



**UNIVERSIDADE FEDERAL DO CEARÁ  
CENTRO DE CIÊNCIAS  
DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR  
BACHARELADO EM BIOTECNOLOGIA**

**MATHIAS COELHO BATISTA**

**DESENVOLVIMENTO DE UMA PLATAFORMA WEB PARA ANÁLISE DE  
LIGAÇÕES DE HIDROGÊNIO EM SISTEMAS PROTEÍNA-ÁGUA**

**FORTALEZA  
2019**

MATHIAS COELHO BATISTA

DESENVOLVIMENTO DE UMA PLATAFORMA WEB PARA ANÁLISE DE  
LIGAÇÕES DE HIDROGÊNIO EM SISTEMAS PROTEÍNA-ÁGUA

Monografia apresentada ao curso de Bacharelado em Biotecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Biotecnologia.

Orientador Pedagógico: Prof. Dr. José Hélio Costa.

Orientador Técnico-Científico: Prof. Dr. Marcos Roberto Lourenzoni.

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação

Universidade Federal do Ceará

Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

B337d Batista, Mathias Coelho.

Desenvolvimento de uma plataforma web para análise de Ligações de Hidrogênio em sistemas proteína-água. / Mathias Coelho Batista. – 2019.

45 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências, Curso de Biotecnologia, Fortaleza, 2019.

Orientação: Prof. Dr. José Hélio Costa.

Coorientação: Prof. Dr. Marcos Roberto Lourenzoni.

1. Ligação de Hidrogênio. 2. Dinâmica Molecular. 3. g\_ener\_hb. I. Título.

CDD 661

---

MATHIAS COELHO BATISTA

DESENVOLVIMENTO DE UMA PLATAFORMA WEB PARA ANÁLISE DE  
LIGAÇÕES DE HIDROGÊNIO EM SISTEMAS PROTEÍNA-ÁGUA

Monografia apresentada ao curso de Bacharelado em Biotecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Biotecnologia.

Aprovada em: \_\_\_\_/\_\_\_\_/\_\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. José Hélio Costa (Orientador Pedagógico)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Marcos Roberto Lourenzoni (Orientador Técnico-Científico)  
Fundação Oswaldo Cruz - CE

---

Dr. Raphael Trevizani  
Fundação Oswaldo Cruz - CE

Aos meus pais, Glória e Leonardo.

## AGRADECIMENTOS

Meus primeiros agradecimentos cabem àqueles que constituem a minha base. Os responsáveis pela minha formação moral, e que nunca mediram esforços para garantir que eu tivesse acesso à educação, meus pais, Leonardo e Glória, pois não teria chegado aqui sem o apoio deles. Agradeço aos meus irmãos Matheus e Marcos por todas as nossas vivências. Não podendo esquecer minhas tias Elisangela e Tiá, por quem tenho muito apreço.

Agradeço ao Prof. Dr. Marcos Roberto Lourenzoni, por ter acreditado na minha capacidade, pela orientação, e por estar sempre disposto a ajudar no desenvolvimento deste trabalho. A todos os membros do Grupo de Desenvolvimento De Fármacos e Biofármacos: Evolução, In Sílico e In Vitro, de Biomoléculas, em especial Cássio, Ana Júlia, Virgínia, Alison, Maísa e Natália.

À Fundação Oswaldo Cruz por apoiar o desenvolvimento deste trabalho.

Agradeço ao Laboratório de Genômica Funcional e Bioinformática, por ter me acolhido. Ao Prof. Dr. José Hélio Costa, por ter me apresentado o mundo da Bioinformática, e por ter dado a liberdade necessária para o meu desenvolvimento. Deste laboratório guardo as melhores lembranças, em especial dos amigos Clesivan, Lyndefânia e Rachel. Faço menção especial à Thais Andrade, pelas longas e proveitosa reflexões, por ser compreensiva, paciente, atenciosa e amiga.

Agradeço ao Laboratório de Biotecnologia Vegetal, por ter me proporcionado a primeira experiência com pesquisa científica, em especial a Profª Dra. Cristina Paiva da Silveira Carvalho, pela orientação paciente.

Agradeço àqueles com quem partilhei a vida acadêmica e que tornaram o caminhar mais leve e rico, Cristiano, Ruthyelle, Jonathan, Pâmella, Mel, Clauds, Gaby, Jéssica, Deborah, Jamily, João Felipe, Ewerton, Lucas Lopes e Marcilene.

Agradeço aos amigos Arley, Jéssica e Luís pelos bons conselhos, e pela amizade que transcende o espaço e o tempo. Aos parceiros Expedito, Esley, Ismael, Samara, Bené, Cleylson, Thyago e Allan pela amizade, e pela alegria contagiante.

Aos professores João Evangelista, Elder, Jarbas e Rochelle.

Agradeço à Universidade Federal do Ceará (UFC); Aos professores e funcionário do Departamento de Bioquímica e Biologia Molecular. Em especial à Profª. Drª Márjory Lima Holanda, e ao Gilmar Ferreira, coordenadora e secretário do Bacharelado em Biotecnologia, respectivamente.

**“Maravilhar-se é o primeiro passo para  
um descobrimento”**

**Louis Pasteur**

## RESUMO

As Ligações de Hidrogênio (LHs) são interações atrativas geralmente estabelecidas entre um dipolo formado por um átomo altamente eletronegativo (X) e um Hidrogênio (X--H), como outro átomo eletronegativo (X—H···Y). Têm grande importância na formação das estruturas secundária e terciária das proteínas, além de serem responsáveis por muitas das características da água que tornam a vida possível. Devido à importância das LHs, diversos estudos foram feitos com a intenção de entender os princípios que regem a sua formação e quebra, bem como a sua influência na estruturação das biomoléculas. Um dos principais vieses para estudo das LHs é a abordagem computacional, através da utilização de ferramentas voltadas para Dinâmica Molecular, como o pacote GROMACS, que permite identificar a formação das LHs, mas não sua energia. Deste modo, a ferramenta *g\_ener\_hb* foi desenvolvida, apresentando a capacidade de identificar e mensurar as energias das LHs formadas em sistemas oriundos de simulação feita pelo GROMACS. Contudo, devido ao fato de que sua instalação e configuração podem ser desafiadoras para pessoas sem experiência com sistemas operacionais Linux, este trabalho foi executado objetivando facilitar o estudo das LHs, ao produzir uma plataforma online, baseada no processo de análise das ferramentas *g\_ener\_hb* e GROMACS, capaz de analisar a energia das LHs, de forma simples e eficaz. A plataforma foi desenvolvida utilizando-se o Python como linguagem de programação principal, junto do framework *Django*, para organização e otimização do desenvolvimento web. Utilizou-se ainda o gerenciador de banco de dados *SQLite3* para armazenamento dos parâmetros usados no processamento e dados dos usuários. Utilizou-se HTML, CSS e *JavaScript* no desenvolvimento da interface. As bibliotecas JS *JQuery* e *NGL* foram usadas para interação com o servidor, e para visualização de proteínas, respectivamente. Por fim, a plataforma foi instalada e configurada em um servidor com o sistema operacional Ubuntu Server 16.04, utilizando as ferramentas *Gunicorn*, *Nginx* e Supervisor, de modo a possibilitar o acesso pelos usuários. A plataforma apresentou funcionamento coerente, permitindo a escolha dos parâmetros pelo usuário em telas intuitivas, executando o processamento e retornando os arquivos resultantes.

