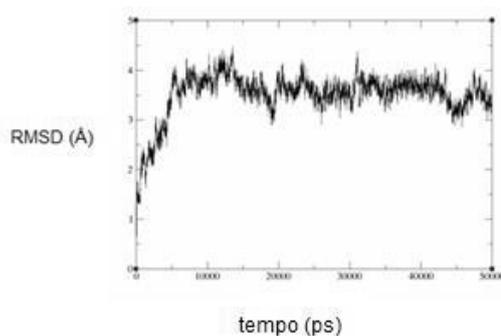


Figura 16 Representação gráfica da análise do RMSD considerando a variação média das distâncias dos carbonos α da proteína.

Na tabela 6 podemos verificar que a Arg se destaca como resíduo com mais interações com a proteína tendo as distâncias médias inferiores a 2 Å a comprovar a sua

especificidade (Pingoud et al., 2005). A amida do *backbone* da Arg 242 forma uma ponte de hidrogénio interagindo com o primeiro oxigénio do grupo fosfato do nucleótido DC 1382 (+18). Já a Arg 588 forma duas pontes de hidrogénio em que interage o azoto NE do aminoácido com o oxigénio 6 do nucleótido DC 1366 (+2) e o azoto N do aminoácido com o primeiro oxigénio do grupo fosfato do nucleótido. O azoto N da Leu 376 interage com o primeiro oxigénio do grupo fosfato do nucleótido DG 1386 (+22) através de uma ponte de hidrogénio. A Gln forma duas pontes de hidrogénio ao interagir com a DG 1404 (+40) entre do azoto N do aminoácido e o primeiro oxigénio do grupo fosfato e entre o azoto NE2 do aminoácido e o oxigénio 3' do nucleótido. Por fim, o oxigénio da Ser 583 forma uma ponte de hidrogénio com o oxigénio 3' do nucleótido DC 1365 (+1).

Após a descrição do tipo de interações feitas pela enzima *SgrAI* com a molécula de ADN observamos que estas se mantêm altamente conservadas em relação aos diferentes modelos em estudo. Prevaecem as pontes de hidrogénio e a presença da Arg que se mostra como resíduo com mais interações específicas entre a enzima e o ADN, sendo assim, e pela sua importância este resíduo torna-se altamente conservado (Dunten et al., 2008).

Tabela 6 Tabela com distância médias das interações entre os resíduos da proteína *SgrAI* e o ADN.

Aminoácido	Nucleótido	Distância média (Å)
Arg 242 -N	DC 1382 -O1P	2.76
Arg 588 -NE	DC 1366 -O6	2.82
Arg 588 -N	DC 1366 -O1P	2.76
Leu 376 -N	DG 1386 -O1P	2.82
Gln 717 -N	DG 1404 -O1P	3.04
Gln 717 -NE2	DG 1404 -O3'	3.40
Ser 583 -OG	DC 1365 -O3'	3.80

Uma vez identificados os resíduos e as suas interações, passou-se ao estudo do mecanismo por QM/MM. Os resíduos selecionados para pertencer à camada alta para serem tratados com QM foram: Asp 242, Glu 442, Asp 527, Lyn581, DC 1364 (DC com o grupo fosfato que será clivado- nucleótico 0), a DA 1363(DA do grupo de saída -nucleótico

-1), Mg^{2+1425} (M1) e Mg^{2+1426} (M2) e as águas WAT1773 (molécula de água nucleófila), WAT1593 WAT1567, WAT1594 (moléculas de água coordenadas a iões Mg^{2+}).

Para estudar a reação a coordenada reacional definida foi a distância entre o oxigénio da água nucleófila e o átomo de fósforo do grupo fosfato que é clivado (DC 1364). Uma constante de força de 150 Kcal foi aplicada ao sistema. Apesar da aproximação a molécula WAT1773 ficou a cerca de 1.8 Å e, a reação não ocorreu. Recorrendo ao WHAM e sabendo que a energia de corte da enzima *SgrAI* é de 19.15 Kcal/mol, podemos observar, no gráfico da figura 17, que a barreira energética obtida é demasiado elevada, mostrando que esta reação não vai ocorrer numa escala biologicamente acessível. Uma causa para isso será o mau posicionamento dos Mg^{2+} (Dunten et al., 2008; Hingorani-Varma & Bitinaite, 2003).

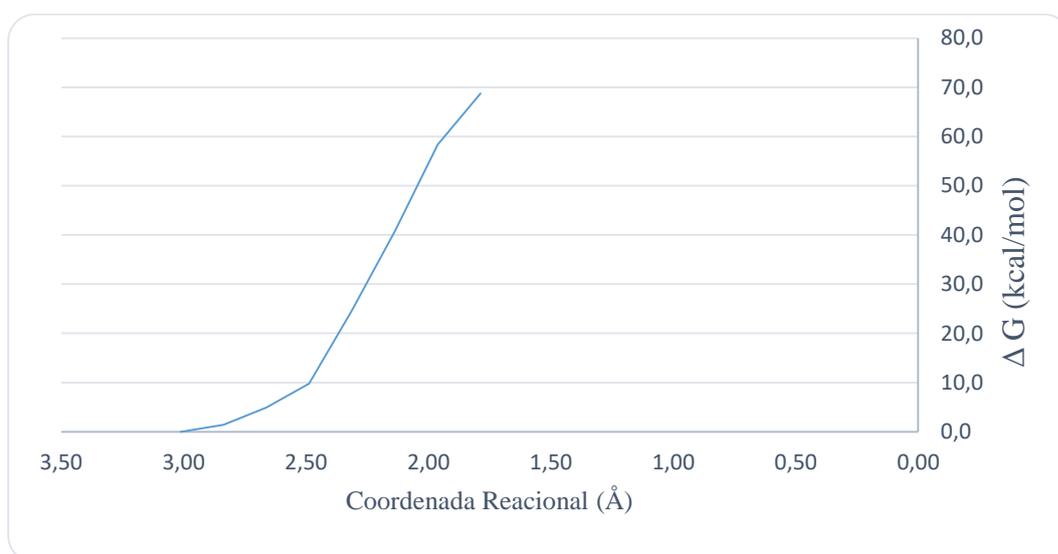


Figura 17 Gráfico representativo do perfil da energia durante a aproximação da WAT1773 ao átomo de fósforo do nucleótido DC 1364.

