



CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS

Recredenciado pela Portaria Ministerial nº 3.607, de 17/10/05, D.O.U. nº 202, de 20/10/2005
ASSOCIAÇÃO EDUCACIONAL LUTERANA DO BRASIL

Ranyelson Neres Carvalho

**APLICAÇÃO DE FERRAMENTA OLAP PARA IDENTIFICAR QUEBRAS DE LIGAÇÕES QUÍMICAS EM
MOLÉCULAS DE ANTOCIANINA**

Palmas - TO

2016

Ranyelson Neres Carvalho

**APLICAÇÃO DE FERRAMENTA OLAP PARA IDENTIFICAR QUEBRAS DE LIGAÇÕES QUÍMICAS EM
MOLÉCULAS DE ANTOCIANINA**

Trabalho de Conclusão de Curso (TCC) elaborado e apresentado como requisito parcial para obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientadora: Prof(a). M.Sc. Madianita Bogo Marioti

Co-orientadora: Prof(a). M.Sc. Walkiria Régis de Medeiro

Palmas – TO

2016

Ranyelson Neres Carvalho

**APLICAÇÃO DE FERRAMENTA OLAP PARA IDENTIFICAR QUEBRAS DE LIGAÇÕES QUÍMICAS EM
MOLÉCULAS DE ANTOCIANINA**

Trabalho de Conclusão de Curso (TCC) elaborado e apresentado como requisito parcial para a obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientadora: Prof(a). M.Sc. Madianita Bogo Marioti

Co-orientadora: Prof(a). M.Sc. Walkiria Régis de Medeiro

Aprovada em: ____/____/____

BANCA EXAMINADORA

Prof(a). M.Sc. Madianita Bogo Marioti
Centro Universitário Luterano de Palmas

Prof. M.Sc. Fabiano Fagundes
Centro Universitário Luterano de Palmas

Prof(a). M.Sc. Walkiria Régis de Medeiro
Centro Universitário Luterano de Palmas

Palmas – TO

2016

AGRADECIMENTOS

Aos meus pais, irmão e a toda a minha família que, com muito carinho e apoio, não mediram esforços para que chegasse até esta etapa de minha vida.

As minhas orientadoras, que carinhosamente chamo de mãe (Madianita) e babá (Walkiria), por seus ensinamentos, paciência e confiança ao longo do desenvolvimento deste projeto. Ganhei duas pessoas maravilhosas na minha vida, que torceram e acreditaram durante todo o projeto. Eu posso dizer que o resultado alcançado aqui, não teria sido o mesmo sem vocês.

Aos professores Edeilson, Jackson, Cristina, Fernando, Parcilene, Fabiano e Madianita, que foram tão importantes na minha vida acadêmica e que juntos formam uma família e tanto.

Aos meus amigos: Leomar, Jhonatan, Jesiel, Marcus Germano, Cleypson, Marcos Vinicius, Welles, William e Robson.

Enfim, agradeço a todos aqueles que de alguma forma estiveram e estão próximos de mim, fazendo esta vida valer cada vez mais a pena. Muito obrigado!

RESUMO

CARVALHO, Ranyelson Neres. **Aplicação de ferramenta OLAP para identificar quebras de ligações químicas em moléculas de antocianina**. 2016. 67 f. TCC (Graduação) – Curso de Sistemas de Informação, Centro Universitário Luterano de Palmas, Palmas, 2016.

A simulação de dinâmica molecular é uma técnica importante para o estudo de sistemas moleculares. Esta técnica possibilita obter informações de uma molécula por meio de sua evolução temporal, retornando informações como, por exemplo, as distâncias entre os átomos ao longo do processo de simulação. Estas distâncias podem ser utilizadas para identificar como ocorre o processo de quebra de ligações de uma molécula. A existência de uma base de dados de referência disponibilizada pelo NIST (*National Institute of Standards and Technology*) permite que as distâncias extraídas da simulação sejam comparadas com os valores de referência, identificando assim o momento em que as distâncias apresentam valores fora do padrão, evidenciando a quebra de ligações da molécula. Este projeto tem como finalidade aplicar uma ferramenta OLAP para automatizar o processo de análise dos dados, identificando os momentos em que ocorre este fenômeno em moléculas de antocianinas, tendo como base os valores disponibilizados pelo NIST.

PALAVRAS-CHAVE: Dinâmica Molecular, Ligações Químicas, NIST, OLAP.

LISTA DE FIGURAS

Figura 1: Processo ETL.....	18
Figura 2: Padronização dos dados.....	19
Figura 3: Cubo multidimensional OLAP para análise.....	21
Figura 4: Modelagem multidimensional e relacional.....	21
Figura 5: Etapas para o desenvolvimento do projeto.....	27
Figura 6: Script Perl - Topologia.....	29
Figura 7: Estrutura molecular: Antocianina; Estrutura Molecular - NIST; Tabela com os valores aceitáveis.....	31
Figura 8 – Molécula Antocianina: cianidina-3-glicosídeo formato PDB.....	32
Figura 9: Molécula Antocianina: cianidina-3-glicosídeo.....	33
Figura 10: arquivo trajetoria.pdb.....	34
Figura 11: Distâncias extraídas.....	36
Figura 12: Processo de leitura, limpeza e inserção dos dados.....	36
Figura 13: Modelo conceitual.....	37
Figura 14 - <i>drill up</i> - distâncias.....	38
Figura 15: Resultado ferramenta OLAP.....	38
Figura 16: parâmetro minimização de energia.....	54
Figura 17: parâmetros nvt.mdp.....	55
Figura 18: parâmetro npt.mdp.....	56
Figura 19: parâmetro md.mdp.....	58
Figura 20: parâmetro minim2.mdp.....	60
Figura 21: criar dashboard.....	64
Figura 22: importar os dados.....	64
Figura 23: conexão com o banco de dados.....	65
Figura 24: tabelas a serem carregadas.....	65
Figura 25: tipo de visualização.....	66
Figura 26: organização dos dados.....	66
Figura 27: resultado da comparação dos dados.....	67

LISTA DE QUADROS

Tabela 1: Valores para as ligações da antocianina	31
Tabela 2: valores de referência NIST	47
Tabela 3: Valores máximos e mínimos para as distâncias de ligações.....	47
Tabela 4: frame cianidina-3-glicosideo (distâncias).....	62

LISTA DE ABREVIATURAS E SIGLAS

BI - Business Intelligence

DOLAP – Desktop OLAP

ETL – Extraction, Transform, Load

GROMACS - Groningen Machine for Chemical Simulations

LaRC – Laboratório de Redes de Computadores

MOLAP – Multidimensional OLAP

NIST - National Institute of Standards and Technology

NPT - Number of Particules, Volume and Temperature

NVT - Number of Particules, Pressure and Temperature

OLAP - Online Analytical Processing

OPLS - Optimized Potential for Liquid Simulations

PDB – Protein Data Bank

RCSB - Research Collaboratory for Structural Bioinformatics

ROLAP – Relacional OLAP

SQL - Structured Query Language

WOLAP – Web OLAP

SUMÁRIO

1. INTRODUÇÃO	10
2. REFERENCIAL TEÓRICO	14
2.1 Dinâmica Molecular	14
2.2 Simulação Computacional de Dinâmica Molecular	15
2.3 Processo ETL	18
2.4 OLAP (Processamento Analítico Online)	19
2.4.1 Classificação quanto aos métodos de armazenamento	22
2.4.2 Ferramentas OLAP	23
3. METODOLOGIA.....	25
3.1 <i>Hardware</i>	25
3.2 <i>Software</i>	25
3.3 Métodos	26
3.4 Preparação do ambiente para obtenção dos dados	28
4. RESULTADOS E DISCUSSÃO.....	30
4.1 Base de dados experimentais do NIST.....	30
4.2 Resultados da dinâmica molecular	32
4.3 Processo ETL	35
4.3.1 Extração	35
4.3.2 Transformação e Carga (<i>Load</i>)	36
4.4 OLAP	37
4.4.1 Uso da ferramenta OLAP – <i>Microstrategy Analytics Desktop</i>	37
5. CONSIDERAÇÕES FINAIS	40
6. REFERÊNCIAS.....	41
ANEXOS	46
APÊNDICES.....	51

1. INTRODUÇÃO

O uso de técnicas computacionais aplicadas à física, matemática, biologia e química possibilitam um entendimento mais amplo de fenômenos ligados à essas áreas. Estas técnicas podem auxiliar na resolução de modelos matemáticos, na realização de simulações *in silico*¹ ou no desenvolvimento de novos meios para o diagnóstico de doenças, dentre outros contextos.

A química é uma área em que as técnicas computacionais são muito utilizadas para estudar virtualmente sistemas, o que antes era realizado, apenas, através de experimentos laboratoriais. Especificamente na química teórica, técnicas computacionais são utilizadas para estudar virtualmente sistemas que podem ser reproduzidos em laboratório, representando, assim, sistemas reais. Por exemplo, a realização de simulação de dinâmica molecular permite analisar o comportamento dos átomos por meio de uma evolução temporal, através de equações de movimento das partículas, criando uma trajetória descrita pelos átomos (SHERER, 2010).

A partir das simulações da dinâmica molecular, é possível avaliar diferentes hipóteses sobre o comportamento do sistema e extrair informações para o desenvolvimento de novos medicamentos, formulações cosméticas, entendimento de mecanismos de reações, dentre outros. Com os dados gerados, evita-se a realização de inúmeros experimentos em laboratórios, custos com a aquisição de equipamentos caros, reagentes e, ainda, contaminação ambiental com os dejetos gerados.

Voltado para este cenário, no CEULP/ULBRA está em andamento o projeto Quanto, que visa obter propriedades moleculares de antocianinas, extraídas dos frutos de pimenta malagueta (*Capsicum frutescens*) para a avaliação da atividade antioxidante desses compostos em formulações cosméticas. A motivação do projeto é a busca das indústrias farmacêuticas e cosméticas em produzir cosméticos verdes, isto é, produtos elaborados com base em ativos naturais, com o menor impacto ambiental possível (MIGUEL, 2011).

O problema é que os componentes naturais tendem a ser instáveis, inviabilizando o uso nas formulações cosméticas, assim, são necessários estudos para a otimização da sua estrutura molecular. A modelagem molecular, que consiste

¹ Simulações computacionais que modelam um processo natural ou de laboratório.

em um conjunto de ferramentas computacionais que possibilitam a construção, visualização e armazenamento de sistemas moleculares (BARREIRO et al., 1996), é uma maneira de buscar essa otimização. Uma forma de obter informações para analisar o comportamento da molécula é através do uso de *softwares* que realizam simulações da interação da molécula com o meio. Nesse projeto, foram realizadas simulações sobre a molécula de cianidina-3-glicosídeo, que é a antocianina retirada da pimenta malagueta, visando estabilizar essas moléculas de forma que poderão ser utilizadas como antioxidantes em cosméticos.

Para a realização de simulações de dinâmica molecular, dentre vários *softwares*, foi escolhido o Gromacs, por ser livre e de código aberto. O Gromacs é um pacote que oferece recursos para realizar simulações de dinâmica molecular (SPOEL et al., 2005). Estas simulações são processos demorados, que demandam muito tempo de processamento, em outro trabalho vinculado ao projeto Quanto, foi construído um *cluster* que, segundo Buyya, “consiste de uma coleção de computadores interconectados e trabalhando juntos como único recurso computacional integrado” (1999 apud Conti 2009). Dessa forma, o Gromacs é executado no cluster, de forma que é feita a paralelização das tarefas da simulação, o que reduz o tempo de resposta das simulações.

Nas simulações, a molécula de cianidina-3-glicosídeo foi exposta a um único ambiente por determinados períodos de tempo, com isso, obteve-se informações das distâncias entre os átomos, em vários momentos. A análise dessas distâncias permite chegar a conclusões sobre a estabilidade da molécula, pois quando as distâncias aumentam muito pode significar que ocorreu quebra de ligações, causando a perda do seu poder antioxidante. Assim, a descoberta de momentos que ocorrem quebras pode ajudar a entender por que elas quebraram, e assim, encontrar meios de estabilizá-las, ou seja, evitar que as ligações sejam rompidas e que o poder antioxidante seja mantido.

O NIST (*National Institute of Standards and Technology*), agência governamental responsável por promover padrões de medidas para o avanço da ciência e tecnologia, entre outras, fornece uma base de dados experimentais de espectroscopia molecular, com informações sobre faixas de valores ótimos para distâncias de ligações (NIST, 2016). Assim, as distâncias dos átomos das moléculas de antocianinas retornadas pelo Gromacs podem ser comparadas aos valores de

referência oferecidos pelo NIST, de forma que, quando a distância for maior que a indicada pelo NIST, pode-se detectar a possível quebra da ligação.

Um problema é que os dados retornados na simulação de um período de 4 nanosegundos chega a aproximadamente a 460 linhas. No decorrer do projeto, serão simulados períodos mais longos, de forma que a quantidade de dados aumentará proporcionalmente à quantidade de tempo. Esse fator inviabiliza uma comparação manual dos resultados da simulação com os valores de referência do NIST.

Devido a esse fator, decidiu-se utilizar uma ferramenta OLAP para a manipulação dos dados. As ferramentas OLAP têm o objetivo de transformar grandes volumes de dados em informações capazes de auxiliar no processo de tomada de decisão, apresentando diferentes formas de visualização. Com isso, podem auxiliar a identificar as distâncias entre as ligações químicas por meio de um cruzamento de dados (SINGH, 2001).

Existem vários aplicativos que oferecem ferramentas OLAP, uma delas é a ferramenta *Microstrategy Analytics Desktop*, que oferece recursos como: *slide and dice*, o qual permite fatiar as informações para visualizar os dados a partir de várias perspectivas, esta operação possibilita modificar a posição de uma informação bem como, alternar entre linhas e colunas de maneira a facilitar a compreensão dos dados; e o *drill down/up*, que consiste em explorar os dados por diferentes níveis de detalhes (SINGH, 2001). Esta ferramenta foi utilizada no projeto para a obtenção de informações a partir dos resultados da simulação da dinâmica molecular sobre a molécula cianidina-3-glicosídeo.

Para que a ferramenta OLAP seja aplicada é necessário que os dados retornados da simulação passem por um processo ETL (Extração, Transformação e Carga (*Load*)), pois os dados retornados não estão em um formato compatível com a entrada de dados da ferramenta OLAP. Este processo tem a finalidade extrair dados, transformá-los de forma que fiquem compatíveis com a ferramenta OLAP e carregá-los para uma base de dados de entrada. Desta forma, é possível obter dados estruturados e consistentes, que podem ser utilizados em ferramentas que fornecem suporte ao processo de tomada de decisão (PICHIALINI, 2008). Nesse trabalho, foi realizada a criação de uma base de dados relacional, com os valores das distâncias de ligação resultantes da dinâmica molecular e os valores de referência disponibilizados pelo NIST.