**Palavras-chave:** Ligações de Hidrogênio. Dinâmica Molecular. *g\_ener\_hb*.

## ABSTRACT

Hydrogen Bonds (HBs) are attractive interactions usually established between a dipole formed by a highly electronegative atom (X) and a Hydrogen (X--H), with another electronegative atom (X—H $\cdots$ Y). The HBs are very important in the formation of the secondary and tertiary structures of proteins and are responsible for many of the characteristics of water that makes life possible. Due to the importance of HBs, several studies have been done with the intention of understanding the principles that govern their formation, as well as their influence on the structure of biomolecules. One of the main biases for the study of LHS is the computational approach, through the use of Molecular Dynamics tools, such as the GROMACS package, which allows the identification of LHS, but not their energy. In this way, the tool g\_ener\_hb was developed, presenting the ability to identify and measure the energies of the LHS formed in systems derived from simulation made by GROMACS. However, due to the fact that its installation and configuration can be challenging for people with no experience in Linux operating systems, this work was executed in order to facilitate the study of LHS by producing an online platform based on the g\_ener\_hb and GROMACS tools analysis process , capable of analyzing the energy of the LHS, in a simple and effective way. The platform was developed using Python as the main programming language, along with the Django framework, for organizing and optimizing web development. The SQLite3 database manager was used to store the parameters of processing and user data. HTML, CSS and JavaScript were used in the development of the interface. The JS JQuery and NGL libraries were used for interaction with the server, and for visualizing proteins, respectively. Finally, the platform was installed and configured on a server with the Ubuntu Server 16.04 operating system, using the Gunicorn, Nginx and Supervisor tools, in order to allow users access. The platform presented coherent operation, allowing the user to choose the parameters in intuitive screens, executing the processing and returning the resulting files.

**Keywords:** Hydrogen Bonds. Molecular Dynamics. g\_ener\_hb.

## LISTA DE FIGURAS

Figura 1- Ligação de Hidrogênio.....	14
Figura 2 - Esquema da execução da análise pelo script hb-analyser.py. Em verde, ferramentas pertencentes ao pacote GROMACS.. ..	22
Figura 3 -- Descrição dos campos que compõem o registro “ATOM” de um arquivo PDB. ....	23
Figura 4 – As quatro primeiras linhas dos três principais tipos de arquivos gerados pela ferramenta g_ener_hb.....	28
Figura 5 -- Representação da relação ângulo $\theta$ com a distância das LHs, tendo Oxigênio como átomo doador (a) e Carbono como doador (b).....	31
Figura 6 – Visualização da estrutura tridimensional de uma proteína com a biblioteca NGL. ....	35
Figura 7 -- Perfil de hidropatia observado após obter os resultados da utilização dos scripts para as ligações de hidrogênio intermoleculares.....	37
Figura 8 -- Processo de utilização da plataforma.....	38
Figura 9 -- Tela de login.....	39
Figura 10 --Tela de cadastro de usuário.....	39
Figura 11. Tela de upload e seleção do estado de protonação dos resíduos de Histidina e N e C-terminal.....	40
Figura 12 -- Tela de inserção dos parâmetros de funcionamento da ferramenta g_ener_hb. ....	40
Figura 13 -- Tela de escolha do campo de força e modelo atômico da água que deverão ser utilizados pelo GROMACS. ....	41
Figura 14 -- Tela de resultados. ....	42

## **LISTA DE TABELAS**

Tabela 1 – Critérios para classificação das Ligações de Hidrogênio .....	15
Tabela 2 – Diversos modos de oferecimento de serviços em nuvem. ....	18
Tabela 3 – Índices de hidropatia dos aminoácidos segundo cinco escalas diferentes: WW, KD, HK, MF e ZL, em que as letras são as iniciais dos dois primeiros autores dos trabalhos que deram origem as escalas.....	31

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>14</b>
1.1	Ligações de Hidrogênio.....	14
1.2	Dinâmica Molecular.....	16
1.3	Computação em nuvem.....	18
<b>2</b>	<b>OBJETIVO GERAL .....</b>	<b>20</b>
2.1	Objetivos Específicos .....	20
<b>3</b>	<b>METODOLOGIA.....</b>	<b>21</b>
3.1	Automatização do processamento .....	21
3.1.1	<i>pdb2gmx</i> .....	21
3.1.2	<i>editconf</i> .....	25
3.1.3	<i>grompp</i> .....	25
3.1.4	<i>g_ener_hb</i> .....	26
3.1.5	<i>pdb-color.py</i> .....	29
3.1.6	<i>energy-graph.py</i> .....	30
3.2	Desenvolvimento da plataforma.....	31
3.2.1	<i>Back-end</i> .....	32
3.2.2	<i>Front-end</i> .....	34
3.3	Configuração do servidor.....	36
<b>4</b>	<b>RESULTADOS.....</b>	<b>37</b>
4.1	Processamento .....	37
4.2	Telas .....	38
<b>5</b>	<b>DISCUSSÃO .....</b>	<b>43</b>
<b>6</b>	<b>CONCLUSÃO.....</b>	<b>44</b>
	<b>REFERÊNCIAS .....</b>	<b>45</b>
	<b>APÊNDICE A – CAMPOS DE FORÇA DISPONÍVEIS PARA USO COM O GROMACS 2016.1 .....</b>	<b>48</b>