De forma a corrigir o sucedido fez-se a aproximação do MG1466 (M2), uma vez que este se encontrava a 7 Å de distância do átomo de oxigénio do grupo de saída. O objetivo foi posicioná-lo a cerca de 3 Å, o que está mais de acordo com a catálise por 2-iões-metálicos. Segundo o gráfico da figura 18 podemos observar que na aproximação

do MG1466(M2) ao local preterido a barreira de energia subiu consideravelmente ao invés do que seria esperado.

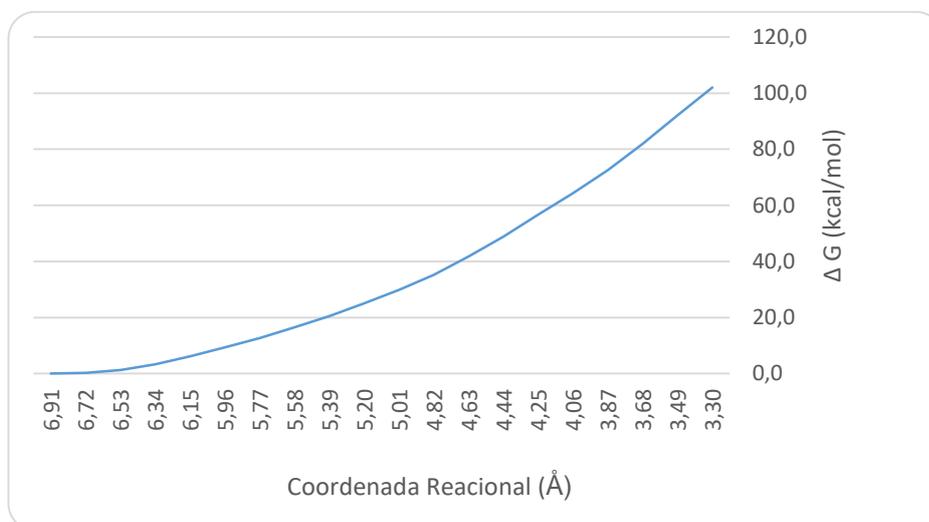


Figura 18 Gráfico representativo do perfil da energia durante a aproximação do MG1426 ao *backbone* do ADN.

Após esta reação, e tendo os iões metálicos nas posições corretas a aproximadamente 3 Å (Pingoud et al., 2005), partimos do ficheiro com as coordenadas da conformação da enzima *SgrAI* após a aproximação do MG21426 e procedeu-se a uma nova tentativa de aproximação da WAT1773 ao fosfato. No entanto, mais uma vez, a reação não ocorreu.

Esta reação foi testada com diferentes constantes de força, de 50 Kcal a 200 kcal e em todos os casos a molécula WAT1773 não se aproxima mais do que 1.9 Å e o MG1426 voltou à posição inicial a 7 Å (figura 19).

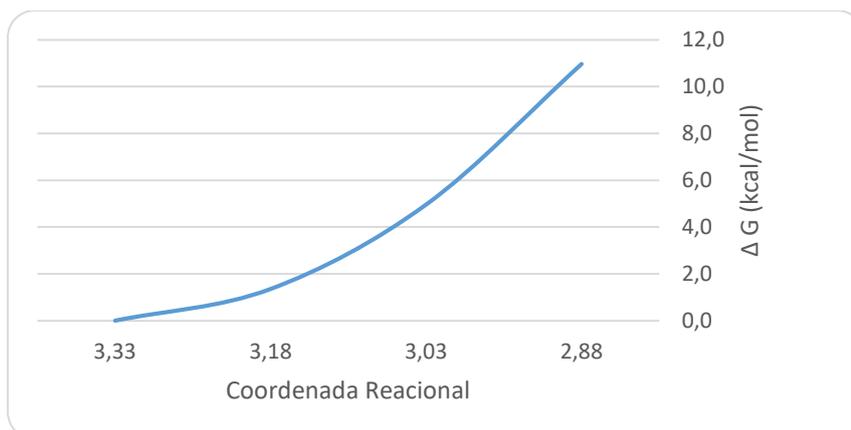


Figura 19 Gráfico representativo do perfil da energia durante a aproximação do da WAT1773 ao fosfato do nucleótido DC 1364 partindo da estrutura com os íons Mg^{2+} na posição correta para a catálise.

A fraca interação do resíduo Glu 442 com o MG21426 (M2) pode ser uma das justificações para o afastamento do íon Mg^{2+} . Para além disso, apesar da Lyn 581 estar disponível para aceitar o protão vindo da água 1773 aquando da sua desprotonação para o ataque ao fosfato, o resíduo Lyn 581 encontra-se muito afastado e esse pode ser o motivo para a WAT1773 não perder um protão e consequentemente não se ligar ao fósforo (figura 20).

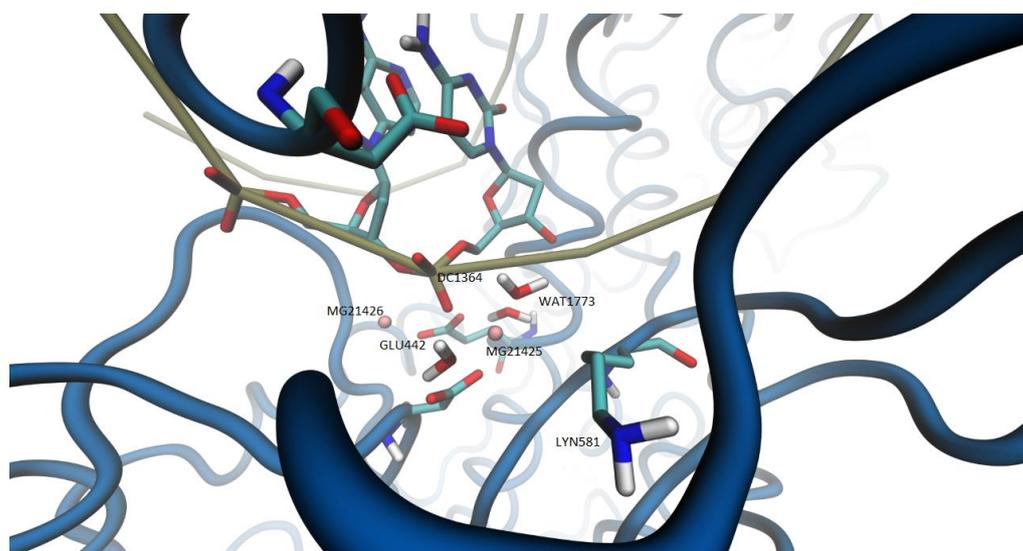


Figura 20 Imagem representativa do centro ativo no modelo tetramérico da enzima *SgrAI* com a Lys neutra.