Desta forma, foi possível realizar um paralelo entre o resultado da simulação e os dados de referência do banco de dados do NIST, que apresentou os momentos em que as distâncias fugiram do padrão. Esta informação possibilitará identificar se houve quebra das ligações químicas na molécula analisada. Caso tenha ocorrido, também pode-se identificar os momentos em que as moléculas de antocianinas sofreram modificações estruturais. Isso auxiliará os pesquisadores a encontrar o motivo da quebra e tentar sanar o problema.

Este documento apresenta os procedimentos realizados para a identificação da quebra de ligações químicas em moléculas de antocianinas, através da ferramenta OLAP *Microstrategy Analytics Desktop*.

2. REFERENCIAL TEÓRICO

Nesta seção são abordados conceitos relacionados à dinâmica molecular, simulação computacional de dinâmica molecular, e como o processo ETL e os sistemas OLAP podem auxiliar na transformação e análise dos dados obtidos durante o processo de simulação de dinâmica molecular.

2.1 Dinâmica Molecular

A dinâmica molecular, segundo Sherer (2010), “é uma técnica computacional para resolver as equações de movimentos dos átomos, considerados como partículas puntiformes que interagem com os demais átomos do sistema”. Esta técnica permite analisar o comportamento dos átomos por meio de uma evolução temporal, através das equações de movimento das partículas, criando uma espécie de trajetória descrita pelos átomos.

A dinâmica molecular é baseada nos princípios da Mecânica Clássica, com a qual as moléculas são tratadas como uma coleção de partículas mantidas unidas por forças harmônicas ou elásticas (NAMBA, SILVA, SILVA, 2008). Estas forças resultam em movimentos de vai e vem, em torno de uma posição central do sistema em função do tempo, criando assim a interação necessária entre os átomos durante o processo de simulação.

As primeiras aplicações utilizando o método da Dinâmica Molecular foram realizadas por Alder e Wainwright no final dos anos 1950 e início dos anos 1960, para estruturar as interações entre esferas rígidas. A partir deste método passou-se a estudar o comportamento de sistemas moleculares por meio das equações de Newton, analisando todas as moléculas de um dado conjunto, a partir de posições e velocidades iniciais e do conhecimento da força que atua em cada molécula. As posições e velocidades de cada molécula do sistema são calculadas em um espaço de tempo de maneira contínua, para que seja obtida uma mudança conforme o tempo de simulação do sistema (FERNANDES, 1985).

As equações utilizadas podem ser descritas para um sistema pelas seguintes fórmulas (NAMBA, SILVA, SILVA, 2008):

$$F_i(t) = m_i a_i$$

Equação 1

$$F_i(t) = -\frac{\partial V(r_i)}{\partial \vec{r}_i}$$

Equação 2

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)\Delta t^2$$

Equação 3

[...] onde é calculada a força F_i que atua sobre cada partícula do sistema em um instante de tempo t , e a_i é a aceleração do átomo i de massa m_i (equação 1). Uma vez definido o campo de força, é possível calcular as forças que atuam sobre cada átomo, calculando-se a derivada primeira da energia potencial, obtida do campo de força escolhido, em relação às posições desses átomos (equação 2). A equação 2 gera diretamente a aceleração da partícula, a partir desta, integrando-se as equações de movimento (equação 3), pode-se obter velocidades, que por sua vez, proporciona a mudança de posição do átomo. Aplicando-se continuamente esse procedimento, obtém-se o que se denomina “trajetória”, que é o conjunto de posições e velocidades de cada partícula ao longo do tempo (NAMBA, SILVA, SILVA, 2008).

Os métodos de simulação de dinâmica molecular possibilitaram um entendimento mais amplo dos fenômenos físicos e químicos (MORGON, 2007). Tais métodos auxiliam na compreensão dos acontecimentos observados experimentalmente, o que possibilita uma análise do comportamento do sistema sobre diferentes perspectivas, sejam elas através de uma análise de temperatura, pressão, densidade dentre outros elementos, buscando a equilíbrio do sistema.

A seção a seguir apresenta como é realizado este método através da simulação computacional.

2.2 Simulação Computacional de Dinâmica Molecular

Para Shannon (1975, apud Leal, 2003) a simulação é uma metodologia de resolução de problemas, aplicada para implementar e analisar um procedimento real (físico) ou proposto em um computador (de forma virtual). Através dela é possível a realização de estudos em que a teoria e prática sejam partes complementares para explicar os problemas propostos. Ela atua como uma ponte entre os resultados obtidos experimentalmente e as teorias desenvolvidas para explicar as características ou comportamento de algo.

Harrel (2000) afirma que “em um modelo de simulação computacional podem ser testados diferentes valores para variáveis, que podem ser controladas e modificadas pelo projetista”. Desta forma, é possível ter o controle sobre as variáveis a serem estudadas possibilitando modificá-las de acordo com a necessidade do projetista ao longo do desenvolvimento da pesquisa como, por exemplo, determinar sob quais temperaturas, pressões ou ambientes serão realizadas as simulações.

A simulação computacional tem sido muito utilizada no estudo de sistemas moleculares, sejam eles isolados ou que interagem com um meio, em particular aqueles em fase líquida ou com problemas de conformação de biomoléculas (COSTA, 2005). Neste contexto, pode-se destacar a dinâmica molecular, que utiliza ferramentas computacionais para analisar o comportamento dos átomos por meio de sua evolução temporal através das equações de movimento de Newton.

Com a utilização de ferramentas computacionais é possível aumentar o nível de precisão dos cálculos. Possibilita obter resultados em um tempo hábil, além de contar com o uso de recursos visuais, como a geração de gráficos ou até mesmo *frames* durante o processo de simulação, o que possibilita uma análise mais interativa sobre o sistema em questão.

A realização da simulação computacional, por meio da técnica de dinâmica molecular, encontra-se como uma técnica computacional consolidada que pode correlacionar e/ou prever o comportamento de substâncias e misturas e o tempo computacional necessário para executá-las (MELO JUNIOR; HABERT, 2005). O uso desta técnica permite a realização de simulações em sistemas que são de difícil reprodução em laboratório como, por exemplo, reações a altíssimas pressões ou temperaturas.

A simulação computacional de dinâmica molecular, normalmente, inclui algumas etapas como: geração das configurações iniciais, preparação da estrutura, definição do ambiente de simulação, minimização, equilíbrio e a produção da dinâmica molecular em si (GUIDO, 2010).

O primeiro passo consiste em obter as configurações iniciais do sistema que se deseja estudar, o qual é composto pelo número de átomos presentes nas moléculas. Estas configurações podem ser obtidas em banco de dados especializados como o RCSB (*Research Collaboratory for Structural*

*Bioinformatics*²), ou geradas através de algum programa como, por exemplo, UCSF Chimera ou Avogadro.

A etapa de preparação envolve a definição do campo de força, que descreve as interações entre as moléculas de um determinado arranjo de configurações e as interações eletrostáticas entre os átomos que atuarão sobre a molécula. Isto permite que seja calculada a energia potencial total do sistema a partir da estrutura da molécula (OLIVEIRA, 2013).

O ambiente de simulação é realizado em uma caixa virtual, podendo ser cúbica, paralelepípedica ou octaédrica, na qual é possível adicionar outros componentes que irão interagir com o sistema como, por exemplo, moléculas de água ou álcool (NAMBA, SILVA, SILVA, 2008).

A minimização de energia do sistema visa à remoção de possíveis tensões existentes na configuração inicial das moléculas (OLIVEIRA, 2013). Esta etapa possibilita encontrar um conjunto de coordenadas que minimizam a energia potencial do sistema.

A equilíbrio do sistema tem como finalidade obter uma conformação favorável ao sistema. No qual são analisados pontos como temperatura, energia potencial, energia total, densidade etc., para que possa ser realizada a produção da dinâmica molecular.

A etapa de produção da dinâmica molecular envolve a integração das equações de movimento de Newton, provocando o deslocamento dos átomos, resultando na dinâmica molecular.

Pode-se acrescentar às etapas apresentadas a fase de análise, que consiste na verificação dos resultados obtidos durante a simulação, sendo que estes podem ser expressos por meio de gráficos.

Existem diversas ferramentas computacionais livres como: Gromacs, Tinker, Mopac e Charmm que realizam as simulações de dinâmica molecular. Estas possibilitam ao usuário um entendimento mais amplo de como ocorre este processo.

Como resultado das simulações pode-se obter informações sobre o comportamento das moléculas quando submetidas a diferentes condições de pressão, temperatura e densidade do sistema. Por exemplo, as distâncias entre os átomos da molécula em vários momentos de uma evolução temporal, que podem

² <http://www.rcsb.org/pdb/home/home.do>

ser filtradas por meio de um processo ETL (*Extract, Transform and Load*) e analisadas através de ferramentas OLAP, descritas nas próximas seções.

2.3 Processo ETL

O processo de Extração, Transformação e Carga, segundo Pichiliani (2008), tem o papel de extrair os dados de diversos sistemas, transformá-los conforme as regras de negócios e carregá-los para uma base de dados, de acordo com a necessidade. Desta maneira, a partir de dados de fontes heterogêneas é possível obter dados estruturados e consistentes, que podem ser utilizados em ferramentas que fornecem suporte ao processo de tomada de decisão. A Figura 1 apresenta as etapas do processo ETL.

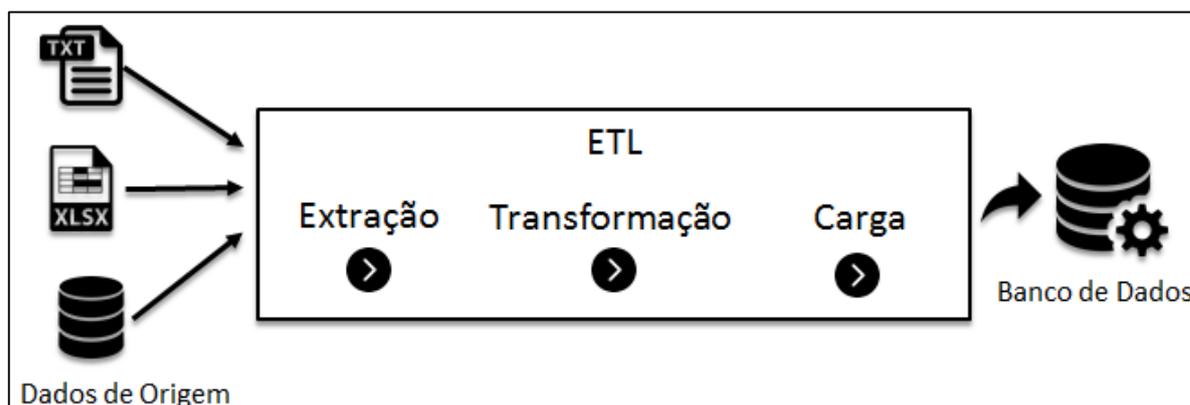


Figura 1: Processo ETL

Na etapa de extração, os dados são retirados dos sistemas de origem, que podem ser provenientes de diversas fontes como banco de dados, arquivos em texto, planilhas, páginas *html*, entre outros. Após a extração, podem existir dados desconfigurados e irrelevantes para o propósito com o qual se deseja trabalhar, sendo necessário realizar o processo de transformação e filtragem dos dados.

A etapa de transformação dos dados tem como objetivo a conversão dos dados extraídos para um formato adequado ao contexto estabelecido, para que sejam armazenados no banco de dados de destino (Turban et al. 2009). Nesta etapa são aplicadas algumas regras do negócio aos dados que foram extraídos como, por exemplo, definições de quais colunas serão extraídas, correção de erros de digitação, inconsistências e a padronização dos dados, já que os mesmos provêm de sistemas distintos e uma mesma informação pode estar em diferentes formatos.

Conforme mostra a Figura 2, podem existir diferentes classificações para definir um mesmo tipo de dado.

Masculino	Feminino
M	F
H	M
1	0

Figura 2: Padronização dos dados

Em situações como a apresentada na Figura 2, é necessário realizar a uniformização dos dados, de maneira a garantir a integridade, deixando-os em um estado ideal para que seja iniciado o processo de carga.

A última etapa do processo ETL é a fase de carga dos dados, que consiste em gravar os dados extraídos, tratados e transformados nas etapas anteriores, no banco de dados de destino. Nesta etapa ocorrem mapeamentos que respeitam as restrições de integridade para criar uma visão unificada das bases, obedecendo às descrições do dado original (OLIVEIRA, 2011). O processo de restrição de integridade garante a consistência dos dados a serem trabalhados como, por exemplo, a restrição de domínio que impede que um campo definido como chave primária no banco de dados seja preenchido com valor nulo.

O processo ETL atua como uma ponte que possibilita o envio dos dados extraídos e estruturados para as ferramentas de apoio a tomada de decisões como OLAP e *DataMining*, sendo empregado como parte integrante na coleta dos dados para tais ferramentas.

A seção a seguir apresenta o conceito sobre a tecnologia OLAP que, neste trabalho, será aplicada no processo de identificação das quebras de ligações de moléculas antocianinas.

2.4 OLAP (Processamento Analítico Online)

O termo OLAP (*Online Analytical Processing*) foi proposto por Edgar F. Codd, para descrever uma classe de tecnologias destinadas ao acesso e a consulta de dados em sistemas *online*, com o objetivo de transformar dados em informações capazes de auxiliar no processo de tomada de decisão (SINGH, 2000). Os sistemas

OLAP possibilitam aos usuários analisarem e confrontarem informações sobre diferentes pontos de vista, dado um conjunto de dados.

Possibilitar a análise das informações sobre diferentes ângulos é a principal característica dos sistemas OLAP, que está relacionada à manipulação multidimensional de dados (OLIVEIRA, 2002). Esta forma de manipulação proporciona uma análise sobre os dados de maneira mais intuitiva e flexível, possibilitando realizar o cruzamento de informações e fornecendo uma visão do todo sobre o sistema.

Em um modelo de dados multidimensional, a base de dados é uma coleção de medidas que contém valores a serem analisados como, médias, totais e quantidade. Cada medida depende de um conjunto de dimensões, que classifica ou agrupa estas medidas, determinando possíveis consultas ou cruzamentos de informações a serem realizadas (RAMAKRISHNAN; GEHRKE, 2008). Por exemplo, as dimensões baseadas em dados de vendas de uma empresa podem ser Produto, Local e Tempo, que poderiam gerar a quantidade (medida) de um produto “P” vendido em um local “L” no tempo “T”. Desta forma, é possível obter inúmeros resultados a partir de medidas que agregam informações de uma ou mais dimensões.

A multidimensionalidade encontrada nos sistemas OLAP permite que os dados sejam visualizados em diversas faces, comumente chamadas de “cubos” que, segundo Hokama (et al. 2004), armazenam todas as informações relacionadas a um determinado assunto, de maneira que sejam montadas diversas combinações entre elas, resultando na extração de visões distintas sobre o mesmo contexto. Isto permite que os usuários explorem e analisem uma coleção de dados de muitas perspectivas diferentes, de forma que cada lado do cubo apresenta uma definição, delimitando o assunto que se deseja analisar.