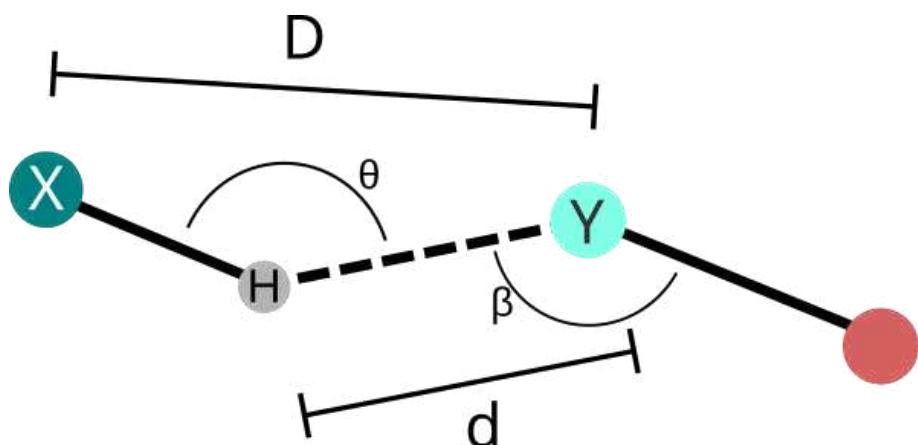
## 1 INTRODUÇÃO

### 1.1 Ligações de Hidrogênio

As ligações de hidrogênio (LHs) têm sido referenciadas na literatura desde o início do século XX, em trabalhos como os de Werner (1902) e Pfeiffer (1913), que se utilizaram dos termos *Nebervalens* (Do alemão, valência fraca ou secundária) e *Innere Komplexsalzbildung* (Ponte de sal interna complexa) para nomeá-las, tendo sido o termo “Ligação de Hidrogênio” popularizado por Pauling (1939) em seu livro *A natureza das ligações químicas* (ARUNAN et al., 2011).

Arunan et al. (2011) definem a Ligação de Hidrogênio como uma interação atrativa entre um átomo de Hidrogênio (H) ligado covalentemente à um átomo mais eletronegativo ( $X - H$ ), e outro átomo, também eletronegativo ( $X - H \cdots Y$ ), estando evidenciada a formação de ligação. Esta é a definição atualmente aceita pela IUPAC (*International Union of Pure and Applied Chemistry*). O átomo X (Doador de próton) é comumente um elemento altamente eletronegativo como Nitrogênio, Oxigênio ou Flúor, embora possa ser também menos eletronegativo como Carbono. X fica com carga parcial negativa ao atrair para si o elétron do Hidrogênio, que, consequentemente, fica com carga parcial positiva. Forma-se, então, um dipolo, que por sua vez irá interagir com um terceiro átomo, Y (Aceptor de próton) (figura 1).

Figura 1- Ligação de Hidrogênio.



Fonte: Elaborado pelo autor.

As LHS são denominadas bifurcadas ou três centros quando envolvem dois aceptores ou doadores. Às LHS formadas por três aceptores ou doadores dá-se o nome de quatro centros ou trifurcadas. Cada LH pode ser classificada como: muito forte, forte e fraca segundo os critérios geométricos, espaciais e energéticos mostrados na tabela 1 (DESIRAJU; STEINER, 2001).

Tabela 1 – Critérios para classificação das Ligações de Hidrogênio.

	<b>Distância (nm)</b>	<b>Ângulo θ (º)</b>	<b>Energia (kcal/mol)</b>
<b>Muito forte</b>	0,12 – 0,15	175 – 180	15 – 40
<b>Forte</b>	0,15 – 0,22	130 – 180	4 – 15
<b>Fraca</b>	0,22 – 0,32	90 – 180	0 – 4

Fonte: Desiraju e Steiner (2001).

Desde a sua descoberta, a importância do entendimento das LHS é crescente e evidenciada pelo gigantesco volume de estudos que as envolvem. Bases de trabalhos científicos como *Web of Science* reportam quase 40 mil trabalhos que as citam somente de 2015 a 2018, com temas que vão de Nanotecnologia a Química de Proteínas. Essa importância atribuída é justificada, quando observa-se, por exemplo, o papel desenvolvido pelas LHS estabelecidas pelas moléculas de água que permitem a interação de grupos distantes, além de serem influenciadoras da conformação proteica, estando o correto enovelamento da proteína diretamente relacionado com o desempenho da sua função biológica. Considera-se que de 40% a 70% dos grupos C=O e N-H de proteínas globulares estão envolvidos em ligações de hidrogênio, assim como as cadeias laterais de aminoácidos como Asp, Glu, Asn, Gln e Arg, sendo estas interações importantes contribuintes da manutenção da sua estrutura secundária (BAKER; HUBBART, 1984). Além disso, o reconhecimento do substrato por enzimas é muitas vezes dependente da ação das LHS, que estabilizam os aceptores e doadores de elétrons carregados no ligante e no sítio de ligação, evitando repulsões eletrostáticas. Assim, as LHS contribuem significativamente para a energia de ligação entre enzima e substrato (FERSHT, 1987).

## 1.2 Dinâmica Molecular

O método de DM foi desenvolvido por Alder e Wainwright (1957) ao criarem um sistema em que os átomos foram modelados como esferas rígidas, onde as colisões são perfeitamente elásticas. O princípio da DM baseia-se na integração da equação de movimento de Newton (equação 1), utilizando a força resultante da derivada do potencial de interação (equação 2). A energia potencial,  $V$ , é composta pela soma dos potenciais de átomos ligados e não-ligados. O potencial entre átomos ligados é resultado da adição dos potenciais associados aos comprimentos, ângulos de ligação e diedros da molécula. Já o potencial entre átomos não ligados, leva em consideração as interações eletrostática e de van der Waals. Vide a equação 3:

$$F_i = m_i \frac{\partial^2 r_i}{\partial t^2}, i = 1, 2, 3 \dots N \quad (1)$$

$$F_i = -\frac{\partial V}{\partial r_i}, i = 1, 2, 3 \dots N \quad (2)$$

Em que  $r_i$  e  $m_i$  são a posição e a massa da partícula  $i$ , respectivamente.  $t$  representa o tempo; e  $F_i$  é a força exercida sobre a partícula  $i$ .