Podemos concluir que este modelo não apresenta nenhum resíduo próximo da água catalítica que possa funcionar como base e, portanto, não se encontra funcional cataliticamente.

10.4.3 *SgrAI* tetramérica com Lys protonada

Uma vez que o modelo com a Lys neutra não respondeu como esperado, optou-se por seguir o mesmo protocolo, mas com a Lys protonada.

A literatura (Dunten et al., 2008) propõe que o próton envolvido na desprotonação da molécula nucleófila se liga ao fosfato do nucleótido seguinte, figura 21.

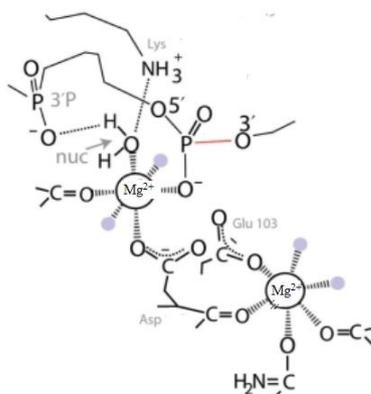


Figura 21 Esquema das interações em torno do Mg^{2+} ões da estrutura do *SgrAI*.

Após o tratamento do PDB com o *Xleap* e da simulação DM podemos analisar as variáveis temperatura, pressão e energia do sistema (Anexo 5) e ver que tudo decorreu conforme o esperado. No gráfico representativo do RMSD observa-se que houve oscilações consideráveis, no início da simulação DM o valor do RMSD tem uma subida considerável, isto pode justificar-se pelo próton adicionado à Lys e de alguma forma este obrigar a enzima a ajustar a conformação. No final da simulação DM o valor do RMSD já se encontra mais estável.

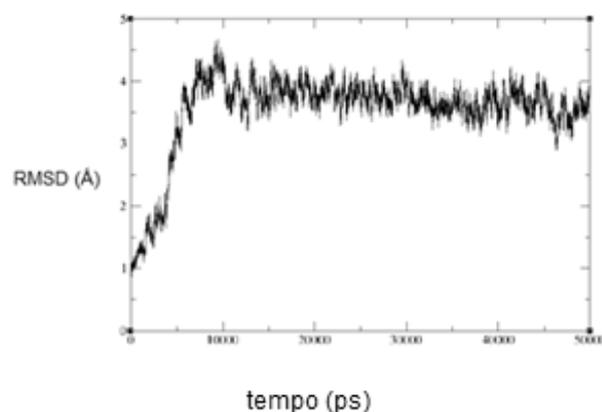


Figura 22 Representação gráfica da análise do RMSD considerando a variação média das distâncias dos resíduos de interesse.

Observando agora a tabela 7 onde estão representadas as interações intermoleculares e a variação média da distância dos mesmos, vemos que estas interações se mantêm iguais às do modelo anterior com muito poucas variações. O azoto do *backbone* da Arg 242 forma uma ponte de hidrogénio interagindo com o primeiro oxigénio do grupo fosfato do nucleótido DC 1382 (+18). Já a Arg 588 forma uma ponte de hidrogénio em que interage o azoto NE do aminoácido com o oxigénio 6 do nucleótido DC 1366 (+2). O azoto N da Leu 376 interage com o primeiro oxigénio do grupo fosfato do nucleótido DG 1386 (+22) através de uma ponte de hidrogénio. A Gln forma duas pontes de hidrogénio ao interagir com a DG 1404 (+40) entre do azoto N do aminoácido e o primeiro oxigénio do grupo fosfato e entre o azoto NE2 do aminoácido e o oxigénio 3' do nucleótido. Seria de esperar que os resíduos se mantivessem os mesmos uma vez que a conformação do modelo enzimático é igual com única diferença na protonação da Lys 241.

Tabela 7 Tabela com distância médias das interações entre os resíduos da proteína *SgrAI* e o ADN.

Aminoácido	Nucleótido	Distância média (Å)
Arg 242 -N	DC 1382 -O1P	2.72
Arg 588 -NE	DC 1366 -OG	3.03
Leu 376 -N	DG 1386 -O1P	2.91
Gln717 -N	DG 1404 -O1P	3.45
Gln 717 -NE2	DG 1404 -O3'	3.40

Com o foco mais uma vez no centro ativo e no mecanismo de catálise, passou-se a definir a camada alta tratada por QM. Os resíduos Asp 242, Glu 442, Asp 527, Lys 581, DC 1364 (DC com o grupo fosfato de interesse), DA 1363 (DA ligada ao grupo de saída), MG21425 e MG21426 (M1 e M2, respetivamente) WAT1769 (molécula de água nucleófila) e WAT1589 WAT1590, WAT1559 (outras moléculas coordenadas a iões Mg^{2+}). Apesar de, pela análise visual da conformação do centro ativo não ser a teoricamente esperada, deu-se início à aproximação da WAT1569 ao fosfato do nucleótido DC 1364. O resíduo WAT1569 não se aproximou mais que 2Å. Assim, tendo em conta o mau posicionamento dos iões Mg^{2+} , procedeu-se à aproximação do Mg^{2+} MG21426 ao átomo O3' do grupo de saída, de forma a passar dos 6.4 Å a aproximadamente 3 Å.