A Figura 3 apresenta a reprodução de um cubo OLAP, que possibilita uma visualização multidimensional dos dados. Como pode ser observado na imagem, é possível analisar um produto sobre diferentes perspectivas.



Figura 3: Cubo multidimensional OLAP para análise
Fonte: Casas, Garcia, 2007 – pág 48

A análise de dados sob diversas perspectivas possibilita ao usuário formular questionamentos para que possa ter um entendimento completo sobre o negócio em questão como, por exemplo, descobrir quais produtos possuem um maior número de vendas nos feriados ou quais produtos são mais vendidos na loja matriz. Com a visualização multidimensional em cubo, perguntas como estas podem ser respondidas facilmente, de modo a identificar tendências ou padrões acerca dos dados apresentados.

Apesar de aplicações de OLAP apresentarem os dados em visões lógicas multidimensionais, estes dados não estão necessariamente armazenados em estruturas multidimensionais. Estruturas relacionais podem ser usadas para a representação e armazenamento de dados. A Figura 4 apresenta um modelo multidimensional em um modelo relacional.

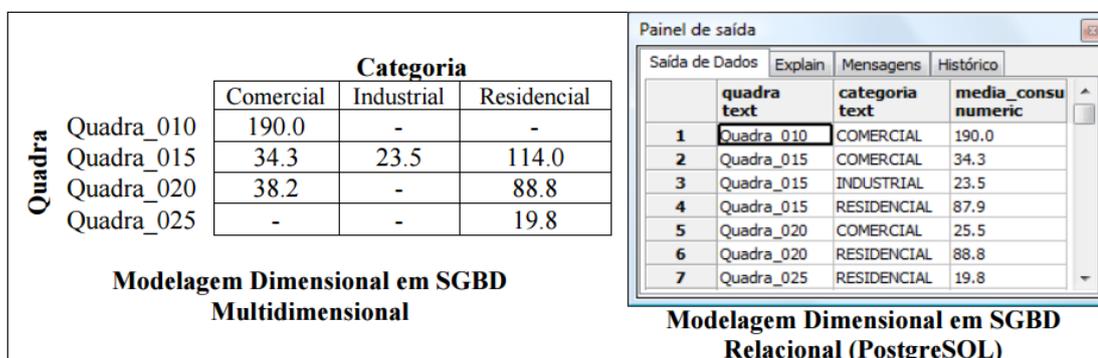


Figura 4: Modelagem multidimensional e relacional
Fonte: (GOUVEIA, 2009)

A utilização de estruturas relacionais para representar um modelo multidimensional possibilita que células que ocupam espaços em disco, mas que não possuem dados cadastrados, sejam eliminadas, como é o caso das quadras 10, 20 e 25 apresentadas no modelo multidimensional na Figura 4.

Os sistemas OLAP não são caracterizados como uma tecnologia de armazenamento de dados e sim uma forma de cruzamento e visualização de dados para os usuários. No entanto, essas ferramentas utilizam dados armazenados em uma base de dados no qual podem ser classificados segundo a forma de armazenamento.

2.4.1 Classificação quanto aos métodos de armazenamento

Os sistemas OLAP podem ser caracterizados quanto à forma como os dados são armazenados e manipulados. Turban et al. (2009) os classificam em:

- **MOLAP (Sistema OLAP Multidimensional):** os dados são armazenados em um banco de dados multidimensional, no qual pode-se trabalhar com várias perspectivas de uma mesma base de dados como, por exemplo, analisar a taxa de rendimento escolar de instituições públicas por período, localidade, faixa etária, dentre outras medidas (Turban et al. 2009). Todo o processo de manipulação das dimensões é realizado diretamente no servidor, o que proporciona uma rapidez nas consultas ou análises dos dados. Estes dados são previamente processados antes de serem apresentados pelo sistema, o que dificulta a inserção de novos dados ao modelo em questão, limitando o número de dados com o qual o sistema pode trabalhar.
- **ROLAP (Sistema OLAP Relacional):** os dados são armazenados em um banco de dados relacional (Turban et al. 2009), ou seja, os dados são guardados em tabelas relacionadas entre si. Neste tipo de sistema o usuário envia um pedido de análise multidimensional, que é convertido em declarações do tipo SQL (*Structured Query Language*) e apresentada ao banco de dados. Este sistema é capaz de lidar com um grande número de dados, porém, um dos problemas encontrados está relacionado ao seu desempenho, que pode se tornar baixo, pelo fato dos relatórios serem gerados a partir de consultas SQL no banco

de dados e o tempo de resposta desta consulta pode ser longo dependendo do tamanho da base de dados (BONEL, 2015).

- **DOLAP (Desktop OLAP):** os dados são armazenados na máquina do cliente, podendo ser um banco de dados relacional ou multidimensional (Turban et al. 2009). Desta maneira, o cliente realiza uma análise local sobre os dados, evitando uma sobrecarga no servidor e reduzindo o tráfego na rede. A desvantagem deste tipo de sistema está na limitação do tamanho do cubo, já que todo o processo é realizado na máquina do usuário e a mesma pode não suportar a análise de grandes quantidades de dados em função de sua configuração.
- **WOLAP (Web OLAP):** os dados são armazenados em um servidor, podendo ser um banco de dados relacional ou multidimensional e estão acessíveis por meio de um navegador *web*, permitindo o acesso remoto e possibilitando a análise dos dados independentemente da plataforma utilizada pelo usuário (Turban et al. 2009).

A definição de qual tipo de sistema OLAP será empregado não é uma tarefa fácil, pois é preciso levar em conta alguns elementos como o volume de dados a ser manipulado, a complexidade das consultas a serem realizadas, a plataforma no qual será aplicada, dentre outros fatores.

Os sistemas OLAP fornecem um modelo para que as ferramentas OLAP sejam desenvolvidas, proporcionando assim condições ideais para a análise dos dados. A seção a seguir apresenta as principais características de uma ferramenta OLAP.

2.4.2 Ferramentas OLAP

As ferramentas OLAP podem ser definidas como uma categoria de *softwares* cujo principal objetivo é oferecer recursos de modelagem, análise e visualização de grandes conjuntos de dados. Elas possibilitam aos usuários analisar e navegar sobre os dados para detectar tendências, exceções e obter detalhes para entender melhor os pontos fortes e fracos de uma determinada atividade (SINGH, 2000).

A capacidade destas ferramentas em retornar informações relevantes sobre um grande volume de dados proporciona uma agilidade na maneira como eles serão

analisados. Por isso, atuam como uns dos componentes principais no processo de análise de informações para as ferramentas de BI (*Business Intelligence*) permitindo o levantamento de questões sobre o negócio de modo a obter respostas de maneira rápida e concreta (BATISTA, 2004 apud REGINATO; NASCIMENTO, 2007).

Os dados apresentados pelas ferramentas OLAP podem ser analisados de acordo com a necessidade do usuário. SINGH (2001) destaca alguns dos principais recursos encontrados nestas ferramentas no que se refere à análise dos dados:

- *Slice and dice*: permite aos usuários fatiar as informações para visualizar os dados a partir de várias perspectivas, na qual podem “cortar” ou “girar” determinados dados ao longo de qualquer dimensão (SINGH, 2000). Esta operação possibilita modificar a posição de uma informação bem como, alternar entre linhas e colunas de maneira a facilitar a compreensão dos dados;
- *Drill down/up*: consiste em realizar exploração em diferentes níveis de detalhes da informação (SINGH, 2000). O *drill down* refere-se ao aumento do nível de detalhe da informação. Por exemplo, considerando um conjunto de dados sobre vendas de uma empresa, no qual o usuário deseja obter informações mais detalhadas, sobre qual vendedor realizou a venda de um determinado produto ou quais itens compõe esta venda, assim é possível ter uma visão mais detalhada sobre os dados. Já o *drill up* é quando o usuário diminui o nível de detalhe da informação, por exemplo, visualizar todos os vendedores de uma loja, considerando apenas o total de vendas de cada vendedor;

Pode-se observar que o uso de uma ferramenta OLAP permite realizar diferentes análises sobre um mesmo dado, porém a maneira como os dados são manipulados é que faz toda a diferença na hora de obter os resultados desejados.

Atualmente, as ferramentas OLAP podem ser empregadas a diferentes contextos como, por exemplo, pode ser utilizada por um comércio para realizar análises de vendas por diferentes níveis (por região, produtor, vendedor etc.). Outro exemplo de uso é em instituições educacionais, que podem usar as ferramentas OLAP para analisar o perfil de alunos e/ou seus resultados durante a aplicação do exame nacional do ensino médio ou superior, possibilitando uma avaliação da qualidade do ensino.

3. METODOLOGIA

O presente trabalho teve como objetivo utilizar a ferramenta OLAP *Microstrategy Analytics Desktop*, para identificar os momentos em que as distâncias de ligações entre as moléculas fogem do padrão indicados nos valores de referência, o que pode indicar que ocorreram quebras nas moléculas de antocianinas durante o processo de simulação da dinâmica molecular, tendo como base para comparação os valores de referência disponibilizados pelo NIST.

As seções a seguir apresentam as tecnologias de *hardwares* e *softwares* que foram utilizados, bem como os métodos aplicados para o desenvolvimento do trabalho.

3.1 Hardware

Para a realização das simulações de dinâmica molecular foi utilizado um *cluster* de computadores, composto por dois nós (um mestre e um escravo):

- Nó mestre: processador *Intel Core i5-4570 CPU @ 3.20GHz*x4 e memória de 4GB;
- Nó escravo: *Intel Celeron CPU 420 @ 1.60GHz* e memória de 500MB.

O *cluster* utilizado foi desenvolvido em outro trabalho relacionado ao projeto Quanto, e encontra-se pronto em funcionamento no LaRC (Laboratório de Redes de Computadores do CEULP/ULBRA), para que sejam realizadas as simulações.

Ainda, foi utilizado um computador pessoal para aplicação da ferramenta OLAP e realização do processo de extração, transformação e carga dos dados.

3.2 Software

Para a realização das simulações de dinâmica molecular, extração, transformação e carga dos dados e análise dos resultados, foram utilizados os *softwares*:

- Gromacs: pacote de *softwares* que oferece recursos para realizar simulação da dinâmica molecular, projetado para trabalhar com moléculas, proteínas e lipídios. Este não possui interface gráfica e sua utilização é feita apenas por linhas de comando (SPOEL et al., 2005). O *software* foi executado no *cluster*, que paraleliza a execução da

simulação, que melhora o desempenho da simulação, reduzindo o tempo de execução das simulações;

- *UCSF Chimera*: ferramenta *open source* multiplataforma, desenvolvida pelo Grupo de Pesquisas em Biocomputação, Visualização e Informática (*Resource for Biocomputing, Visualization and Informatics*) da Universidade da Califórnia, São Francisco - EUA. Permite a visualização gráfica e análise de estruturas moleculares. Oferece recursos para captar distâncias entre átomos, calcular centroides, executar trajetória de dinâmica molecular e navegar entre as estruturas através de movimentações em três dimensões (UCSF Chimera, 2015). Esta ferramenta será utilizada para a retirada das distâncias de ligações entre os átomos, resultantes da simulação de dinâmica molecular;
- *Kettle Spoon*: ferramenta para a realização do processo de limpeza, transformação e carga dos dados extraídos durante o processo de simulação de dinâmica molecular. Permite a manipulação de dados através de múltiplas fontes (PENTAHO, 2015). Os dados podem ser extraídos de fontes como, arquivos *CSV*, *TXT*, *XML*, planilhas ou tabelas de um banco de dados;
- *Microstrategy Analytics Desktop*: ferramenta OLAP capaz de gerar gráficos, grades de planilhas interativas, mapas de calor, entre outros recursos, permitindo aos usuários pesquisarem e realizarem análises sobre os dados a partir de uma grande variedade de fontes, incluindo *Data Warehouses*, arquivos do Excel, entre outros (MICROSTRATEGY, 2014). Esta ferramenta oferece recursos como o *slice and dice*, que permite realizar diferentes cruzamentos entre os dados e navegar por vários níveis de detalhes;

3.3 Métodos

A Figura 5 apresenta as etapas realizadas para o desenvolvimento do projeto.

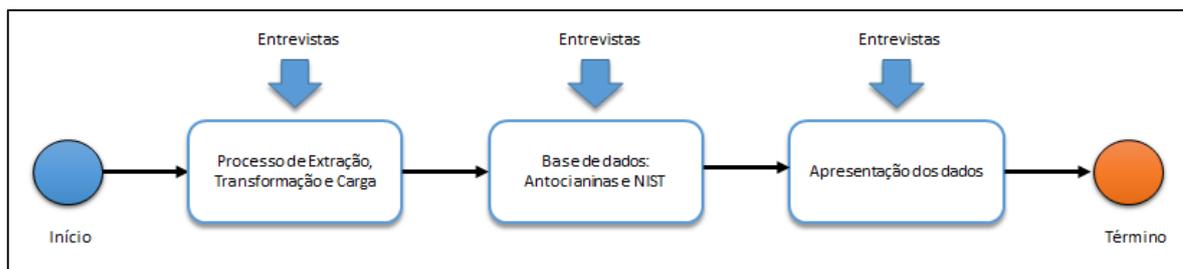


Figura 5: Etapas para o desenvolvimento do projeto

Para a elaboração deste projeto, foram realizadas entrevistas com a especialista da área de Química para obter informações a respeito das necessidades e os desafios encontrados no projeto Quanto. O principal desafio apontado pela especialista está em identificar o momento em que ocorrem as quebras das ligações das moléculas, fazendo com que elas percam seu poder antioxidante.

Foram realizadas entrevistas com a especialista da área e definiu-se que seriam utilizados os valores de referência do NIST, no que se refere as distâncias mínimas e máximas aceitáveis para as ligações da molécula. Os valores foram comparados aos dados retornados da simulação da dinâmica molecular realizada pela ferramenta Gromacs, que correspondem as ligações obtidas durante o processo de simulação. A especialista da área acompanhou todo o desenvolvimento do projeto.

O *software* Gromacs possibilita simular diferentes ambientes e compostos, desta forma pode-se obter diferentes resultados durante o processo de simulação. Porém, as simulações foram feitas em um único ambiente, alterando apenas as configurações de entradas do sistema, que estão disponíveis no Apêndice A. Antes da simulação, foi feita a preparação do ambiente, mostrada na seção 3.4. O resultado das simulações são arquivos que contém os dados da trajetória da molécula, com o qual podem ser extraídas as distâncias ligações entre os átomos.

Após a obtenção dos dados, foi realizado o processo ETL:

- **Extração:** retirada dos dados resultantes da simulação de dinâmica molecular sobre as antocianinas, extraindo apenas as distâncias formadas entre os átomos durante o processo de simulação.
- **Transformação:** processo de limpeza e transformação dos dados da simulação, com a retirada de dados irrelevantes para a análise.

- **Carga:** os dados foram carregados para uma base de dados relacional, juntamente com os dados extraídos do NIST.