$$V = \frac{1}{2} \sum_{ligações} K_b (b_{eq} - b)^2 + \frac{1}{2} \sum_{ângulos} K_\theta (\theta_{eq} - \theta)^2 + \frac{1}{2} \sum_{torções} V_n [\cos(n\phi - \gamma)] \quad (3)$$

$$+ \frac{1}{2} \sum_{pares\ não\ ligados} i-j \left[ \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}} \right]$$

No cálculo de  $V$ , os componentes de ligação e ângulo, variam com o desvio dos valores de comprimento da ligação,  $b$ , e ângulo,  $\theta$ , entre duas ligações químicas concomitantes, em relação aos valores de equilíbrio,  $b_{eq}$  e  $\theta_{eq}$ , respectivamente.  $K_b$  e  $K_\theta$  são as constantes de Hook da força para restituição de  $b$  e  $\theta$  aos seus valores de equilíbrio.  $V_n$  é a barreira energética para torção, enquanto  $n$ ,  $\phi$  e  $\gamma$ , representam o máximo (ou o mínimo) de energia de uma torção completa, ângulo diedro e ângulo de fase, respectivamente. Sendo  $i$  e  $j$  os átomos não ligados,  $r_{ij}$  representa a distância entre eles enquanto  $q_i$  e  $q_j$  suas respectivas cargas.

Objetivando viabilizar e, consequentemente, popularizar a aplicação dos princípios da DM no estudo de sistemas biológicos, foram desenvolvidas diversas ferramentas voltadas para simulação, cada uma com características ímpares, sendo

os mais populares: AMBER, NAMD, CHARMM, LAMMPS, HOOMD-Blue, Towhee, Cassandra e a utilizada neste trabalho, GROMACS. Com exceção de AMBER, NAMD e CHARMM, todos são de código aberto, podendo ser usados tanto para fins acadêmicos, quanto comerciais, além de recebem apoio da comunidade de desenvolvedores (CUMMINGS; GILMER, 2019).

O GROMACS (Groningen Machine for Chemical Simulations) (BERENDSEN; VAN DER SPOEL; VAN DRUNEN, 1995) é um pacote composto de diversas ferramentas voltadas para simulação molecular. É produto de um projeto inicialmente desenvolvido da parceria entre o Departamento de Química e o Departamento de Ciência da Computação da Universidade de Groningen, Holanda. O objetivo era a criação de um sistema de computação paralela para simulação molecular. Tendo em vista o crescente abandono de linguagens de programação como Pascal entre os cientistas da computação, na época, decidiu-se utilizar a linguagem C para construí-lo. Hoje, os programas que compõem o pacote estão sendo portados para C++, alcançando por volta de dois milhões de linhas de código (ABRAHAM *et al.*, 2015). O GROMACS veio com a proposta de possibilitar a execução de simulações de DM, sendo especialmente direcionado para simulação de macromoléculas biológicas em meio aquoso e membranar (VAN DER SPOEL *et al.*, 2005). Desde então, seu uso tornou-se possível não apenas em supercomputadores, mas também em máquinas comuns, como computadores pessoais. Assim, o GROMACS tem sido largamente utilizado em diversos projetos científicos e tecnológicos.

Embora eficiente na execução de simulações de DM, o GROMACS nem sempre é facilmente instalável, podendo requerer etapas de configuração, que podem se tornar complexas para quem não tem familiaridade com o funcionamento de sistemas operacionais Linux e com execução de programas por linha de comando. Além disso, sua utilização não é trivial, uma vez que envolve a execução de diversas ferramentas pertencentes ao pacote, que por sua vez geram uma série de arquivos que precisam ser corretamente gerenciados. Uma solução para este problema é a computação em nuvem, através da qual, o processamento passa a ser realizado em um servidor, eliminando a necessidade de instalações e configurações pelo usuário.

### 1.3 Computação em nuvem

O termo “computação em nuvem” ou *cloud computing* pode ser aplicado tanto a serviços prestados na internet, como também à disponibilização de infraestrutura para execução, gerência e desenvolvimento de projetos (ARMBRUST *et al.*, 2010). Existem diversos modelos de prestação de serviço por computação em nuvem: Software como Serviço (*Software as a Service* - SaaS), Plataforma como Serviço (*Platform as a Service* - PaaS), Infraestrutura como Serviço (*Infrastructure as a Service* - IaaS) e Dados como Serviço (*Data as a Service* – DaaS), são formas genéricas de denominá-los, uma vez que diversos modelos mais especializados têm surgido, como os mostrados na tabela 2 (EBEJER *et al.*, 2013).

Tabela 2—Diversos modos de oferecimento de serviços em nuvem.

X-as-a-Service terms	Abbreviation
Infrastructure as a Service	IaaS
Platform as a Service	PaaS
Software as a Service	SaaS
Network as a Service	NaaS
Storage as a Service	STaaS
Security as a Service	SECaS
Data or Desktop as a Service	DaaS
Database as a Service	DBaaS
Test Environment as a Service	TEaaS
API as a Service	APIaaS
Backend as a Service	BaaS
Integrated Development Environment as a Service	IDEaaS
Integration Platform as a Service	IPaaS

Fonte: Ebejer *et al.* (2013).

O SaaS é a forma geral de denominar os modelos que permitem a utilização de aplicações na internet pelo consumidor final, compreendendo ferramentas como e-mails, serviços de armazenamento, redes sociais, etc. Essa abordagem têm se popularizado principalmente devido a esse tipo de serviço eliminar a necessidade de instalações e configurações, por vezes complexas, o que acarreta em uma economia de tempo que pode ser aplicado na utilização da ferramenta. Outro fato é que os softwares que funcionam na máquina do usuário estão constantemente expostos a instabilidades, uma vez que atualizações nos sistemas operacionais

provocam uma série de modificações, que podem afetar o bom funcionamento do programa (HAYES, 2008).

É importante ressaltar que as aplicações que funcionam na nuvem não demandam hardware potente do usuário, podendo ser acessadas de qualquer máquina e em qualquer local, desde que se tenha acesso à internet, resultando em menos gastos (HASHEM *et al.*, 2015). Além disso, os serviços oferecidos desta forma costumam ter interface amigável, o que é uma vantagem quando comparado, neste sentido, a ferramentas que funcionam por linha de comando, tendo em vista que nem todos tem familiaridade com este ambiente.