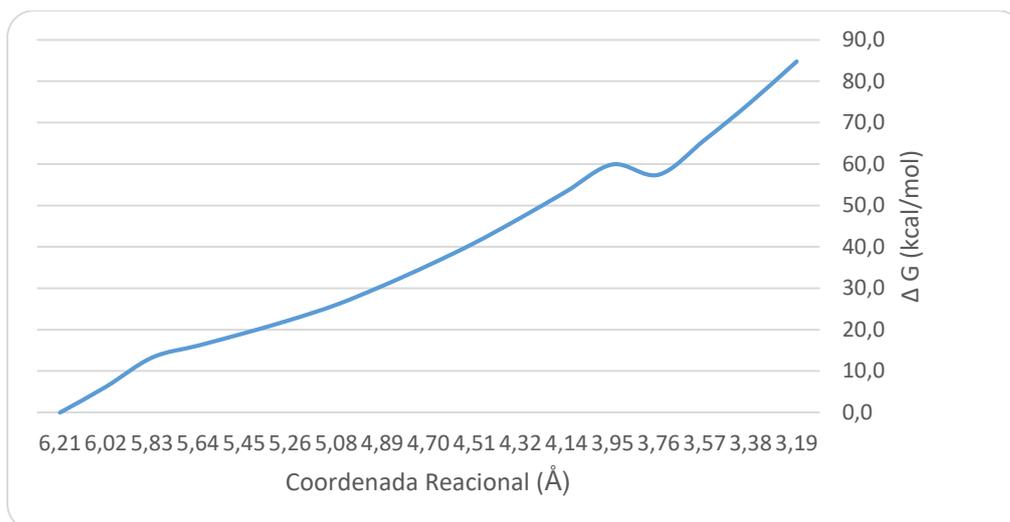


Figura 23 Gráfico representativo do perfil da energia durante a aproximação do íon Mg^{2+} para a posição correta para a catálise.

Apesar do gráfico do perfil energético da reação de aproximação do Mg^{2+} nos sugerir que a barreira energética está demasiada elevada, foi feita nova tentativa do ataque da molécula de água nucleófila WAT1763 partindo da conformação em que ambos os íons Mg^{2+} se encontram bem posicionados.

Novamente não houve reação, mesmo aumentando a constante de força aplicada. A molécula WAT1763 não avançou para além dos 2 Å e novamente a barreira energética manteve-se demasiado elevada.

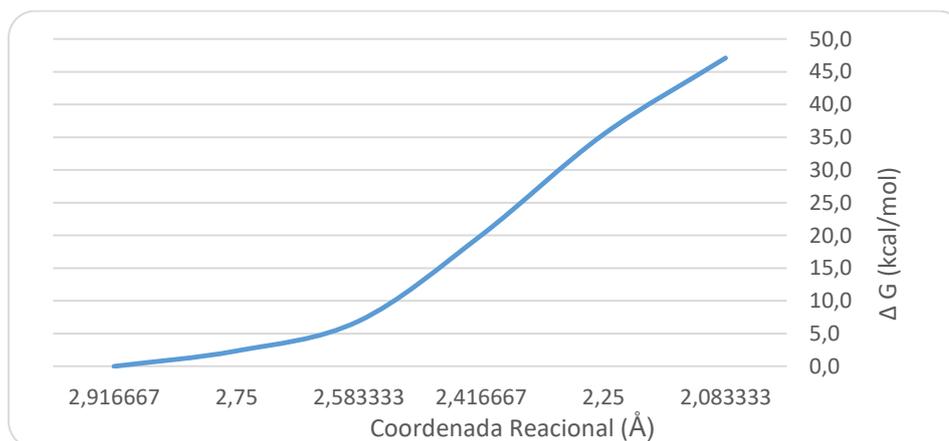


Figura 24 Gráfico representativo do perfil da energia durante a aproximação do resíduo WAT 1763 ao fosfato da DC 1364.

Atentando à figura 25, podemos reafirmar a má conformação geral do centro ativo do complexo enzima/ADN. O íon Mg^{2+} que devia suportar o grupo de saída após a reação de catálise, encontra-se muito distante para essa função. Para além disso, a distância entre os íões metálicos é bastante superior a 4 Å, tornando-os não funcionais na redução da repulsão eletrostática do estado de transição.

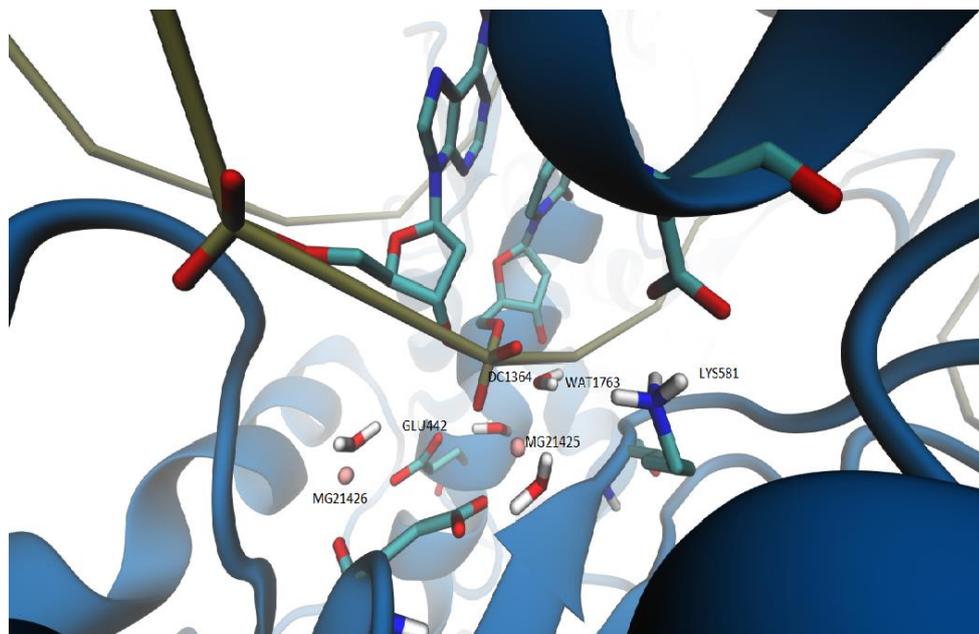


Figura 25 Imagem representativa do centro ativo no modelo tetramérico da enzima *SgrAI* com a Lys protonada.