Após o processo ETL, os dados da base relacional foram carregados para a ferramenta OLAP com o objetivo de identificar o momento em que as distâncias fogem do padrão, de acordo com os dados do NIST, o que pode indicar uma possível quebra de ligações entre as moléculas, resultado este que será visto pelo especialista da área. Os resultados foram apresentados por meio de gráficos e planilhas dinâmicas.

Como última etapa do projeto, as informações obtidas com a utilização da ferramenta OLAP foram apresentadas aos especialistas da área de Química, vinculadas ao projeto Quanto.

3.4 Preparação do ambiente para obtenção dos dados

No início deste projeto, a ferramenta Gromacs já estava instalada nas suas configurações padrões, descritas na página oficial da ferramenta³, e executando em um *cluster* de computadores. Porém, a ferramenta havia sido usada apenas para testes com moléculas genéricas disponibilizadas no banco de dados do RCSB, assim, foi necessário realizar alguns ajustes para a simulação com as antocianinas de pimenta.

Na tentativa de realizar a simulação de dinâmica molecular foi identificado que o Gromacs não era capaz de analisar a estrutura inicial da cianidina-3-glicosídeo e determinar qual campo de força, apresentado na seção 2.2, seria empregado. Assim, não foi possível gerar a topologia da molécula, arquivo fundamental para a simulação, que contém as coordenadas iniciais dos átomos que compõem a estrutura da molécula.

Para que a topologia pudesse ser gerada, foi adicionado um *script* desenvolvido na linguagem Perl por Lemkul (2015) e disponível na seção de contribuições do Gromacs⁴, este está disponibilizado no Anexo II. A função do *script* é analisar a estrutura de uma molécula, descrita em um arquivo no formato PDB (*Protein Data Bank*), que permite uma representação padrão para dados de

³ http://www.gromacs.org/Documentation/Installation_Instructions_5.0

⁴ http://www.gromacs.org/Downloads/User_contributions/Other_software

estruturas moleculares. Desta forma, o *script* determina a partir do arquivo PDB qual o campo de força irá atuar na molécula durante o processo de simulação.

O *script* utilizado foi alterado para ser executado na versão 5.0 do Gromacs, utilizada no projeto, modificando a biblioteca **FFOPLSAA**, a qual faz referência ao campo de força **OPLSAA**, renomeando para **FORCEFIELD**. A alteração consistiu na substituição de um trecho de código, conforme apresenta a Figura 6. A figura apresenta apenas o trecho do *script* que foi alterado: 1) *script* com a antiga biblioteca; e 2) *script* com a nova biblioteca.

The figure shows two side-by-side code snippets. The left snippet, titled '1) Script Antigo', contains a Perl code block with an if-statement. The second line of the code is highlighted with a red box and reads: `print HEADER_OUT ";#include \"ffoplsaa.itp\"";`. The right snippet, titled '2) Script Novo', contains a similar Perl code block. Its second line is also highlighted with a red box and reads: `print HEADER_OUT ";#include \"oplsaa.ff/forcefield.itp\"";`. Both snippets start with `if ($out type eq "top")` and end with a closing curly brace.

Figura 6: Script Perl - Topologia

Na Figura 6, pode-se observar, na parte em destaque dos *scripts*, as referências para as bibliotecas do campo de força presentes no Gromacs.

4. RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os resultados na simulação da dinâmica molecular com a molécula cianidina-3-glicosídeo, como foram definidos os valores de referência do NIST, de que maneira o processo ETL foi aplicado para a retirada das distâncias e, por fim, como estes dados foram inseridos na ferramenta OLAP *Microstrategy Analytics Desktop* e apresentados para análise.

4.1 Base de dados experimentais do NIST

Os dados disponibilizados pelo NIST são fornecidos por cientistas cadastrados, com experimentos validados, em várias áreas da ciência (NIST, 2016). A grande área da química recebe dados tanto experimentais quanto calculados.

Foram extraídos do NIST dados que fornecem valores de referência para as distâncias de ligações de moléculas com configurações parecidas às da cianidina-3-glicosídeo, apresentados na Figura 7. A figura apresenta: 1) a estrutura molecular da antocianina; 2) estrutura molecular do NIST correspondente a uma das ligações da antocianina; e 3) a tabela com os valores aceitáveis disponibilizados pelo NIST, com a descrição dos átomos e o valor para cada ligação presente na estrutura molecular. Por exemplo, para a ligação entre os átomos Carbono (C) e Hidrogênio (H), o valor a ser considerado é de 1.078Å(Angström).

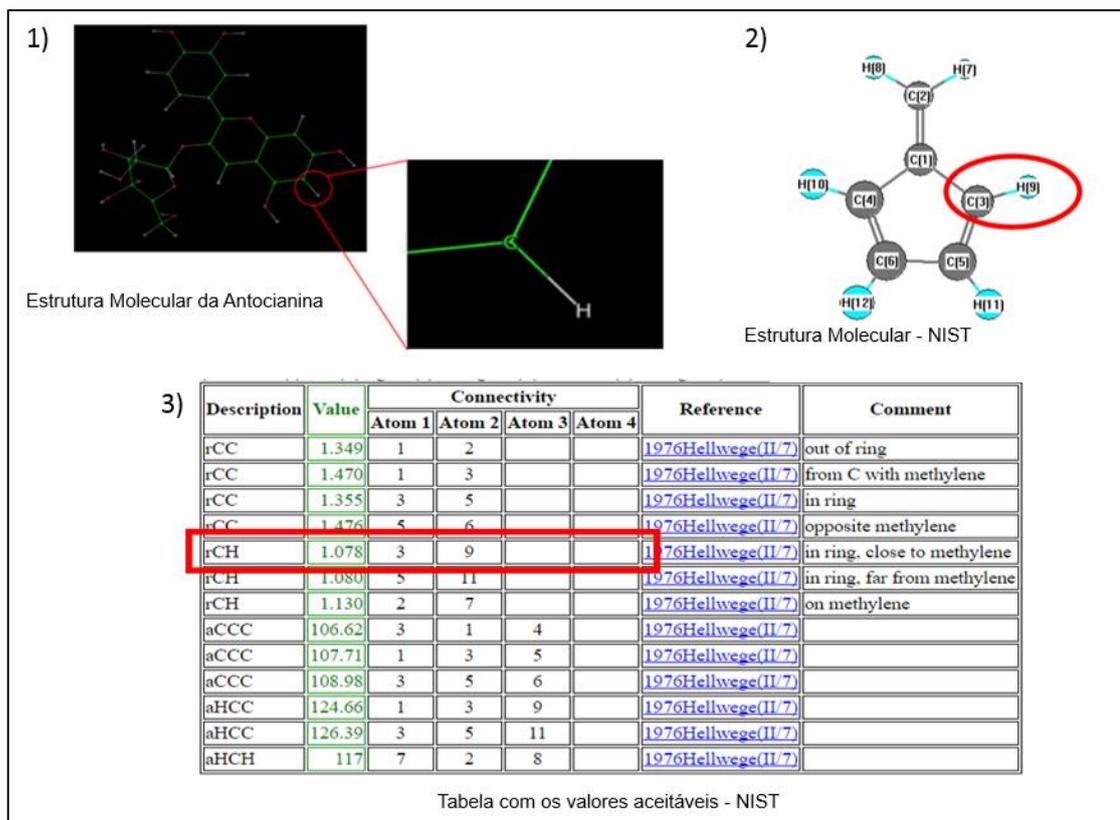


Figura 7: Estrutura molecular: Antocianina; Estrutura Molecular - NIST; Tabela com os valores aceitáveis

Foram extraídas apenas as distâncias, que representam as ligações químicas presentes na molécula utilizada no projeto. A tabela 1 apresenta as distâncias aceitáveis para as ligações presentes na molécula. Estes dados foram carregados para a base de dados relacional.

Tabela 1: Valores para as ligações da antocianina

Ligações	Valores(A)
C-H	1.078
C-C	1.531
C=C	1.398
C-O	1.420
O-H	0.956

* (-) Ligação simples

* (=) Ligação dupla

*C – Carbono

*H – Hidrogênio

*O - Oxigênio

Os dados da tabela apresentam os valores em Angström para as distâncias presentes na molécula cianidina-3-glicosídeo.

4.2 Resultados da dinâmica molecular

A dinâmica molecular consistiu em seis etapas:

1. Configurações Iniciais
2. Preparação da estrutura
3. Definição do ambiente de simulação
4. Minimização de Energia
5. Equilibração
6. Produção de Dinâmica Molecular

A fase de configurações iniciais consistiu em desenvolver a estrutura molecular da antocianina, indicando os átomos e as ligações que compõem a molécula, em um arquivo no formato PDB, conforme apresenta a Figura 8.

COMPND	UNNAMED									
AUTHOR	GENERATED BY OPEN BABEL 2.3.2									
HETATM	1	C	LIG	1	-6.625	7.998	-0.790	1.00	0.00	C
HETATM	2	C	LIG	1	-5.575	8.662	-0.138	1.00	0.00	C
HETATM	3	C	LIG	1	-4.293	8.691	-0.709	1.00	0.00	C
HETATM	4	O	LIG	1	-3.278	9.327	-0.088	1.00	0.00	O
HETATM	5	O	LIG	1	-7.858	7.969	-0.242	1.00	0.00	O
HETATM	6	C	LIG	1	-6.382	7.362	-2.021	1.00	0.00	C
HETATM	7	H	LIG	1	-5.754	9.154	0.811	1.00	0.00	H
HETATM	8	H	LIG	1	-8.050	8.399	0.604	1.00	0.00	H
HETATM	9	H	LIG	1	-3.402	9.777	0.759	1.00	0.00	H
HETATM	10	C	LIG	1	-5.104	7.391	-2.590	1.00	0.00	C
HETATM	11	C	LIG	1	-4.062	8.055	-1.934	1.00	0.00	C
HETATM	12	O	LIG	1	-4.869	6.777	-3.780	1.00	0.00	O
HETATM	13	H	LIG	1	-3.074	8.077	-2.376	1.00	0.00	H

Figura 8 – Molécula Antocianina: cianidina-3-glicosídeo formato PDB

Na Figura 8, pode-se observar no trecho em destaque, o átomo, a sua ligação e suas posições nas dimensões x , y e z , em relação ao centro da estrutura molecular. Este arquivo pode ser utilizado por programas gráficos como o PyMOL para apresentar uma imagem tridimensional da molécula, conforme mostra a Figura 9.

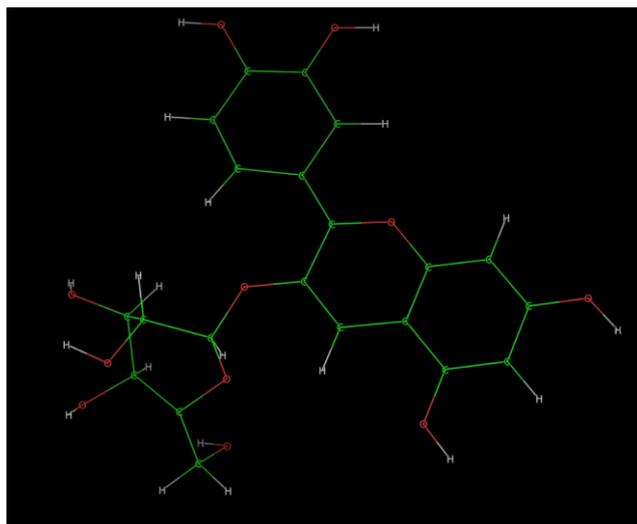


Figura 9: Molécula Antocianina: cianidina-3-glicosídeo

A Figura 9 apresenta a imagem tridimensional da estrutura inicial da cianidina-3-glicosídeo gerada através do *software* PyMOL, a molécula possui 52 átomos e 55 ligações e foi desenvolvida por pesquisadores do CEULP/ULBRA.

Depois, foi realizada a preparação do ambiente descrita na seção 3.4. Nesta etapa o campo de força definido foi OPLS (*Optimized Potential for Liquid Simulations*), que é utilizado para simulações que envolvem potenciais otimizados em líquidos, pois a análise sobre a cianidina-3-glicosídeo foi com a presença do solvente do tipo água (MARK, 2007).

Na definição do ambiente de simulação é pré-determinado que seja em uma caixa virtual, conforme explicado na seção 2.2. Para o trabalho, foi definida uma caixa cúbica, preenchida com o solvente de água do tipo TIP3P, que corresponde a uma solução aquosa para a interação entre os átomos presentes na molécula.

A etapa de minimização do sistema tem como objetivo remover as possíveis tensões na configuração inicial da molécula, conforme explicado na seção 2.2. As tensões ocorrem quando os átomos da molécula estão muito próximos, assim, é necessário aumentar as distâncias entre esses átomos. Nesta etapa, o Gromacs indicou quais eram as tensões que precisavam ser removidas e corrigiu as distâncias entre os átomos da molécula. As distâncias consideradas pequenas são configuradas por padrão na ferramenta e não são alteradas.

Na etapa de equilíbrio da antocianina, que visa obter uma conformação favorável do sistema, analisando pontos como temperatura, energia total, densidade etc., conforme explicado na seção 2.2, foi realizada a equilíbrio em duas fases. A

primeira consistiu em analisar o NPT (*Number of Particules, Volume and Temperature*), que corresponde a equilibração da temperatura da molécula. Na segunda fase foi analisada a densidade e a pressão da molécula, por meio do NVT (*Number of Particules, Pressure and Temperature*).

A etapa de produção da dinâmica molecular envolve a integração dos átomos, provocando o seu deslocamento, conforme explicado na seção 2.2. Foram definidos 50 mil passos e 4 nanosegundos para a simulação. Como resultado da simulação foi obtida a trajetória descrita pelos átomos, que corresponde as suas movimentações através da sua evolução temporal. A Figura 10 apresenta o conteúdo do arquivo **trajetoria.pdb**, que foi extraído no processo de simulação da molécula de antocianina que está sendo estudada.