Na Bioinformática, a utilização dos diversos modelos de computação em nuvem vem sendo aplicada visando facilitar e simplificar a utilização das ferramentas, fornecendo recursos computacionais sob demanda para o pesquisador (EBEJER *et al.*, 2013), trazendo celeridade e eficiência ao desenvolvimento dos projetos científicos. Contudo, as soluções em nuvem para análises de bioinformática estrutural são escassas, principalmente no que diz respeito a análise de Ligações de Hidrogênio em biomoléculas. Assim, este trabalho vem com a finalidade de apresentar uma solução, ao prover uma ferramenta de análise de LHS online, capaz de identificar e quantificar a energia das LHS em sistemas proteína-água.

## 2 OBJETIVO GERAL

Desenvolver uma plataforma web voltada para detecção e análise geométrica e energética das Ligações de Hidrogênio em sistemas proteína-água.

### 2.1 Objetivos Específicos

- Automatização do processamento dos arquivos de entrada e geração dos resultados.
- Desenvolvimento de interface intuitiva para entrada e visualização dos dados.
- Criação e configuração do servidor de hospedagem da plataforma.

### 3 METODOLOGIA

A plataforma desenvolvida neste trabalho tem seu funcionamento baseado em três etapas principais: Recebimento do arquivo e dos parâmetros de entrada, processamento, e visualização dos resultados. Na próxima seção será abordada a automatização do processamento (figura 2) através da construção de um *script* (*hb-analyser.py*), responsável pela execução de ferramentas pertencentes ao pacote GROMACS, da ferramenta *g\_ener\_hb* e dos *scripts* *pdb-color.py* e *energy-graph.py*, respectivamente.

#### 3.1 Automatização do processamento

No processo de análise das LHS pelo *hb-analyser.py*, o GROMACS é utilizado para a preparação dos arquivos que deverão ser recebidos pela ferramenta *g\_ener\_hb*, o que envolve uma série de etapas (figura 2), descritas nas próximas seções.

##### 3.1.1 *pdb2gmx*

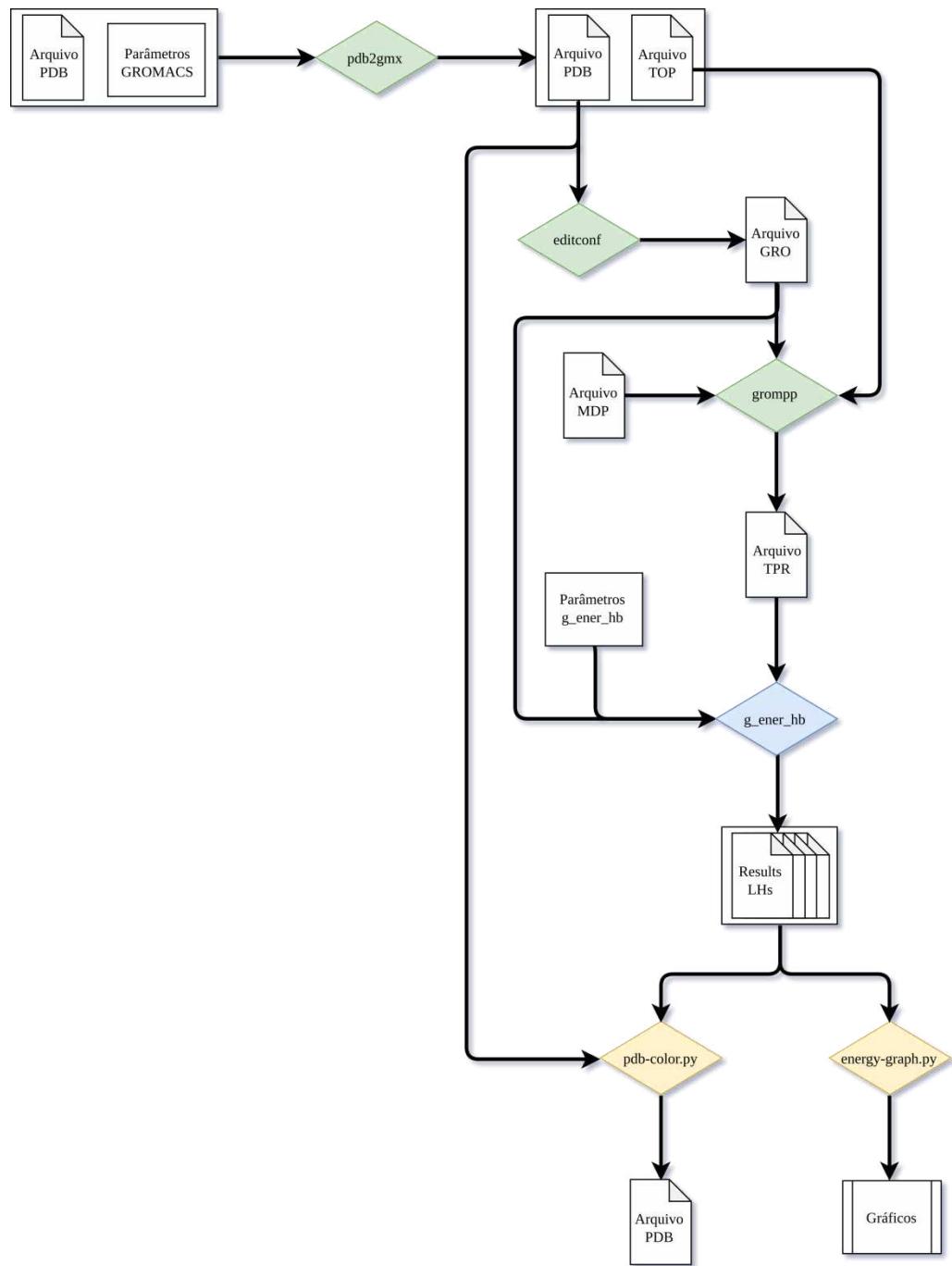
A primeira etapa do processo de análise consiste na execução da ferramenta *pdb2gmx*, pertencente ao pacote GROMACS, com o arquivo de entrada (PDB) e os parâmetros escolhidos pelo usuário: Campo de força, modelo atômico da água, e estado de protonação dos resíduos N e C-terminal e dos resíduos de Histidina. Abaixo, cada um destes é explicado:

###### 3.1.1.1 Arquivo PDB

PDB é o formato de arquivo geralmente usado para representação das estruturas tridimensionais de macromoléculas, sendo o padrão requerido para deposição de estruturas no *Protein Data Bank* (BERMAN *et al.*, 2006). Um arquivo PDB fornece diversas informações a respeito da(s) molécula(s) e/ou sistema representado(s): nome das moléculas, informações das estruturas primária e secundária, referência a bancos de dados de sequências, detalhes da coleta de dados, resolução da estrutura, citações bibliográficas, assim como, fornece as coordenadas, nos eixos x, y e z de cada átomo. Embora rico em informações, somente os dados contidos nos registros ATOM (figura 3) são essenciais para o funcionamento do GROMACS. As demais informações podem atrapalhar o processamento, de modo que, uma filtragem deve ser feita no arquivo, sendo

mantidos apenas os registros ATOM (figura 3) e os registros HETATM (registro semelhante ao ATOM, mas contendo dados de átomos não pertencentes ao polímero principal), quando referentes às moléculas de água.