Após este resultado e numa tentativa de perceber se o erro estaria na primeira fase do processo (desprotonação da molécula de água nucleofílica), foi substituída a molécula de água por um grupo hidroxilo, o que pressupõe que o protão é primeiro transferido para a Lys e depois ocorre ataque nucleófilo do OH^- . Esta abordagem também não teve sucesso. Como podemos ver no gráfico da figura 26 a barreira energética continua a ser muito superior ao desejado.

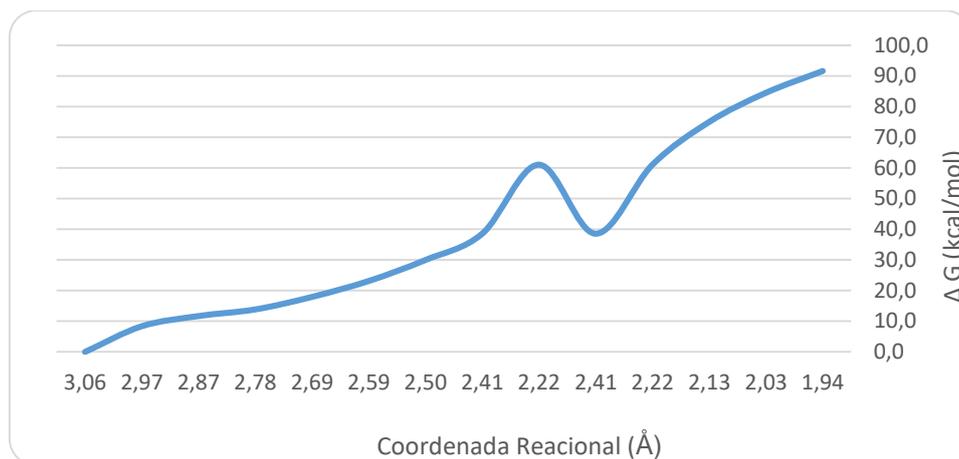


Figura 26 Gráfico representativo do perfil da energia durante o ataque nucleófilo do grupo hidroxilo ao fosfato do nucleótido DC 1364.

Seria interessante, numa continuação deste estudo, testar o modelo de mecanismo com três íons metálicos proposto Horton e Perona (Horton & Perona, 2004) e até mesmo o mecanismo com um único íon metálico (Pingoud et al., 2005).

11 Conclusão

Com a realização deste trabalho, concluímos que o mecanismo da reação de catálise da enzima *SgrAI* terá de continuar a ser alvo de estudo, embora se tenham feito alguns progressos.

O mecanismo catalítico dos 2-iões-metálicos sugerido por (Dunten et al., 2008) não se mostrou eficaz durante este estudo. Uma das hipóteses para a falha deste mecanismo será o facto deste complexo enzimático precisar de três iões metálicos para estabilização do centro ativo. Ao analisar todas as tentativas de reação com as diferentes abordagens, verificou-se que o segundo Mg^{2+} apresentava sempre uma posição distante do local M2 e aquando das tentativas da sua aproximação ele voltava sempre ao ponto de partida.

Assim, sugere-se que este ião Mg^{2+} se encontrava no local M3 (Horton & Perona, 2004; Pingoud et al., 2005) sendo necessário mais um ião Mg^{2+} , de forma a poderem ocupar as posições M1, M2 e M3, passando assim a ser um mecanismo catalítico 3-iões-metálicos. Este mecanismo já abordado por (Horton & Perona, 2004) é utilizado por enzimas *EcoRV* (Horton & Perona, 2004; Little, Dunten, Bitinaite, & Horton, 2011; Pingoud et al., 2005).

Esta hipótese poderá ser o ponto de partida para um estudo futuro envolvendo esta enzima.

12 Bibliografia

- Aggarwal, A. K. (1995). Structure and function of restriction endonucleases. *Curr Opin Struct Biol*, 5(1), 11-19.
- Ahmad, S., Keskin, O., Sarai, A., & Nussinov, R. (2008). Protein-DNA interactions: structural, thermodynamic and clustering patterns of conserved residues in DNA-binding proteins. *Nucleic Acids Res*, 36(18), 5922-5932. doi:10.1093/nar/gkn573
- Alder, B. J., & Wainwright, T. E. (1959). Studies in Molecular Dynamics. I. General Method. *The Journal of Chemical Physics*, 31(2), 459-466. doi:10.1063/1.1730376
- Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4), 2384-2393. doi:10.1063/1.439486
- Antunes, J. M. P. d. D. (2016). *Modelação molecular/bioinformática estrutural da hemaGlutina do vírus influenza*. (Mestrado em Bioquímica Dissertação em Mestrado). Universidade de Lisboa, Lisboa. <https://repositorio.ul.pt/handle/10451/23614>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235-242. doi:10.1093/nar/28.1.235
- Bickle, T. A., & Kruger, D. H. (1993). Biology of DNA restriction. *Microbiol Rev*, 57(2), 434-450.
- Bitinaite, J., & Schildkraut, I. (2002). Self-generated DNA termini relax the specificity of SgrAI restriction endonuclease. *Proceedings of the National Academy of Sciences*, 99(3), 1164-1169. doi:10.1073/pnas.022346799
- Carvalho, A. T., Barrozo, A., Doron, D., Kilshtain, A. V., Major, D. T., & Kamerlin, S. C. (2014). Challenges in computational studies of enzyme structure, function and dynamics. *J Mol Graph Model*, 54, 62-79. doi:10.1016/j.jmglm.2014.09.003
- Cheatham, T.E., Miller, J.L., Fox, T., Darden, T.A., Kollman, P.A. (1995). Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *Journal of the American Chemical Society*, 117 (14), 4193-4194. doi: 10.1021/ja00119a045

Connolly, B. A., Eckstein, F., & Pingoud, A. (1984). The stereochemical course of the restriction endonuclease EcoRI-catalyzed reaction. *J Biol Chem*, 259(17), 10760-10763.

D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. (2018). *AMBER 2018*.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12), 10089-10092. doi:10.1063/1.464397

Deibert, M., Grazulis, S., Sasnauskas, G., Siksnyis, V., & Huber, R. (2000). Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat Struct Biol*, 7(9), 792-799. doi:10.1038/79032

Dryden, D. T., Murray, N. E., & Rao, D. N. (2001). Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res*, 29(18), 3728-3741. doi:10.1093/nar/29.18.3728

Dunten, P. W., Little, E. J., Gregory, M. T., Manohar, V. M., Dalton, M., Hough, D., Horton, N. C. (2008). The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res*, 36(16), 5405-5416. doi:10.1093/nar/gkn510

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biol*, 9, 71. doi:10.1186/1741-7007-9-71

Fuxreiter, M. (2014). *Computational Approaches to Protein Dynamics: From Quantum to Coarse-Grained Methods*. Boca Raton: Taylor & Francis Group.