REMARK	GENERATED BY TRJCONV									
TITLE	LIG topology, generated by TopolGen in water t= 50.00000									
REMARK	THIS IS A SIMULATION BOX									
CRYST1	32.831	32.831	32.831	90.00	90.00	90.00	P 1			1
MODEL	1									
ATOM	1	C	LIG	1	16.705	19.166	20.498	1.00	0.00	C
ATOM	2	C	LIG	1	17.337	20.388	20.727	1.00	0.00	C
ATOM	3	C	LIG	1	17.629	21.167	19.561	1.00	0.00	C
ATOM	4	O	LIG	1	19.713	33.949	29.969	1.00	0.00	O
ATOM	5	O	LIG	1	16.045	18.894	21.672	1.00	0.00	O
ATOM	6	C	LIG	1	17.094	18.395	19.303	1.00	0.00	C
ATOM	7	H	LIG	1	18.020	20.522	21.554	1.00	0.00	H
ATOM	8	H	LIG	1	17.198	19.051	15.457	1.00	0.00	H
ATOM	9	H	LIG	1	27.779	33.363	36.253	1.00	0.00	H
ATOM	10	C	LIG	1	18.052	18.918	18.388	1.00	0.00	C
ATOM	11	C	LIG	1	18.444	20.242	18.704	1.00	0.00	C
ATOM	12	H	LIG	1	19.491	20.265	18.893	1.00	0.00	H
ATOM	13	C	LIG	1	17.994	17.122	16.515	1.00	0.00	C
ATOM	14	C	LIG	1	18.698	16.438	15.392	1.00	0.00	C
ATOM	15	C	LIG	1	16.661	16.901	17.147	1.00	0.00	C
ATOM	16	C	LIG	1	16.519	17.294	18.526	1.00	0.00	C
ATOM	17	O	LIG	1	15.461	16.373	16.547	1.00	0.00	O
ATOM	18	H	LIG	1	15.145	17.891	17.723	1.00	0.00	H

Figura 10: arquivo trajetoria.pdb

A Figura 10 apresenta parte do arquivo extraído, que possui cerca de 3424 átomos e 2350 ligações. O número de átomos e ligações podem variar de acordo com o número de passos e o tempo de integração dos átomos definidos na produção de dinâmica molecular. Analisando o arquivo, é possível observar o deslocamento dos átomos e extrair as distâncias entre eles. Por exemplo, na Figura 9, o trecho destacado apresenta o átomo, a sua ligação e as suas posições nas dimensões x, y e z, em relação ao centro de sua estrutura molecular.

O arquivo no formato PDB possibilita que as ferramentas realizem análises por meio de *frames*, isto é, uma análise quadro a quadro dos resultados da

simulação. Desta forma, eles poderão ser analisados separadamente para um estudo mais detalhado.

Durante a execução desta simulação foi observado que a molécula sofreu algumas oscilações, o que fez com que ela perdesse parte de sua estrutura molecular. Isto resultou em distâncias com valores fragmentados, sendo necessário realizar uma nova simulação.

A segunda simulação teve um tempo inferior, cerca de 2 nanosegundos, com o qual é possível identificar a quebra da molécula, porém com o dobro de passos executados na primeira simulação, cerca de 100 mil passos. Nesta simulação analisou-se somente a etapa de minimização de energia, para descobrir quais fatores provocaram as oscilações encontradas na primeira simulação, a partir deste ponto foram retiradas as distâncias de ligações da molécula, para a análise dos especialistas da área.

A partir dos dados do arquivo PDB, foi criado o banco de dados relacional, no processo ETL descrito a seguir.

4.3 Processo ETL

4.3.1 Extração

No processo de extração dos dados, o arquivo **trajetoria.pdb**, resultante da simulação de dinâmica molecular, foi carregado para a ferramenta *UCSF Chimera* apresentada na seção 3.2. Essa ferramenta apresenta a visualização gráfica das estruturas moleculares e oferece recursos para extrair as distâncias entre os átomos.

As distâncias foram extraídas através do comando 1, com o qual são selecionados todos os átomos de um *frame* específico da molécula.

```
~distance all
```

Comando 1: extrair todas as distâncias

As informações das distâncias são retornadas em um arquivo no formato TXT, apresentado na Figura 11.

```

Model #0 is cianidina_otim.pdb
Distance information
1  LIG 1.het C <-> LIG 1.het H:  1.0707
2  LIG 1.het C <-> LIG 1.het C:  1.3421
3  LIG 1.het H <-> LIG 1.het C:  1.0865
4  LIG 1.het C <-> LIG 1.het C:  1.4838
5  LIG 1.het O <-> LIG 1.het C:  1.3520
6  LIG 1.het O <-> LIG 1.het H:  0.9677
7  LIG 1.het C <-> LIG 1.het C:  1.3423
8  LIG 1.het O <-> LIG 1.het C:  1.3516
9  LIG 1.het H <-> LIG 1.het O:  0.9679
10 LIG 1.het C <-> LIG 1.het C:  1.5055

```

Figura 11: Distâncias extraídas

Na Figura 11 pode-se observar um trecho das ligações presentes na molécula e suas respectivas distâncias. A parte em destaque apresenta os átomos, sua ligação e o valor da distância entre eles em relação à estrutura molecular. No processo descrito a seguir, as informações passaram por um processo de transformação e foram carregadas para uma base de dados relacional.

4.3.2 Transformação e Carga (Load)

Para o processo de transformação e carga das distâncias, foi utilizada a ferramenta *Kettle – Spoon*, apresentada na seção 3.2. Como mostra a Figura 12, o processo consistiu nas tarefas de leitura, limpeza e inserção dos dados.

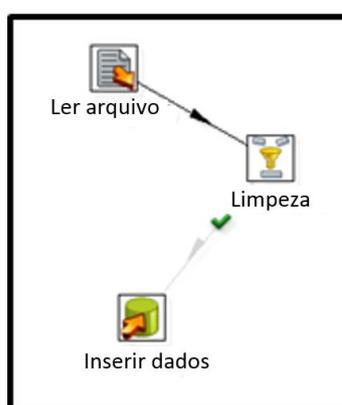


Figura 12: Processo de leitura, limpeza e inserção dos dados

O processo iniciou-se com a leitura de arquivo no formato TXT, mostrado na Figura 11, obtido na etapa de extração dos dados. Depois, foi realizada a limpeza de registros indesejados, dentre eles estão: “**LIG 1.HET**”, “<->”, “:”. Assim, restaram

apenas os dados referentes aos átomos e as suas respectivas distâncias. A tarefa foi encerrada com a inserção dos registros na base de dados relacional.

Para o armazenamento dos dados extraídos foi elaborado o modelo conceitual apresentado na Figura 13:

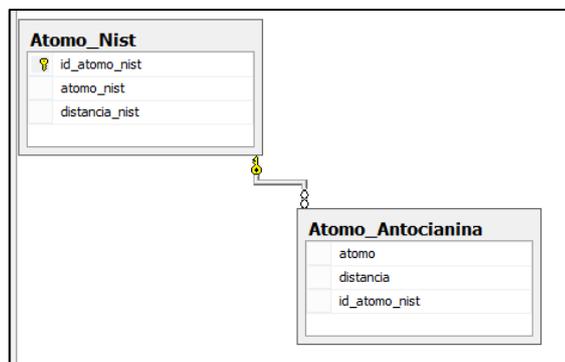


Figura 13: Modelo conceitual

A Figura 13 apresenta o modelo conceitual da estrutura elaborada para armazenar os dados extraídos, com as principais entidades e atributos. A entidade **Atomo_Nist** representa as distâncias extraídas do NIST, já a entidade **Atomo_Antocianina** representa as distâncias extraídas durante o processo de simulação. O relacionamento entre as tabelas é de “1” para “n”, sendo que uma distância do NIST abrange diversas ligações da molécula, porém uma ligação entre os átomos da molécula possui apenas uma distância específica.

Após o processo ETL, os dados armazenados na base de dados relacional foram carregados para a ferramenta OLAP.

4.4 OLAP

4.4.1 Uso da ferramenta OLAP – *Microstrategy Analytics Desktop*

A base de dados relacional foi inserida na ferramenta OLAP *Microstrategy Analytics Desktop*, que oferece recursos como: *slice and dice* e *drill up/down*, apresentados na seção 2.4.2.

Utilizando o recurso *drill up*, pode-se determinar que apenas algumas ligações sejam analisadas, aumentando o nível de detalhe da informação como, por exemplo, filtrando apenas as distâncias que contém os átomos de carbono da molécula, conforme mostra a Figura 14.

Atomo	Distanc	Distancia Nist
C C	1.102	1.531
C C	1.200	1.531
C C	1.342	1.531
C C	1.342	1.531
C C	1.593	1.531
C C	1.600	1.531
C C	1.620	1.531
C C	1.685	1.531
C C	1.698	1.531
C C	1.704	1.531
C C	1.721	1.531
C C	1.752	1.531
C C	1.779	1.531
C C	1.780	1.531
C C	1.782	1.531
C C	1.784	1.531
C C	1.786	1.531
C C	1.792	1.531
C C	1.850	1.531
C C	1.892	1.531
C C	1.995	1.531

Figura 14 - *drill up* - distâncias

Os dados apresentados na Figura 14 podem ser visualizados também de forma gráfica, conforme mostra a Figura 15. Como pode ser observado no trecho destacado, o valor de referência do NIST para a ligação é de 1.531, existem 3 ligações de átomos de carbono (C C), sendo que uma delas ultrapassou o valor de referência e está com distância de 1.600.

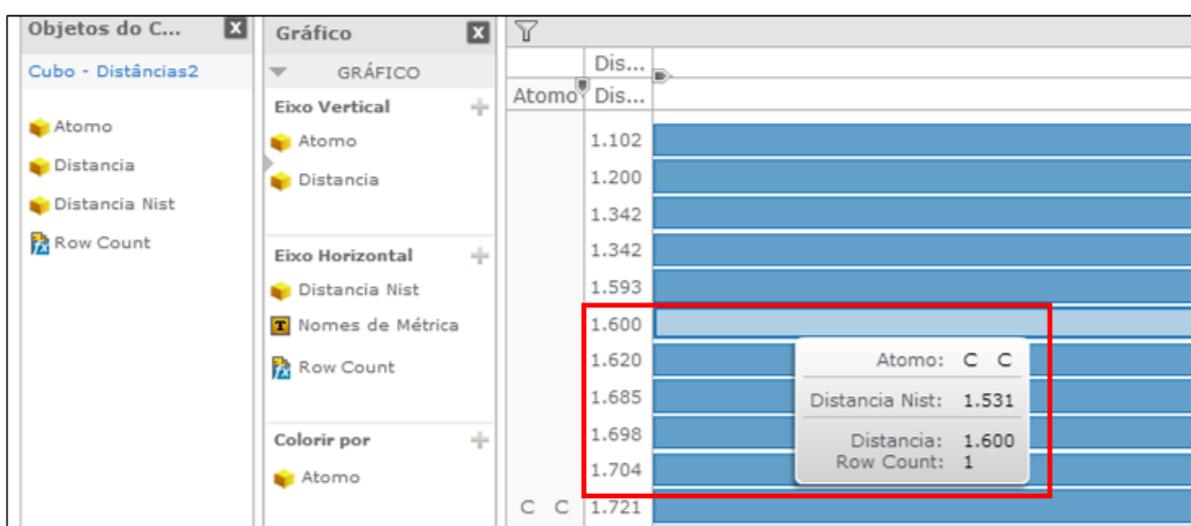


Figura 15: Resultado ferramenta OLAP

Para cada ligação presente na molécula são determinados valores máximos e mínimos, de maneira que, se uma ligação ultrapassar o valor máximo, pode indicar um possível rompimento entre as ligações, caso contrário, pode-se inferir que ela se manteve estável durante o processo de simulação, respeitando o valor mínimo estabelecido. Estes valores são determinados pelo NIST e apresentam um comparativo entre os resultados de diversos experimentos, para que os resultados obtidos sejam validados experimentalmente.

No exemplo apresentado na Figura 15, pode-se observar que a ligação ultrapassou o limite máximo estabelecido para a ligação entre os átomos de carbono, que possui o valor máximo de 1.557, o que pode indicar que a ligação tenha sofrido um rompimento a partir deste ponto. Com base nestas informações, podem ser adicionados novos componentes para que a molécula se mantenha estável durante o processo de simulação, obedecendo os limites definidos para cada ligação.

5. CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo utilizar a ferramenta OLAP *Microstrategy Analytics Desktop* para identificar modificações nas distâncias nas ligações químicas em moléculas de antocianinas de pimenta malagueta extraídas no CEULP/ULBRA, com base nos valores de referência disponibilizados pelo NIST. Os resultados apresentados com recursos gráficos auxiliarão os pesquisadores da área de Química/Farmácia do CEULP, vinculados ao projeto Quanto, no processo de análise dos dados.

Os procedimentos executados serão validados futuramente em laboratório, assim a importância da simulação é verificar de forma rápida e barata várias possibilidades, indo para o laboratório apenas quando identificar situações relevantes.

O processo ETL desenvolvido no trabalho atuou como um dos fatores fundamentais, para que os dados pudessem ser carregados de maneira consistente para a ferramenta OLAP. Nesta fase pode-se destacar a etapa de transformação, que permitiu com que os dados extraídos, pudessem ser formatados de acordo com o padrão estabelecido para a análise das distâncias.

Os resultados retornados pela ferramenta OLAP possibilitarão aos pesquisadores identificar quais ligações apresentaram valores fora do padrão estabelecido pelo NIST, já que para cada ligação são determinados valores máximos e mínimos, identificando o momento da quebra de ligação na molécula. Com isso poderão descobrir o que desestabiliza as moléculas, estabilizá-las e utilizá-las em formulações cosméticas.

Como trabalhos futuros, pretende-se automatizar todo o processo de construção da base de dados relacional, de forma que os pesquisadores consigam realizar o processo sem o auxílio de alguém de Informática.

Espera-se criar processos para a realização de novas análises e também adicionar novos compostos no processo de simulação, como por exemplo, o uso do álcool como solvente.

As configurações geradas nos frames podem ser estudadas a nível quântico, para elucidar os eventos ligados a estrutura eletrônica, cargas atômicas e energias envolvidas na formação e quebra de ligações.

6. REFERÊNCIAS

BATISTA E. O. **Sistemas de informação**. São Paulo: Saraiva, 2004.

BONEL, Claudio. **Afinal, o que é Business Intelligence?:** Micros, pequenas, médias e grandes Empresas com o poder da informação nas mãos. Rio de Janeiro: Clube dos Autores, 2015. 133 p.

BUYA, R, **High Performance Cluster Computing: Architectures and Systems**, Prentice Hall, volume 1 (1999).

CAIÇARA JUNIOR, Cícero. **Sistemas Integrados de Gestão ERP**. 3. ed. Curitiba: Editora Ibpex, 2008. 197 p.

CASAS, Alexandre Luzzi Las.; GARCIA, Maria Tereza.; **Informação em Marketing**. São Paulo: 2007. 336 p. Novatec Editora.

CONTI, Fabieli de. **“Grades Computacionais para Processamento de Alto desempenho”**. 2009, Laboratório de Sistemas de Computação – LSC. Universidade Federal de Santa Maria (UFSM) - Santa Maria, RS – Brasil. Disponível em: <<http://www-usr.inf.ufsm.br/~andrea/elc888/artigos/artigo3.pdf>>, Acesso em: set2014.

COSTA, Gil da (Org.). **Física: Tendências e Perspectivas**. São Paulo: Editora Livraria da Física, 2005. 342 p.

D. Frenkel and B.Smit. **Understanding Molecular Simulation**. Academic Press, San Diego, 1^a edition, 1996.

DANI, Sílvia. **Simulação Computacional por Dinâmica Molecular de Propriedades de Equilíbrio e Espectroscópias de Misturas Líquidas CS₂/C₆H₆**. 2003. 141 f. Tese (Doutorado) - Curso de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003. Disponível em:

<http://www.lume.ufrgs.br/bitstream/handle/10183/4107/000407616.pdf?sequence=1&locale=pt_BR>. Acesso em: 03 set. 2015.