Figura 2 - Esquema da execução da análise pelo script hb-analyser.py. Em verde, ferramentas pertencentes ao pacote GROMACS. Em azul, ferramenta g\_ener\_hb. Em amarelo, os scripts de pós-processamento.



Fonte: Elaborado pelo autor.

Figura 3 -- Descrição dos campos que compõem o registro “ATOM” de um arquivo PDB.

	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>	<b>e</b>	<b>f</b>	<b>g</b>	<b>h</b>	<b>i</b>	<b>j</b>	<b>k</b>
ATOM	1	N	MET	A	1	81.915	55.351	46.487	1.00	-6.72	N
ATOM	2	H1	MET	A	1	81.171	54.728	46.245	1.00	13.40	H
ATOM	3	H2	MET	A	1	81.643	55.906	47.273	1.00	13.40	H
ATOM	4	H3	MET	A	1	82.115	55.948	45.710	1.00	-6.72	H
ATOM	5	CA	MET	A	1	83.131	54.563	46.833	1.00	13.40	C
ATOM	6	CB	MET	A	1	83.573	53.785	45.611	1.00	13.40	C
ATOM	7	CG	MET	A	1	84.792	52.935	45.815	1.00	13.40	C
ATOM	8	SD	MET	A	1	85.327	52.219	44.252	1.00	13.40	S
ATOM	9	CE	MET	A	1	86.493	53.557	43.700	1.00	13.40	C
ATOM	10	C	MET	A	1	84.155	55.612	47.171	1.00	13.40	C
ATOM	11	O	MET	A	1	84.558	56.345	46.275	1.00	13.40	O
ATOM	12	N	LYS	A	2	84.544	55.739	48.439	1.00	0.00	N
ATOM	13	H	LYS	A	2	84.189	55.120	49.140	1.00	0.00	H
ATOM	14	CA	LYS	A	2	85.502	56.798	48.800	1.00	0.00	C
ATOM	15	CB	LYS	A	2	85.568	57.012	50.330	1.00	0.00	C
ATOM	16	CG	LYS	A	2	86.725	56.349	51.069	1.00	0.00	C
ATOM	17	CD	LYS	A	2	87.095	57.229	52.310	1.00	0.00	C
ATOM	18	CE	LYS	A	2	88.291	56.674	53.166	1.00	0.00	C
ATOM	19	NZ	LYS	A	2	88.010	55.408	54.014	1.00	0.00	N
ATOM	20	HZ1	LYS	A	2	88.838	55.150	54.512	1.00	0.00	H
ATOM	21	HZ2	LYS	A	2	87.275	55.599	54.665	1.00	0.00	H
ATOM	22	HZ3	LYS	A	2	87.735	54.658	53.412	1.00	0.00	H
ATOM	23	C	LYS	A	2	86.882	56.575	48.168	1.00	0.00	C
ATOM	24	O	LYS	A	2	87.381	55.450	48.162	1.00	0.00	O

**a** – serial – Número de série do átomo.

**b** – name – Nome do átomo.

**c** – resName – Nome do resíduo.

**d** – chainID – Identificador da cadeia.

**e** – resSeq – Número de sequência do resíduo .

**f** – x – Coordenada ortogonal para x em angstroms.

**g** – y – Coordenada ortogonal para y em angstroms.

**h** – z – Coordenada ortogonal para z em angstroms.

**i** – occupancy – Ocupância.

**j** – tempFactor – Fator de temperatura.

**k** – element – Símbolo do elemento.

Fonte: Elaborado pelo autor.

### 3.1.1.2 Campos de Força

Os campos de força (CF) são responsáveis por prover as equações, a partir das quais são calculadas as energias potenciais e as forças atuantes sobre cada átomo (seção 1.2). Também são definidos pelos CF os parâmetros necessários a estes cálculos. Além disso, os CF são encarregados de definir restrições de movimento, bem como, a nomenclatura para diferentes tipos de átomos e grupos funcionais.

O GROMACS oferece 15 opções diferentes de CF (Apêndice A) divididos em quatro tipos: AMBER (*Assisted Model Building and Energy Refinement*), CHARMM (*Chemistry at HARvard Macromolecular Mechanics*), GROMOS e OPLS (*Optimized Potential for Liquid Simulations*).

### 3.1.1.3 Modelo atômico da água

Outra informação requerida para o processamento é o modelo atômico da água que o usuário deseja utilizar. Classificando-se pelo número de pontos de interação, existem três tipos principais de modelos: três pontos (SPC, SPC/E, TIP3P e TIP3P), quatro pontos (TIP4P, TIP4P-Ew) e cinco pontos (TIP5P). Quanto mais pontos tem o modelo, mas fidedigno estruturalmente deve ser à realidade. Contudo, o aumento de pontos também implica no aumento do poder de processamento requerido na simulação. Outros fatores importantes para este trabalho são a distância e a energia do mínimo do potencial de Lennard-Jones, que variam de acordo com o modelo. Como será visto na seção 3.1.4, tanto o número de pontos e suas cargas parciais, quanto os parâmetros do potencial de LJ são importantes no cálculo da energia das LHs. Os modelos de Carga Pontual Simples (*Simple Point Charge – SPC*) e de Potencial Intermolecular Transferível de 3 pontos (*Transferable Intermolecular Potential with 3 Points – TIP3P*) são os modelos de água padrão do GROMACS, sendo mais simples em relação aos demais, uma vez que possuem menos pontos de interação, desta forma atingindo alto desempenho computacional.