González, M. A. (2011). Force fields and molecular dynamics simulations. *JDN*, 12, 169-200. Retrieved from <https://doi.org/10.1051/sfn/201112009>

Halford, S. E., Bilcock, D. T., Stanford, N. P., Williams, S. A., Milsom, S. E., Gormley, N. A., Szczelkun, M. D. (1999). Restriction endonuclease reactions requiring two recognition sites. *Biochem Soc Trans*, 27(4), 696-699. doi:10.1042/bst0270696

Hingorani-Varma, K., & Bitinaite, J. (2003). Kinetic analysis of the coordinated interaction of SgrAI restriction endonuclease with different DNA targets. *J Biol Chem*, 278(41), 40392-40399. doi:10.1074/jbc.M304603200

Horton, N. C., & Perona, J. J. (2004). DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry*, 43(22), 6841-6857. doi:10.1021/bi0499056

J.M., B. (2004). Molecular Phylogenetics of Restriction Endonucleases. In *Nucleic Acids and Molecular Biology* (Vol. 14). Berlin, Heidelberg: Springer.

Jean-Paul Ryckaert, G. C., Herman J.C Berendsen,. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3), 327-341. doi:https://doi.org/10.1016/0021-9991(77)90098-5

Little, E. J., Dunten, P. W., Bitinaite, J., & Horton, N. C. (2011). New clues in the allosteric activation of DNA cleavage by SgrAI: structures of SgrAI bound to cleaved primary-site DNA and uncleaved secondary-site DNA. *Acta Crystallogr D Biol Crystallogr*, 67(Pt 1), 67-74. doi:10.1107/s0907444910047785

Ma, X., Shah, S., Zhou, M., Park, C. K., Wysocki, V. H., & Horton, N. C. (2013). Structural analysis of activated SgrAI-DNA oligomers using ion mobility mass spectrometry. *Biochemistry*, 52(25), 4373-4381. doi:10.1021/bi3013214

Martins, Lucas & Lima, Anderson. (2016). Métodos semiempíricos de química quântica no ensino da entalpia das reações químicas. *Scientia Plena*. 12. 10.14808/sci.plena.2016.069903.

Niv, M. Y., Ripoll, D. R., Vila, J. A., Liwo, A., Vanamee, E. S., Aggarwal, A. K., Scheraga, H. A. (2007). Topology of Type II REases revisited; structural classes and the common conserved core. *Nucleic Acids Res*, 35(7), 2227-2237. doi:10.1093/nar/gkm045

NobelPrize.org. (2013). The Nobel Prize in Chemistry 2013 [Press release]. Retrieved from <https://www.nobelprize.org/prizes/chemistry/2013/summary/>

Park, C. K., Joshi, H. K., Agrawal, A., Ghare, M. I., Little, E. J., Dunten, P. W., Horton, N. C. (2010). Domain swapping in allosteric modulation of DNA specificity. *PLoS Biol*, 8(12), e1000554. doi:10.1371/journal.pbio.1000554

Park, C. K., Sanchez, J. L., Barahona, C., Basantes, L. E., Sanchez, J., Hernandez, C., & Horton, N. C. (2018). The run-on oligomer filament enzyme mechanism of SgrAI: Part 2. Kinetic modeling of the full DNA cleavage pathway. *J Biol Chem*, 293(38), 14599-14615. doi:10.1074/jbc.RA118.003682

Park, C. K., Stiteler, A. P., Shah, S., Ghare, M. I., Bitinaite, J., & Horton, N. C. (2010). Activation of DNA cleavage by oligomerization of DNA-bound SgrAI. *Biochemistry*, 49(41), 8818-8830. doi:10.1021/bi100557v

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and anaLys is. *J Comput Chem*, 25(13), 1605-1612. doi:10.1002/jcc.20084

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem*, 26(16), 1781-1802. doi:10.1002/jcc.20289

Pingoud, A., Fuxreiter, M., Pingoud, V., & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci*, 62(6), 685-707. doi:10.1007/s00018-004-4513-1

Pingoud, A., & Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res*, 29(18), 3705-3727. doi:10.1093/nar/29.18.3705

Queirós, S. F. d. S. (2016). *Modelação da interação da glicoproteína e do vírus do dengue com ligantes de interesse farmacológico*. (Mestrado em Biofísica e Bionanossistemas Dissertação de Mestrado). Universidade do Minho, <http://hdl.handle.net/1822/45555>

Ramos, R. M. C. (2012). *Determinação computacional dos efeitos da mutagénese em interfaces proteína-ADN*. (Mestre em Química). FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO, <https://repositorio-aberto.up.pt/bitstream/10216/65143/2/24220.pdf>

Smith, H. O., & Nathans, D. (1973). Letter: A suggested nomenclature for bacterial host modification and restriction systems and their enzymes. *J Mol Biol*, 81(3), 419-423. doi:10.1016/0022-2836(73)90152-6

Tautz, N., Kaluza, K., Frey, B., Jarsch, M., Schmitz, G. G., & Kessler, C. (1990). SgrAI, a novel class-II restriction endonuclease from *Streptomyces griseus* recognizing the octanucleotide

sequence 5'-CR/CCGGYG-3' [corrected]. *Nucleic Acids Res*, 18(10), 3087. doi:10.1093/nar/18.10.3087

van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Yu, H. B. (2006). Biomolecular Modeling: Goals, Problems, Perspectives. *Angewandte Chemie International Edition*, 45(25), 4064-4092. doi:10.1002/anie.200502655

Venclovas, C., Timinskas, A., & Siksnyis, V. (1994). Five-stranded beta-sheet sandwiched with two alpha-helices: a structural link between restriction endonucleases EcoRI and EcoRV. *Proteins*, 20(3), 279-282. doi:10.1002/prot.340200308

Vipond, I. B., & Halford, S. E. (1995). Specific DNA recognition by EcoRV restriction endonuclease induced by calcium ions. *Biochemistry*, 34(4), 1113-1119. doi:10.1021/bi00004a002

Vreven, T., Morokuma, K., Farkas, Ö., Schlegel, H. B., & Frisch, M. J. (2003). Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. *Journal of Computational Chemistry*, 24(6), 760-769. doi:10.1002/jcc.10156

Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of Lys ozyme. *J Mol Biol*, 103(2), 227-249. doi:10.1016/0022-2836(76)90311-9

Woodbury, C. P., Jr., Downey, R. L., & von Hippel, P. H. (1980). DNA site recognition and overmethylation by the Eco RI methylase. *J Biol Chem*, 255(23), 11526-11533.