FERNANDES, Fernando M.S Silva. Simulação de microagregados moleculares: O método da Dinâmica Molecular. **Computadores em Química**, p.59-62, dez. 1985. – arquivo: Simulação de microagregados.pdf

GOUVEIA, Roberta Macêdo Marques. **Mineração de dados em data warehouse para sistema de abastecimento de água**. 2009. 147 f. Dissertação (Mestrado) - Curso de Sistemas de Computação, Universidade Federal da Paraíba, João Pessoa, 2009. Disponível em: <<http://tede.biblioteca.ufpb.br/bitstream/tede/6054/1/parte1.pdf>>. Acesso em: 28 nov. 2015

GUIDO, Rafael V. C.. **Modelagem e Engenharia de Proteínas**. São Carlos, 2010. Color. Disponível em: <http://www.gradadm.ifsc.usp.br/dados/20102/FFI0776-1/Aula_05_Campo_de_Forca_dinamica_molecular.pdf>. Acesso em: 27 out. 2015.

HARREL, C. R.; GHOSH, B. K.; BOWDEN, R. **Simulation Using Promodel**. 4^aed. McGraw-Hill. 2000.

HOKAMA, Daniele Del Bianco et al. **A modelagem de dados no ambiente Data Warehouse**. São Paulo: 2004. 121 p. Projeto de Diplomação (Bacharelado em Sistemas de Informação) – Faculdade de Computação e Informática, Universidade Presbiteriana Mackenzie.

LEAL, Fabiano. **Um diagnóstico do processo de atendimento a clientes em uma agência bancária através de mapeamento do processo e simulação computacional**. 2003. 223 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Universidade Federal de Itajubá, Itajubá, 2013. Disponível em: <<http://saturno.unifei.edu.br/bim/0031150.pdf>>. Acesso em: 19 set. 2015.

MARK, James E.. **Physical Properties of Polymers Handbook**. 2. ed. Ohio: Springer, 2007. 1076 p.

MELO JUNIOR, Príamo Albuquerque; HABERT, Alberto Cláudio (Ed.). **Fronteiras da Engenharia Química I**. Rio de Janeiro: E-papers Serviços Editoriais Ltda., 2005. 284 p.

MICROSTRATEGY. **MicroStrategy Analytics Desktop**: Quick Start Guide. 2014. Disponível em: <http://www.microstrategy.com/Strategy/media/downloads/free/AnalyticsDesktop_QuickStart.pdf>. Acesso em: 25 out. 2015.

MIGUEL, Laís Mourão. **Tendências do uso de produtos naturais nas indústrias de cosméticos da França**. Revista Geográfica de América Central, Costa Rica, v. 1, n. 1, p.1-15, abr. 2011. Disponível em: <<http://www.biologia.seed.pr.gov.br/arquivos/File/biotecnologia/cosmeticos.pdf>>. Acesso em: 16 abr. 2016.

MORGON, N. H.; COUTINHO, K. **Métodos de Química Teórica e Modelagem Molecular**. São Paulo: Livraria da Física, 2007.

NAMBA, A. M.; SILVA, V. B. da; SILVA, C. H. T. P. da. **Dinâmica molecular: teoria e aplicações em planejamento de fármacos**. Eclética Química, Ribeirão Preto, v. 33, n. 4, p.13-23, jan. 2008. Disponível em: <<http://www.scielo.br/pdf/eq/v33n4/v33n4a02.pdf>>. Acesso em: 17 set. 2015.

NIST. **National Institute of Standards and Technology**. 2016. Disponível em: <<http://www.nist.gov/>>. Acesso em: 16 abr. 2016.

OLIVEIRA, Guedmiller Souza de. **Análise da interação molecular proteína-herbicida através de simulação computacional: aplicação no desenvolvimento de nanobiossensores**. 2013. 110 f. Tese (Doutorado) - Universidade Federal de São Carlos, 2013.

OLIVEIRA, J. C. P. **Desenvolvimento de um sistema de apoio a decisão para o programa bolsa universitária**. 2011. Trabalho de Conclusão de Curso

(Bacharelado em Sistemas de Informação). Centro Universitário Luterano de Palmas, Palmas, Tocantins, 2011. Disponível em: <<http://ulbra-to.br/bibliotecadigital/publico/home/documento/30>>. Acesso em: 13 out. 2015

OLIVEIRA, Luciana Fontes de. **Uso de imagens digitais e ferramentas quimiométricas para identificação e determinação de Cr (VI) em solos contaminados**. Dissertação (Mestrado). Universidade Federal de São Carlos, 2012.

OLIVEIRA, Wilson José de; **Data Warehouse**. Florianópolis: Editora Visual Books, 2002. 188 p.

PENTAHO: **Data Integration – Kettle - Spoon**. Disponível em: <<http://community.pentaho.com/projects/data-integration/>>. Acesso em: 17 out. 2015.

PICHILIANI, Mauro Carlos. **Conversando sobre banco de dados**. Clube dos Autores, 2008. 687 p.

RAINER JUNIOR, R. Kelly; CEGIELSKI, Casey G.. **Introdução a sistemas de informação**: Apoiando e transformando negócios na era da mobilidade. 3. ed. Elsevier Editora Ltda, 2012.

RAMAKRISHNAN, Raghu; GEHRKE, Johannes. **Sistemas de gerenciamento de banco de dados - 3.ed.** 3. ed. Porto Alegre: Amgh Editora, 2008. 905 p.

REGINATO, Luciane; NASCIMENTO, Auster Moreira. Um estudo de caso envolvendo Business Intelligence como instrumento de apoio a controladoria. **Revista Contabilidade e Finanças**, São Paulo, v. 18, jul. 2007. Disponível em: <<http://dx.doi.org/10.1590/S1519-70772007000300007>>. Acesso em: 01 nov. 2015.

SHANNON, Robert E. **Systems simulation: the art and science**. Englewood Cliffs: PrenticeHall, [s. n.], 1975.

SHERER, Cláudio. **Métodos Computacionais da Física**. Editora Livraria da Física, 2010. 300 p. Disponível em: <<http://lief.if.ufrgs.br/pub/Cursos/Cuda/MetodosC-Scilab.pdf>>. Acesso em: 17 out. 2015.

SINGH, Harry S.. **Data Warehouse: Conceitos, Tecnologias, Implementação e Gerenciamento**. São Paulo: Makron Books, 2001.

SPOEL, D. van der; LINDAHL, E.; HESS, B.; BUUREN, A. R. van; APOL, E.; MEULENHOF, P. J.; TIELEMAN, D. P.; SIJBERS, A. L. T. M.; FEENSTRA, K. A.; DRUNEN, R. van; BERENDSEN, H. J. C. **Gromacs: Fast, Flexible and Free**. 2005.

TURBAN, Efraim et al. **Business Intelligence: Um enfoque gerencial para a inteligência do negócio**. São Paulo: Editora Bookman, 2009. 249 p.

UCSF Chimera: an Extensible Molecular Modelling System. Disponível em: <<https://www.cgl.ucsf.edu/chimera/>>. Acesso em: 17 out. 2015.

ANEXOS

ANEXO I - DISTÂNCIAS EXTRAÍDAS DO NIST E VALORES MÁXIMOS E MÍNIMOS

Tabela 2: valores de referência NIST

Ligações da cianidina-3-glicosídeo	Distância (Angström)	Endereço - NIST
C-H	1.078	http://goo.gl/uAi6gf
C-C	1.531	http://goo.gl/FeRzXj
C=C	1.398	http://goo.gl/3MB8Xg
C-O	1.420	http://goo.gl/FeRzXj
O-H	0.956	http://goo.gl/jZnGKX

Tabela 3: Valores máximos e mínimos para as distâncias de ligações

Ligação	Distância (Angström)	Valor Máximo (Angström)	Valor Mínimo (Angström)
C-H	1.078	1.140	1.054
C-C	1.531	1.557	1.187
C=C	1.398	1.557	1.187
C-O	1.420	1.448	1.191
O-H	0.956	0.980	0.940

* (-) Ligação simples

* (=) Ligação dupla

*C – Carbono

*H – Hidrogênio

*O - Oxigênio

ANEXO II – SCRIPT – PRODUÇÃO DO CAMPO DE FORÇA

```

# Version 1.1_dev - 10/14/2009
#
# Updates since version 1.0
# 1. Made command line options into a hash, requiring flags
# 2. Added printing of atom labels to bonded parameters sections
# 3. Added renumbering option
# 4. Added compatibility with chain identifiers and notes about atom type field
# 5. Added primitive atom typing and charge assignment based on neighboring atoms and connectivity
# 6. Added charge group assignment for each new C or N atom type found
# 7. Added dihedral check - proper dihedrals no longer erroneously generated for amides
# 8. Added ability to output either a .itp or a .top version of the topology
#

# Date and version info for printing to topology
chomp(my $date = `date`);
my $v = "1.1_dev (10/14/2009)";

# Store command line arguments in a hash
my %arguments = @ARGV;

# Parse through command line arguments
if (exists($arguments{"-h"})) {
    print "
#
# TopolGen, version $v
# Written by Justin Lemkul
#
# The program takes an all-atom .pdb file and generates a GROMACS-compatible topology.
# The only force field option is OPLS-AA at this time. Usage:
#
# perl topolgen.pl -f input.pdb -o output.top [-type itp/top] [-renumber]
#
# Questions, comments, bug reports? Email jalemkul[at]vt[dot]edu
#\n\n";
    exit;
}

# Define input file
my $input;

if (defined($arguments{"-f"})) {
    $input = $arguments{"-f"};
} else {
    die "Input not specified. Usage: perl $0 -f <input.pdb> -o <output.top> [-type itp/top] [-renumber]\n";
}

# Define output type
my $out_type;

if (defined($arguments{"-type"})) {
    $out_type = $arguments{"-type"};
} else {
    print "\nNo output type specified. Using default type of .top\n";
    $out_type = "top";
}

# Define output file
my $output;

if (defined($arguments{"-o"})) {
    $output = $arguments{"-o"};
} else {
    if ($out_type eq "top") {
        print "\nNo output filename specified. Using default output filename \"ffoplsaa_TopolGen.top.\n";
        $output = "ffoplsaa_TopolGen.top";
    } elsif ($out_type eq "itp") {
        print "\nNo output filename specified. Using default output filename \"ffoplsaa_TopolGen.itp.\n";
        $output = "ffoplsaa_TopolGen.itp";
    } else {
        printf("\nUnknown output type %s\n", $arguments{"-type"});
    }
}

# Check to make sure output filename and type are consistent
my @check = split('\.', $output);

```

```

if ($check[1] ne $out_type) {
    printf("\nOutput filename and type are inconsistent: file name is of type %s, specified type is %s\n\n", $check[1], $out_type);
    exit;
}

# Define some pre-determined function types required by OPLS
my $funct_bond = 1; # Harmonic bond potential
my $funct_angle = 1; # Harmonic angle
my $funct_dihedral = 3; # Ryckaert-Bellemans dihedral
my $funct_pairs = 1; # Lennard-Jones pair potential
my $funct_improper = 1; # "Proper" dihedral representing the improper

# Generate the [ atoms ] section and header information
open(ATOMS_IN, $input) or die "Problem while reading $input for header and atoms: $!\n";
my @input_file = <ATOMS_IN>;
close(ATOMS_IN);

# strip out only the HETATM records, write to new array
my @atoms;

foreach $ (@input_file) {
    if ($_ =~ /^HETATM/) {
        push(@atoms, $_);
    }
}

#
# Fields in the PDB file:
# HETATM atom# atom_name residue res# x y z beta occ type
#
# Can throw away HETATM, x/y/z coords, beta, and occupancy
#
# Chain identifier may be present!
# atom# atom_name residue chain res# type
#

my @line;
my @lines_edit;

foreach $_ (@atoms) {
    chomp(@line = split(" ", $_));
    shift(@line); # remove HETATM

    # Check to see if chain identifier is present
    if (scalar(@line) > 10) { # line contains chain identifier
        my @coords_b_occ = splice(@line, 5, 5); # remove x/y/z, beta, occupancy
                                                # store in case we need it
    } else { # no chain identifier
        my @coords_b_occ = splice(@line, 4, 5);
    }

    my $string = join(" ", @line);
    push(@lines_edit, $string);
}

#
# So now, in each element of @line, there is a string that contains
# atom# atom_name residue res# atom_type
#

chomp(my @temp_info = split(" ", $lines_edit[0]));
my $res_name = $temp_info[2];
my $first_atom_num = $temp_info[0];
my $renumber_factor;

# Determine if renumbering is necessary, based on command line flag
if (exists($arguments{"-renumber"})) {
    if ($first_atom_num != 1) {
        $renumber_factor = $first_atom_num - 1;
    }
}

# Open output file
open(HEADER_OUT, ">>section_header");
print HEADER_OUT "\n";
print HEADER_OUT "\tOPLS-AA topology, built by TopolGen version $v\n";
print HEADER_OUT "\tScript written by: Justin Lemkul (jalemkul@vt.edu)\n";
print HEADER_OUT "\tThis is your molecule's topology\n";
print HEADER_OUT "\tCheck it carefully for any errors. It is not necessarily perfect!\n";
print HEADER_OUT "\n";
print HEADER_OUT "\tTopology written on $date\n";
print HEADER_OUT "\n";

```

```
# proper [ dihedrals ] section
open(FIFTH_IN, "<section_dihedrals") or die "Cannot open \"section_dihedrals\" while writing output: $!\n";
while (my $line_out = <FIFTH_IN>) {
    print TOPOLOGY_OUT $line_out;
}

close(FIFTH_IN);

# improper [ dihedrals ] section
# only write it out if it contains relevant lines
open(TEST, "<section_impropers") or die "Cannot open \"section_impropers\" while writing output: $!\n";

chomp(my @test_array = <TEST>);

my $test_length = scalar(@test_array);

if ($test_length > 3) {
    open(SIXTH_IN, "<section_impropers") or die "Cannot open \"section_impropers\" while testing for output: $!\n";
    while (my $line_out = <SIXTH_IN>) {
        print TOPOLOGY_OUT $line_out;
    }

    close(SIXTH_IN);
}

if ($out_type eq "top") {
    print TOPOLOGY_OUT "[ system ]\n";
    print TOPOLOGY_OUT "; Name\n";
    print TOPOLOGY_OUT "$res_name topology, generated by TopolGen\n";
    print TOPOLOGY_OUT "\n";
    print TOPOLOGY_OUT "[ molecules ]\n";
    printf TOPOLOGY_OUT ":%9s%13s\n", "Compound", "\#mols";
    printf TOPOLOGY_OUT ":%-6s%15d\n", $res_name, 1;
}

close(TOPOLOGY_OUT);
```

APÊNDICES

APÊNDICE A - PROCEDIMENTOS REALIZADOS NO GROMACS

Foram realizados dois procedimentos de simulação da dinâmica molecular, com tempos diferentes, para identificar o comportamento e a estabilidade da cianidina-3-glicosídeo.