### 3.1.1.4 Protonação das Histidinas e aminoácidos terminais

Após a escolha do modelo atômico da água, a ferramenta requer os estados de protonação de determinados resíduos (Lys, Asp, Glu, Cis ou His) e dos aminoácidos N e C-terminal. No caso desta ferramenta, restringiu-se a escolha da protonação apenas às Histidinas e aos aminoácidos das extremidades N e C-terminal, de modo a tornar o processo de submissão simplificado. Três estados de protonação e um de complexação são dados como opção para cada Histidina, isto, devido à capacidade do seu anel imidazólico de ganhar ou perder dois prótons, um em cada nitrogênio,  $\delta$  e  $\epsilon$ , além de poder estabelecer ligação covalente com um grupo Heme. No caso da Lys, Asp, Glu e Cis, o GROMACS deve utilizar as opções padrão. Quanto aos aminoácidos da extremidade N e C-terminal, dois estados de protonação são dados para cada. Para ambos é possível o estado protonado ( $\text{NH}_3^+$  e  $\text{COOH}$ , respectivamente), assim como o estado desprotonado ( $\text{NH}_2$  e  $\text{COO}^-$ ).

Após o término do processamento, a ferramenta *pdb2gmx* gera um arquivo de estrutura PDB baseado no arquivo de entrada, mas diferindo pela protonação dos

resíduos; gera também um arquivo de topologia (.top), no qual estão descritas, dentre outras informações, os tipos de moléculas, o campo de força escolhido, cada átomo e sua carga parcial, ligações, pares, e diedros formados.

### **3.1.2 *editconf***

A ferramenta *editconf* é responsável pela segunda fase do processamento, pertencente ao pacote GROMACS, tem por função a criação de uma caixa no qual a proteína será alocada com o solvente. Nesta etapa, essa ferramenta é utilizada duas vezes. A primeira para detecção das dimensões da proteína, e a segunda para a criação da caixa com um tamanho adequado e na qual a proteína será centralizada. Isto é feito com base no seguinte cálculo:

$$D = M_d + 3R_c \quad (1)$$

Onde  $D$  é a dimensão da caixa nos três eixos (x, y e z),  $M_d$  é o valor da maior dimensão da proteína, e  $R_c$  é o raio de corte utilizado no sistema, que representa a distância máxima, no qual um átomo pode estabelecer algum tipo de interação com outros. Estabelecer uma caixa de tamanho razoável é importante para evitar que um átomo da proteína ou do solvente interaja com átomos que deveriam estar espacialmente inacessíveis. Embora o GROMACS permita a utilização de caixas nas formas de dodecaedro rômbico e octaedro truncado, neste trabalho estabeleceu-se como padrão a utilização de caixas cúbicas, visando, principalmente, a simplicidade no cálculo das dimensões. Esta etapa tem como resultado um arquivo de estrutura (.gro). O formato de arquivo .gro é padrão do GROMACS, sendo semelhante ao PDB, mas, neste caso, possuindo apenas informações referentes ao posicionamento dos átomos e ao tamanho da caixa na qual a proteína está inserida.

### **3.1.3 *grompp***

Por fim, a ferramenta *grompp* é executada, recebendo um arquivo de topologia (.top), um arquivo de configuração (.mdp) e o arquivo de estrutura (.gro) gerado pela ferramenta *editconf*. Faz, então, a expansão das informações de topologia do nível de descrição molecular para atômico. Tem como saída um arquivo binário (.tpr) que contém a estrutura inicial e a topologia das moléculas do sistema, além de todos os parâmetros da simulação. Tanto o arquivo .tpr aqui gerado, quanto

o arquivo .gro gerado pelo *editconf* são requisitos para execução da próxima ferramenta, *g\_ener\_hb*.

### **3.1.4 *g\_ener\_hb***

Embora não haja um termo explícito que descreva a energia das LHS nos cálculos de energia potencial, estas podem ser modeladas através da combinação dos termos de LJ e Coulomb, reproduzindo bem as geometrias e energias (GUVENCH; MACKERELL, 2008).

É possível analisar as LHS com uma ferramenta do próprio *gromacs*: *hbond*. No entanto, assim como as outras ferramentas encontradas na literatura, para esse fim, a *hbond* é incapaz de calcular a energia envolvida nas ligações formadas. O mesmo ocorre para outras ferramentas instaláveis como *MDAnalysis* (MICHAUD-AGRAWAL et al., 2011) e *CPPTRAJ* (ROE; CHEATHAM, 2013) e para servidores como o HBCS (GURUSARAN et al., 2016). Para sanar esse problema foi desenvolvida a ferramenta *g\_ener\_hb* que além de identificar as possíveis LHS, reportando seus tamanhos e ângulos, calcula a energia média das ligações formadas por cada dipolo C=O e N-H presente na cadeia principal e nas cadeias laterais da proteína. Para isso, recebe tanto o arquivo .tpr gerado pelo *grompp*, quanto o arquivo .gro gerado pelo *editconf* (LOURENZONI, 2005).

Embora o *g\_ener\_hb* seja capaz de processar dados de trajetória, ou seja, de arquivos contendo vários estados de um sistema molecular em função do tempo, a plataforma desenvolvida neste trabalho visa à análise de arquivos PDB contendo somente um único estado do sistema. Essa restrição se deve a três motivos principais:

- a) Arquivos de trajetória são geralmente grandes, o que poderia causar demora no upload a depender da velocidade e qualidade da internet do usuário;
- b) A criação de um arquivo .tpr, neste caso, envolveria a execução de ferramentas como *pdb2gmx* para cada estado do sistema, uma vez que não são voltadas para análise de trajetórias. Deste modo, o processamento poderia se tornar oneroso computacionalmente, sobrecarregando o servidor;

- c) Embora o arquivo .tpr pudesse ser fornecido pelo usuário, teria de ser criado com a mesma versão do GROMACS usada no servidor (2016), havendo, caso contrário, incompatibilidades que inviabilizariam a análise.

A ferramenta g\_ener\_hb gera uma série de arquivos contendo os dados sobre as LHs formadas; esses dados são disponibilizados em três níveis de detalhe: resíduo, dipolo e par dipolo-dipolo. No primeiro caso (resíduo) são gerados arquivos que informam resumidamente dados das LHs formadas por cada resíduo da proteína. No segundo (dipolo) são fornecidos arquivos com dados das LHs formadas por cada dipolo. O último nível, dipolo-dipolo, consiste em arquivos onde são especificados cada par de dipolos formadores de LH da proteína.