Yang, W., Lee, J. Y., & Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg²⁺-ion catalyzed and substrate specificity. *Mol Cell*, 22(1), 5-13. doi:10.1016/j.molcel.2006.03.013

Anexo I

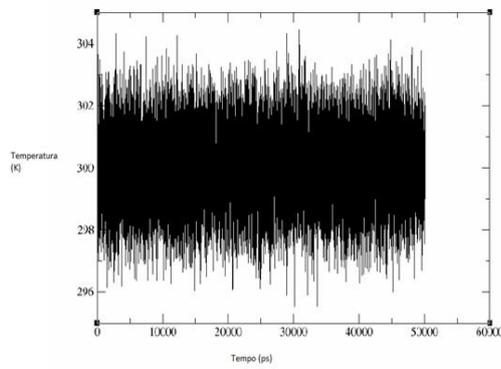


Figura 27 Gráfico representativo da variação da temperatura (K) durante a DM do dímero de *SgrAI*.

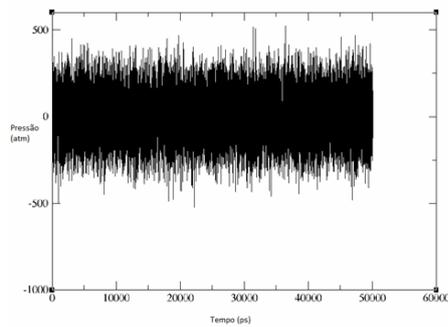


Figura 28 Gráfico representativo da variação da pressão (atm) durante a DM do dímero de *SgrAI*.

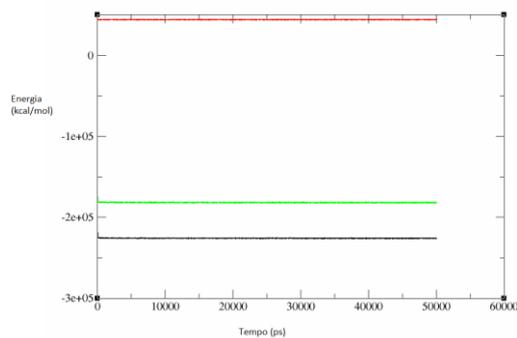


Figura 29 Figura representativa da energia do Sistema durante a DM do dímero de *SgrAI*: a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total.

Anexo II

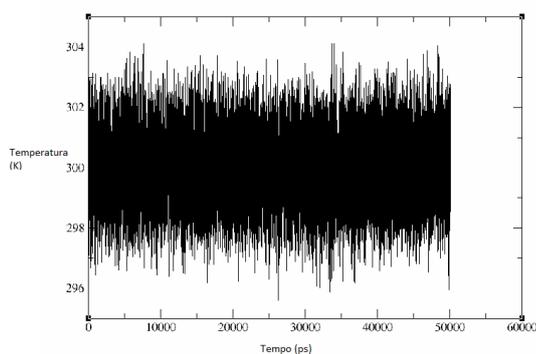


Figura 30 Gráfico representativo da variação da temperatura (K) durante a DM do dímero de *SgrAI* com mutação no par de bases DC-DG.

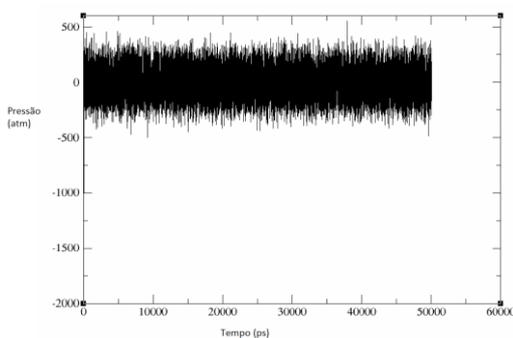


Figura 31 Gráfico representativo da variação da pressão (atm) durante a DM do dímero de *SgrAI* com mutação no par de bases DC-DG.

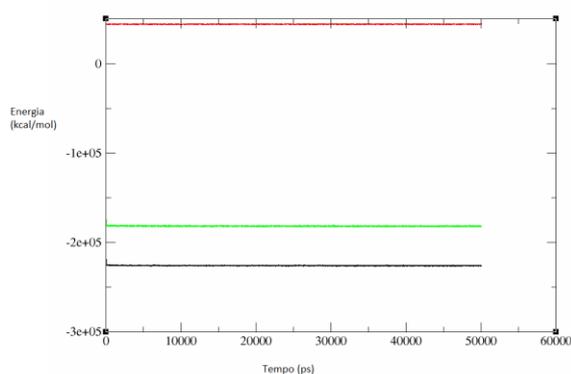


Figura 32 Figura representativa da energia do sistema durante a DM do dímero de *SgrAI* com mutação no par de bases DC-DG: a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total

Anexo III

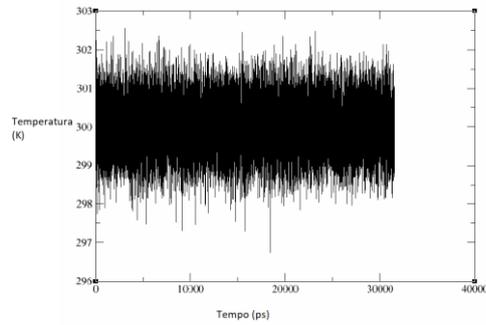


Figura 33 Gráfico representativo da variação da temperatura (K) durante a DM do tetrâmero de *SgrAI* com íons de Ca^{2+} .