Os procedimentos foram realizados de acordo com as seguintes etapas:

- Etapa básica: preparação das propriedades da molécula, comum aos dois procedimentos;
- Etapa específica: definição das variáveis específicas para cada procedimento.

Etapa Básica

Para iniciar o processo de simulação é necessário determinar o campo de força, que irá executar sobre a molécula durante o processo de simulação. Este processo criará a topologia da molécula, que consiste nas coordenadas dos átomos que compõem a cianidina-3-glicosídeo.

A definição do campo de força foi realizada através do *script*, este disponível no anexo II. Para executar o script digita-se no *prompt*, o comando 2, que realiza a leitura do arquivo de entrada **.pdb** e gera a topologia da molécula, chamado de **topol.top**, que é um nome padrão.

```
perl topolgen1.1.pl -f cianidina-3-glicosideo.pdb -o topol.top
```

Comando 2: criar a topologia

Após gerar a topologia da molécula é necessário construir um arquivo com a extensão **.gro**, que corresponde a estrutura molécula, utilizado para gerar a trajetória da molécula. Para criar o arquivo foi utilizado o comando 3, que utiliza um arquivo no formato **.pdb** como entrada.

```
editconf -f cianidina-3-glicosideo.pdb -o cianidina-3-  
glicosideo.gro
```

Comando 3: criar o arquivo .gro

O resultado do comando 2 e 3 são os arquivos que serão utilizados nas próximas etapas da simulação da dinâmica molecular.

Procedimento 1 – Etapa Específica

Inicialmente, foi realizada a etapa de preparação do ambiente, no qual foi definido o campo de força OPLS (*Optimized Potential for Liquid Simulations*).

Depois foi definido o espaço de simulação, executando o programa **editconf**, que permite criar uma caixa, com informações sobre ângulos, coordenadas, velocidades etc., definidas nos parâmetros passados no momento da execução do programa, com o comando 4. A caixa de simulação foi definida como cúbica, parâmetro **cubic**, que possui um raio de 1.0nm (nanômetro). O parâmetro **cianidina-3-glicosideo.gro** se refere ao arquivo gerado na execução do comando 3 e o parâmetro **cianidina-3-glicosideo_newbox.gro** é a caixa de simulação o arquivo a ser gerado.

```
gmx editconf -f cianidina-3-glicosideo.gro -o cianidina-3-  
glicosideo_newbox.gro -c -d 1.0 -bt cubic
```

Comando 4: caixa de simulação

Em seguida, a caixa foi preenchida com solvente, usando o programa **genbox**, que acrescenta o solvente a topologia da molécula. O preenchimento da caixa foi realizado executando-se o comando 5. O parâmetro **cianidina-3-glicosideo_newbox.gro** se refere ao arquivo gerado na execução do comando 4, o parâmetro **spc216.gro** se refere o tipo de solução que foi adicionado a caixa e o parâmetro **cianidia-3-glicosideo_solv.gro** indica o resultado das coordenadas do solvente a topologia da molécula, através do parâmetro **topol.top**.

```
gmx genbox -cp cianidina-3-glicosideo_newbox.gro -cs  
spc216.gro -o cianidina-3-glicosideo_solv.gro -p topol.top
```

Comando 5: adição de solvente

Nesse momento, a caixa foi preenchida com solvente, sendo que o próximo passo foi realizar a minimização de energia do sistema. Primeiro foi criado o arquivo

minim.mdp, que possui os parâmetros de entrada para a simulação, conforme mostra a Figura 16.

```

; minim.mdp - used as input into grompp to generate em.tpr
; realizou-se 50 passos

; Run control
integrator = steep           ; algoritmo utilizado para a minimizacao de energia
nsteps = 500000             ; numero de passos para a integracao ou minimizacao

constraints = none
cutoff-scheme = verlet

; Pre-processamento
define = -DFLEXIBLE         ; apontando que sera uma agua flexivel em sua topologia

; Energia de minimizacao
emtol = 100                 ; a minimizacao eh convergente em que a forza for menor que este valor
emstep = 0.01               ; passo inicial (nm)

; pesquisa de vizinhos
nstlist = 1                 ; frequencia de atualizacao dos vizinhos
ns_type = grid              ; metodo de atualizacao dos vizinhos (grid ou simple)
pbc = xyz                   ; condicoes de contornos periodicas (xyz -> em todas as direcoes)
rlist = 1.0                 ; distancia entre os vizinhos

; eletrostatico
coulombtype = PME           ; Particle-Mesh Ewald electrostatics
rcoulombtype = 1.0

; Van der Waal
rvdw = 1.0                  ; distancia para (LJ cut-off)

```

Figura 16: parâmetro minimização de energia

Após a criação do arquivo **minim.mdp**, realizou-se o pré-processamento para a minimização de energia, através do programa **grompp**, executando-se o comando 6. O parâmetro **minim.mdp** se refere ao arquivo criado anteriormente, visto na Figura 16, o parâmetro **cianidia-3-glicosideo_solv.gro** se refere ao arquivo gerado na execução do comando 5, o parâmetro **topol.top** se refere ao arquivo gerado na execução do comando 2 e o parâmetro **em.tpr** se refere ao arquivo gerado pelo pré-processamento para a minimização de energia.

```

gmx grompp -f minim.mdp -c cianidia-3-glicosideo_solv.gro -p
           topol.top -o em.tpr

```

Comando 6: pré-processamento minimização de energia

O resultado do comando 6 foi a criação do arquivo **em.tpr**, que contém as coordenadas e a topologia geral do sistema. Este arquivo foi utilizado pelo programa **mdrun**, para realizar a minimização de energia, comando 7. O parâmetro **-v** indica o progresso da minimização de energia, o parâmetro **-deffnm** define os nomes de entrada e saída do arquivo **em** a ser gerado.

```
gmx mdrun -v -deffnm em
```

Comando 7: executar a minimização de energia

O comando 7 resultou na criação de três arquivos: **em.log** – log de ações; **em.edr** – contém todos os termos de energia; e **em.tpr** – contém coordenadas geral do sistema.

Após executar a minimização de energia, realizou-se a estabilização da molécula, procedimento realizado em duas etapas. Na primeira etapa realizou-se a estabilização da temperatura, com a qual criou-se o arquivo **nvt.mdp**, que possui os parâmetros para a estabilização da temperatura, conforme mostra a Figura 17.

```
title = OPLS EtAD NVT equilibration - antocianina
define = -DFLEXIBLE
integrator = md
nsteps = 50000
dt = 0.002

nstxout = 100
nstvout = 100
nstenergy = 100
nstlog = 100

continuation = no
constraint_algorithm = lincs
constraints = none
lincs_iter = 1
lincs_order = 4
nstlist = 5
ns_type = grid
rlist = 1.0
rcoulomb = 1.0
rvdw = 1.0
coulombtype = PME
pme_order = 4
fourierspacing = 0.16
tcoupl = V-rescale
tc-grps = System
tau-t = 0.1
ref-t = 300
pcoupl = no
pbc = xyz
DispCorr = EnerPres
gen_vel = yes
gen_temp = 300
gen_seed = -1
cutoff-scheme = Verlet
```

Figura 17: parâmetros nvt.mdp

Após a criação do arquivo **nvt.mdp**, realizou-se o pré-processamento, através do programa **grompp**, executando-se o comando 8. O parâmetro **nvt.mdp** se refere ao arquivo criado, conforme mostra a Figura 17. O parâmetro **em.gro** se refere ao arquivo criado na execução do comando 7, o parâmetro **topol.top** se refere ao arquivo gerado na execução do comando 2 e o parâmetro **nvt.tpr** é o arquivo a ser gerado.

```
gmx grompp -f nvt.mdp -c em.gro -p topol.top -o nvt.tpr
```

Comando 8: pré-processamento equilibração NVT

Em seguida, foi executado o programa **mdrun** sobre o arquivo **nvt.tpr**, para equilibrar a temperatura, executando-se o comando 8. O parâmetro **-deffnm** define os nomes de entrada e saída do arquivo **nvt** a ser gerado.

```
gmx mdrun -deffnm nvt
```

Comando 9: executar a equilibração NVT

Na segunda etapa de equilibração da molécula, foi realizada a estabilização da pressão. Primeiro criou-se o arquivo **npt.mdp**, que possui os parâmetros mostrado na Figura 18.

```
title = OPLS EtAD NPT equilibration
define = -DFLEXIBLE
integrator = md
nsteps = 50000
dt = 0.002
nstxout = 100
nstvout = 100
nstenergy = 100
nstlog = 100
continuation = yes
constraint_algorithm=lincs
constraints = all.bonds
lincs_iter = 1
lincs_order = 4
ns_type = grid
nstlist = 5
rlist = 1.0
rcoulomb = 1.0
rvdw = 1.0
coulombtype = PME
pme_order = 4
fourierspacing = 0.16
tcoupl = V-rescale
tc-grps = Protein Non-Protein ;atencao a este item
tau_t = 0.10.1
ref_t = 300 300
pcoupl = Parrinello-Rahman
pcoupltype = isotropic
tau_p = 2.0
ref_p = 1.0
compressibility = 4.5e-5
refcoord_scaling = com
pbc = xyz
DispCorr = EnerPres
gen_vel = no
```

Figura 18: parâmetro npt.mdp

Após a criação do arquivo **npt.mdp**, realizou-se o pré-processamento através do programa **grompp**, executando-se o comando 9. O parâmetro **npt.mdp** se refere ao arquivo criado, visto na Figura 18, o parâmetro **nvt.gro** se refere ao arquivo criado na execução do comando 9, o parâmetro **topol.top** se refere ao arquivo gerado na execução do comando 2 e o parâmetro **npt.tpr** é o arquivo a ser gerado.

```
gmx grompp -f npt.mdp -c nvt.gro -p topol.top -o npt.tpr
```

Comando 10: pré-processamento NPT

Após realizar o pré-processamento, executou-se o programa **mdrun**, sobre o arquivo **npt.tpr**, para alcançar o equilíbrio da pressão do sistema, através do comando 10. O parâmetro **-deffnm** define os nomes de entrada e saída do arquivo **npt** a ser gerado.

```
gmx mdrun -deffnm npt
```

Comando 11: executar a equilibração NPT

Após concluir a fase de estabilização do sistema, foi realizada a produção da dinâmica molecular, primeiro criou-se o arquivo **md.mdp**, que possui os parâmetros mostrados na Figura 19.

```

title = OPLS EtAD MD
integrator = md
nsteps = 500000
dt = 0.002
nstxout = 1000
nstvout = 1000
nstenergy = 1000
nstlog = 1000
continuation = yes
constraint_algorithm=lincs
constraints = all.bonds
lincs_iter = 1
lincs_order = 4
ns_type = grid
nstlist = 5
rlist = 1.0
rcoulomb = 1.0
rvdw = 1.0
coulombtype = PME
pme_order = 4
fourierspacing = 0.16
tcoupl = V-rescale
tc-grps = Protein Non-Protein ;atencao a este item
tau_t = 0.1 0.1
ref_t = 300 300
pcoupl = Parrinello-Rahman
pcoupltype = isotropic
tau_p = 2.0
ref_p = 1.0
compressibility = 4.5e-5
pbc = xyz
DispCorr = EnerPres
gen_vel = no

```

Figura 19: parâmetro md.mdp

Após a criação do arquivo **md.mdp**, executou-se o pré-processamento da produção da dinâmica molecular, usando o programa **grompp**, através do comando 12. O parâmetro **md.mdp** se refere ao arquivo visto na Figura 19, o parâmetro **npt.gro** e **npt.cpt** se referem ao arquivo gerado na execução do comando 11, o parâmetro **topol.top** se refere ao arquivo gerado no comando 2 e o parâmetro **md_0_1.tpr** é o arquivo a ser gerado.

```

gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o
md_0_1.tpr

```

Comando 12: pré-processamento dinâmica molecular

Com a execução do comando 12, foi gerado o arquivo **md_0_1.tpr**, que foi utilizado para realizar a dinâmica molecular, executando-se o comando 13.

```

gmx mdrun -deffnm md_0_1

```

Comando 13: executar a dinâmica molecular

Para a análise da trajetória da molécula, foi criado o arquivo **filme.pdb**, tal procedimento foi realizado pelo programa **trjconv**, que contém um arquivo com a extensão **.xtc** gerado na execução do comando 13 para o formato **.pdb**. Esta conversão foi realizada executando-se o comando 14. O parâmetro **md_0_1.tpr** se refere ao arquivo gerado na execução do comando 13, o parâmetro **md_0_1.xtc** se refere ao arquivo gerado na execução do comando 13 e o parâmetro **filme.pdb** é o arquivo a ser gerado pelo comando 14.

```
trjconv -s md_0_1.tpr -f md_0_1.xtc -o filme.pdb -ur compact -  
center -pbc mol
```

Comando 14: criar trajetória da dinâmica molecular

Após a execução do comando 14, foi possível visualizar a movimentação da molécula.

Procedimento 2 – Etapa Específica

Inicialmente, foi realizada a etapa básica, executando os comandos 2 e 3 para a definição do campo de força, no qual foi definido o mesmo campo de força do primeiro procedimento OPLS.

O ambiente de simulação foi definido, executando o programa **editconf**, definindo os parâmetros passados no momento da execução do programa, com o comando 15. A caixa possui as mesmas dimensões do primeiro procedimento.

```
gmx editconf -f cianidina-3-glicosideo.gro -o cianidina-3-  
glicosideo_newbox.gro -c -d 1.0 -bt cubic
```

Comando 15: definição do ambiente

Em seguida, a caixa foi preenchida com o mesmo solvente, usando o programa **genbox**, que acrescenta o solvente a topologia da molécula. O preenchimento da caixa foi realizado executando-se o comando 16.

```
gmx genbox -cp cianidina-3-glicosideo_newbox.gro -cs  
spc216.gro -o cianidina-3-glicosideo_solv.gro -p topol.top
```

Comando 16: preencher a caixa com solvente

Após o preenchimento da caixa, foi realizada a minimização de energia do sistema. Neste ponto os parâmetros foram alterados para que as informações pudessem ser analisadas a partir desta etapa. Primeiro foi criado o arquivo **minim2.mdp**, que possui os parâmetros de entrada para a simulação, conforme mostra a Figura 20.

```

;minim.mdp - used as input into grompp to generate em.tpr

;Run control
integrator = steep           ;algoritmo utilizado para a minimizacao de
nsteps = 50000              ;numero de passos para a integracao ou mini

constraints = none
cutoff-scheme = Verlet      ;pesquisa de vizinhos (verlet ou group) - re
atomos vizinhos

;Pre-processamento
define = -POSRES           ;apontando que sera uma agua flexivel em su

;Energia de minimizacao
emtol = 4895
emstep = 273.9
niter = 20

;pesquisa de vizinhos
nstlist = 10               ;frequencia de atualizacao dos vizinhos
ns_type = grid             ;metodo de atualizacao dos vizinhos (grid ou simple
pbc = xyz                  ;condicoes de contornos periodicas (xyz ->
rlist = 1.0                ;distancia entre os vizinhos
periodic-molecules = yes  ; influencia no PBC (yes or no)

;eletroestatico
coulombtype = PME          ;Particle-Mesh Ewald electrostatics
rcoulomb = 1.2

;Van der Waals
rvdw = 1.2                 ;distancia para (LJ cut-off)
DispCorr = Ener
vdwtype = cut-off          ;PME, cut-off,

```

Figura 20: parâmetro minim2.mdp

Após a criação do arquivo **minim2.mdp**, realizou-se o pré-processamento para a minimização de energia, através do programa **grompp**, executando-se o comando 17.

```

gmx grompp -f minim.mdp -c cianidia-3-glicosideo_solv.gro -p
topol.top -o em2.tpr

```

Comando 17: pré-processamento da segunda minimização de energia

O resultado do comando 17 foi a criação do arquivo **em2.tpr**, que contém as coordenadas e topologia geral do sistema. Este arquivo foi utilizado pelo programa **mdrun**, para realizar a minimização de energia, comando 18.

```
gmx mdrun -v -deffnm em2
```

Comando 18: executar a segunda minimização de energia

O comando 18 resultou na criação de três arquivos: **em2.log** – log de ações; **em2.edr** – contém todos os termos de energia; e **em2.tpr** – contém coordenadas geral do sistema.

Para a análise dos resultados da minimização de energia, foi gerado o arquivo **minimizacao.pdb**, que contém a trajetória da minimização, através do comando 19.

```
gmx trjconv -s em2.tpr -f em2.trr -drop em2.xvg -o  
minimizacao.pdb -pbc mol -ur compact
```

Comando 19: trajetória da minimização de energia

O arquivo gerado, foi utilizado para a extração das distâncias, a partir da minimização de energia.

APÊNDICE B – DISTÂNCIAS DA CIANIDINA-3-GLICOSÍDEO – PRIMEIRO PROCEDIMENTO

Tabela 4: frame cianidina-3-glicosideo (distâncias)

Atomos	Distância
C H	1.285
C C	1.600
H C	1.399
C C	1.593
O C	1.650
O H	1.282
C C	1.682
O C	1.961
H O	0.967
C C	1.780
C H	1.210
C C	1.786
C C	1.779
O C	1.985
C O	1.568
C C	1.704
C C	1.995
C H	1.245
C C	1.784
C C	1.620
C C	1.698
C C	1.752
C C	1.685
C H	1.420
H O	1.097
C H	1.235
H O	1.258
O C	1.685
O C	1.965
C H	1.352
C C	1.782
C O	1.652
C H	1.265
O H	1.180
C C	1.721
H C	1.215
O C	1.850
H O	1.450
C H	1.652
O C	1.782

O H	1.478
C C	1.850
C C	1.792
H C	1.352
C H	1.259
O C	1.580
O H	1.560
O C	1.780
O C	1.698
C C	1.200
C C	1.102
O C	1.152
O C	1.215
C C	1.892

APÊNDICE C - UTILIZAÇÃO DA FERRAMENTA OLAP

Para a análise das distâncias utilizou-se a ferramenta OLAP *Microstrategy Analytics Desktop*. Primeiramente foi criada uma *dashboard*, com a qual permite selecionar o conjunto de dados a serem trabalhados, conforme apresenta a Figura 21.

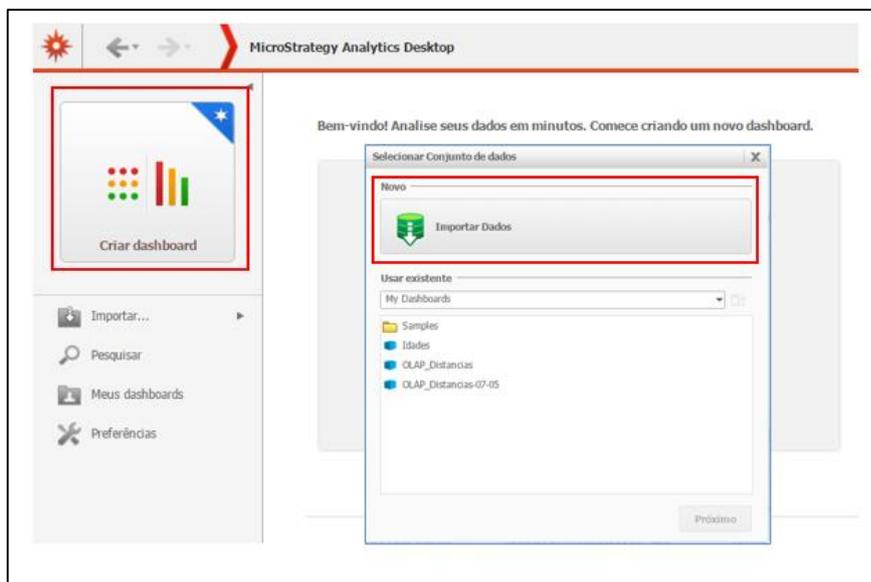


Figura 21: criar dashboard

Ao clicar na opção “Importar Dados”, visto na Figura 21, foi apresentada a página inicial do assistente para a definição do tipo de dado a ser importado pela ferramenta, conforme apresenta a Figura 22.

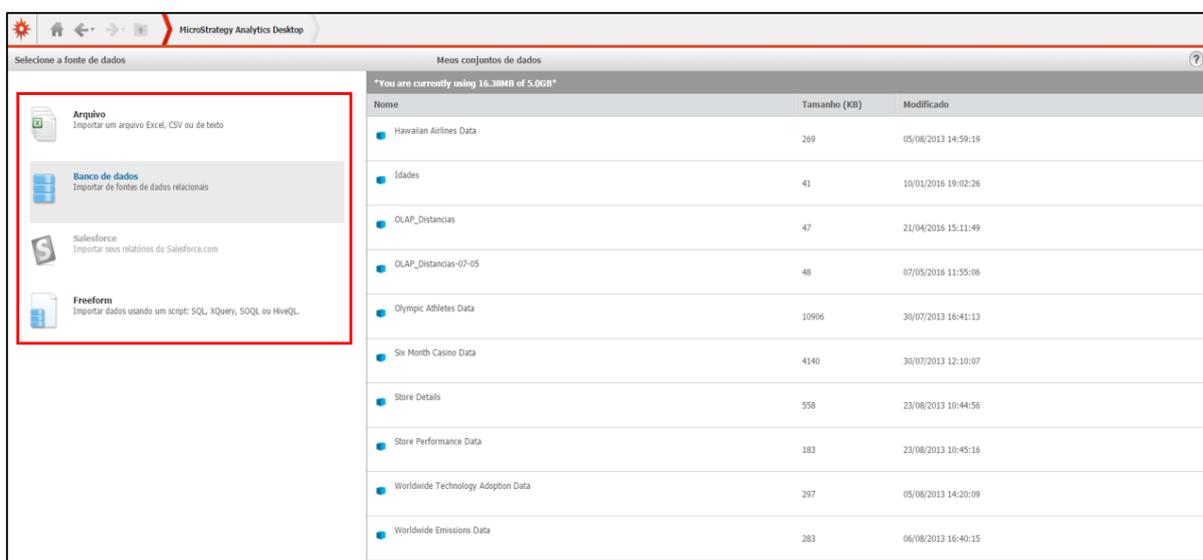


Figura 22: importar os dados

O trecho destacado da Figura 22, apresenta os tipos de dados aceitáveis pela ferramenta, que podem ser, dados de um arquivo do Excel, CSV ou até mesmo de fontes de dados relacionais. Para o projeto desenvolvido, os dados foram importados de uma base de dados relacional, selecionando a opção “Banco de Dados”.

Ao selecionar a opção “Banco de Dados” é necessário criar a conexão com a fonte de dados de origem, clicando no botão “Nova Conexão”. Após este passo a ferramenta carrega uma janela para realizar a conexão com o banco de dados, sendo necessário indicar as informações de conexão necessária, como por exemplo, nome do banco de dados, o tipo de SGBD e o nome da conexão, conforme mostra a Figura 23.



Figura 23: conexão com o banco de dados

Realizada a conexão com a base de dados, é preciso definir quais tabelas e colunas serão carregadas para a criação do CUBO OLAP. Para o projeto foram definidas as seguintes tabelas: Atomo_Antocianina, Atomo_Nist e as colunas: átomo, distancia e distancia_nist, conforme mostra a Figura 24.

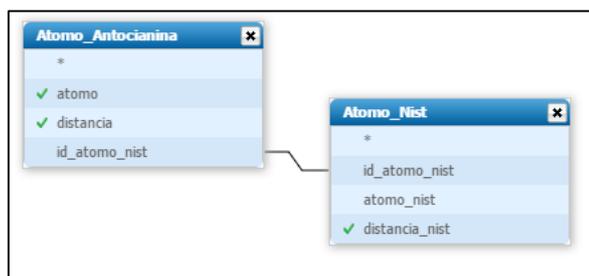


Figura 24: tabelas a serem carregadas

Definido os campos a serem carregados, determina-se o tipo de visualização dos dados, conforme mostra a Figura 25.

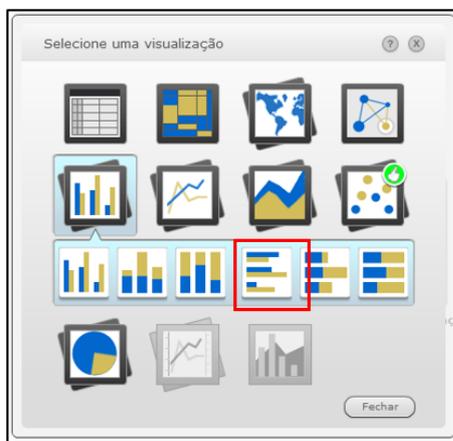


Figura 25: tipo de visualização

Para a análise das distâncias, foi escolhido a visualização em gráficos do tipo barras, como mostra o trecho em destaque na Figura 25. A ferramenta seleciona automaticamente os dados para adicionar a visualização, para otimizar sua exibição.

Para exibir os dados, pode-se arrastar e soltar os atributos ao painel de visualização, conforme mostra a Figura 26.

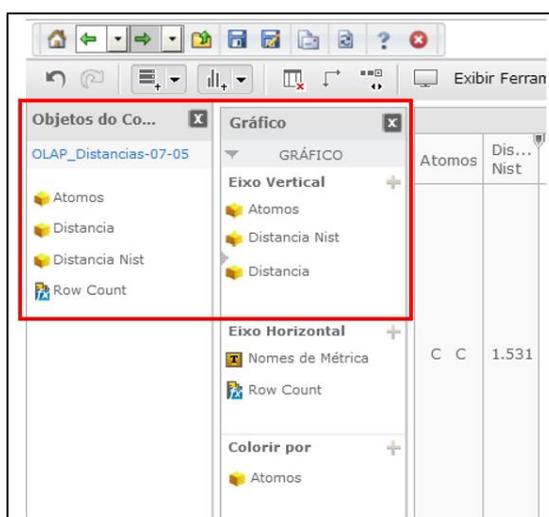


Figura 26: organização dos dados

A Figura 27, apresenta o resultado sobre os atributos adicionados ao painel de visualização, de acordo a sua ordem definida no Eixo Vertical.

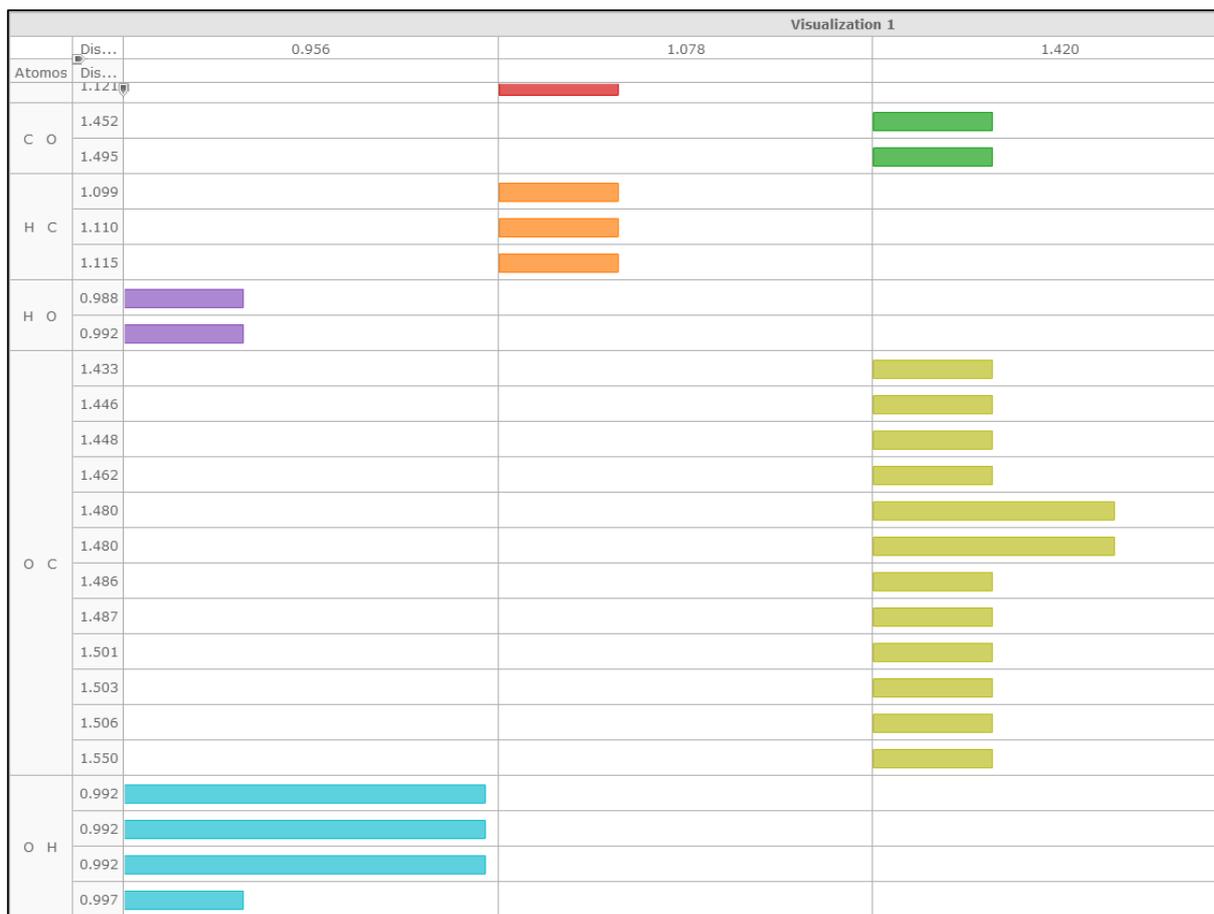


Figura 27: resultado da comparação dos dados

Após salvar este painel de visualização, os dados atualizados no banco de dados serão automaticamente atualizados pela ferramenta, sendo necessário criar um novo cubo.