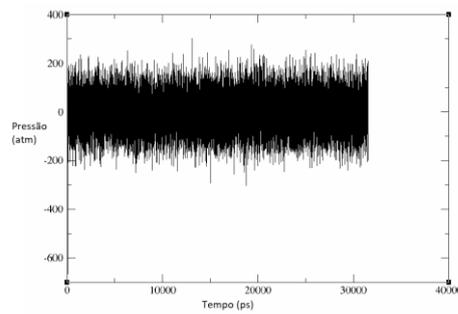


Figura 34 Gráfico representativo da variação da pressão (atm) durante a DM do tetrâmero de *SgrAI* com íons de Ca^{2+} .

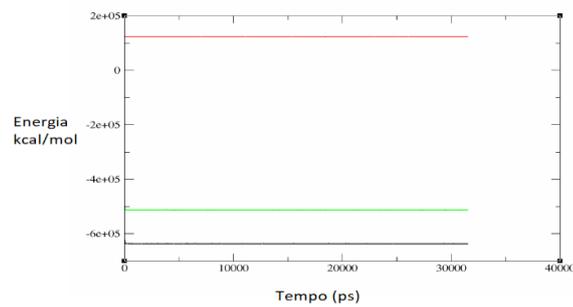


Figura 35 Figura representativa da energia do sistema durante a DM do tetrâmero de *SgrAI* com íons de Ca^{2+} : a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total.

Anexo IV

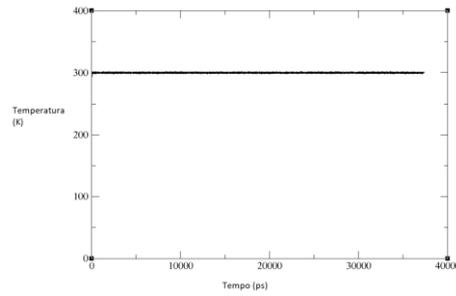


Figura 36 Gráfico representativo da variação da temperatura (K) durante a DM do tetrâmero de *SgrAI* com íons de Mg^{2+} e Lys neutra.

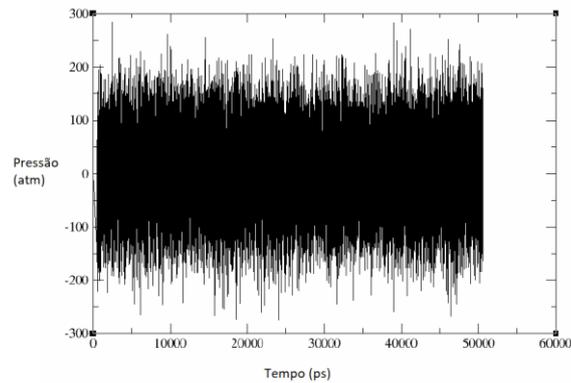


Figura 37 Gráfico representativo da variação da pressão (atm) durante a DM do tetrâmero de *SgrAI* com íons de Mg^{2+} e Lys neutra.

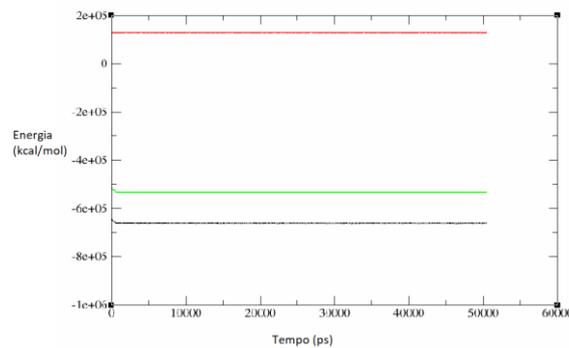


Figura 38 Figura representativa da energia do sistema durante a DM do tetrâmero de *SgrAI* com íons de Mg^{2+} e Lys neutra: a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total.

Anexo V

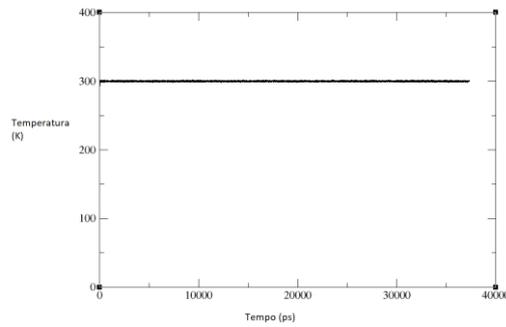


Figura 39 Gráfico representativo da variação da temperatura (K) durante a DM do tetrâmero de *SgrAI* com íões de Mg^{2+} e Lys protonada.

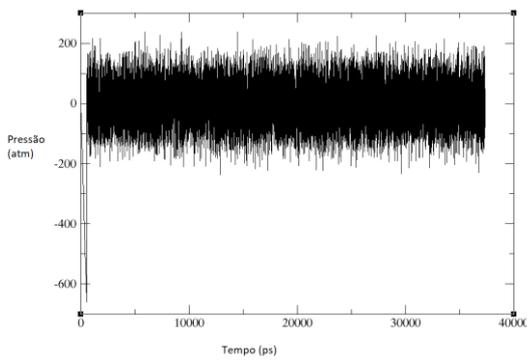


Figura 40 Gráfico representativo da variação da pressão (atm) durante a DM do tetrâmero de *SgrAI* com íões de Mg^{2+} e Lys protonada.

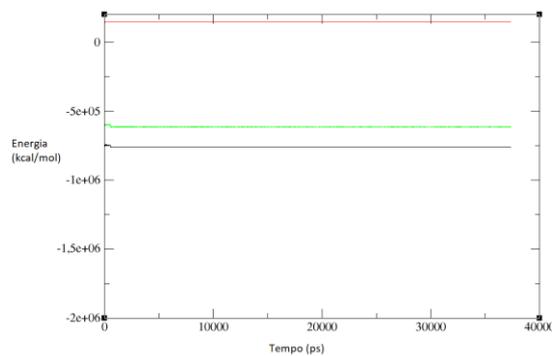


Figura 41 Figura representativa da energia do sistema durante a DM do tetrâmero de *SgrAI* com íões de Mg^{2+} e Lys protonada: a linha vermelha representa a energia cinética, e a linha preta mostra a energia potencial. A linha verde mostra o total.