Nos níveis de resíduos e dipolos, os arquivos são separados em LHs formadas pela cadeia lateral e cadeia principal. São divididos ainda entre: intramoleculares, quando a LH for formada entre resíduos/dipolos da proteína, e intermolecular, quando a LH for formada entre resíduos/dipolos da proteína com a água.

Os arquivos de par dipolo-dipolo estão separados em: interação da cadeia lateral com a cadeia principal, cadeia lateral com cadeia lateral e cadeia principal com cadeia principal. Os arquivos dipolo-dipolo restringem-se às LHs intramoleculares.

Todos os arquivos possuem basicamente as mesmas colunas, com algumas exceções. Onde, V avg, D avg, Theta avg, Beta avg e N avg, são as médias da energia potencial (kcal), distância entre X e Y (angstroms), ângulo  $\theta$  ( $^{\circ}$ ), ângulo  $\beta$  ( $^{\circ}$ ) e número médio de ligações formadas, respectivamente. Enquanto as colunas sd são seus respectivos desvios-padrão. A coluna %OBS presente nos arquivos de nível par dipolo-dipolo, é útil apenas quando analisa-se arquivos de trajetória, uma vez que os valores dessa coluna informam a porcentagem do tempo em que determinada LH esteve formada.

Figura 4 – As quatro primeiras linhas dos três principais tipos de arquivos gerados pela ferramenta g\_ener\_hb. a) Resultado em nível de par dipolo-dipolo. b) Resultado em nível de dipolo. c) Resultado em nível de resíduo.

a)

aa1	at11	at12	at21	at22	aa2	V avg	Vsd	D avg	D sd	theta	theta_sd	beta	beta_sd	N avg	N sd	% OBS	n
29 SER	N	HN	0	C	1 GLU	-2.2334	0.0004	0.1950	0.0010	170.3	0.338	132.8	0.781	0.0040	0.0631	0.40	
4 LEU	N	HN	0	C	2 GLN	-1.2288	0.0008	0.1967	0.0007	150.7	0.721	94.9	0.892	0.0040	0.0631	0.40	
9 ASP	N	HN	0	C	6 SER	-1.0021	0.2850	0.2056	0.0130	159.0	4.531	107.9	1.553	0.0080	0.0891	0.80	

b)

res	at1	at2	HBs	V avg	Vsd	D avg	Dsd	Theta	T sd	Beta	B sd	N avg	N sd
1 (GLU )	0	C	0.004000	-2.2334	0.0004	0.1950	0.0010	170.3	0.338	132.8	0.781	0.0040	0.0630
2 (GLN )	0	C	0.004000	-1.2288	0.0008	0.1967	0.0007	150.7	0.721	94.9	0.892	0.0040	0.0630
3 (LYS )	0	C	0	0	0	0	0	0	0	0	0	0	0

c)

res	HBs	V avg	Vsd	D avg	D sd	Theta	T sd	Beta	B sd	N avg	N sd
1 (GLU )	0.004000	-2.2334	0.0004	0.1950	0.0010	170.3	0.338	132.8	0.781	0.0040	0.0630
2 (GLN )	0.004000	-1.2288	0.0008	0.1967	0.0007	150.7	0.721	94.9	0.892	0.0040	0.0630
3 (LYS )	0.004000	-1.1707	0.0007	0.1988	0.0008	153.0	0.984	140.4	0.398	0.0040	0.0630

### 3.1.5 *pdb-color.py*

O script *pdb-color* foi desenvolvido com a linguagem de programação Python, com a finalidade de possibilitar a visualização de parte dos resultados produzidos pela ferramenta *g\_ener\_hb* na etapa anterior. O script é utilizado duas vezes, uma para as LHs intramoleculares e outra para as intermoleculares. Três arquivos são recebidos: o primeiro é o arquivo PDB derivado da execução da ferramenta *pdb2gmx*; o segundo e terceiro são arquivos criados pelo *g\_ener\_hb* que descrevem os dipolos envolvidos na formação de LHs pela cadeia principal e pela cadeia lateral da proteína, e a energia destas ligações. Recebe ainda, como parâmetro, o identificador de uma das escalas de hidropatia disponíveis (explicado posteriormente). Uma vez identificados os átomos formadores de LHs, o script substitui o valor de *b-factor* da linha referente ao átomo em questão no arquivo PDB (figura 3), pelo valor da energia da LH na qual este átomo está envolvido.

Aos átomos não envolvidos na formação de LHs é atribuído o índice de hidropatia do resíduo no qual estão inseridos. Os índices de hidropatia indicam o quão hidrofóbico são os resíduos, e são oriundos da escala escolhida pelo usuário (tabela 3). Os valores negativos das escalas HK e MF indicavam, ao contrário das demais, maior hidrofobicidade, por isso, todos os valores de ambas as escalas tiveram seus sinais invertidos.

Como pode ser observado na tabela 3, as escalas possuem variações diferentes, tendo mínimos e máximos distintos. Assim, antes de ser usada como fonte dos índices de hidropatia dos resíduos, a escala escolhida pelo usuário deve ser normalizada segundo a Fórmula 2. Deste modo, a escala normalizada terá apenas valores positivos, e seu maior valor será exatamente oposto à energia de LH mais negativa encontrada na estrutura. Assim, após todas as substituições, todos os valores negativos e positivos presentes na coluna de *b-factor* serão referentes às energias das LHs formadas e aos índices de hidropatia dos aminoácidos, respectivamente.

$$HP'_X = \frac{LH_{min}}{HP_{min} - HP_{max}} (HP_X - HP_{min}) \quad (2)$$

Em que  $HP'_X$  é o índice de hidropatia atribuído ao resíduo após a normalização.  $LH_{min}$  representa o valor de energia de LH mais negativo encontrado

no sistema.  $HP_{min}$  e  $HP_{max}$  são os índices de hidropatia mais negativo e mais positivo da escala, respectivamente.  $HP_x$  é o índice de hidropatia do aminoácido antes da normalização.

Tabela 3 – Índices de hidropatia dos aminoácidos segundo cinco escalas diferentes: WW, KD, HK, MF e ZL, em que as letras são as iniciais dos dois primeiros autores dos trabalhos que deram origem as escalas. Quanto maior o valor, mais hidrofóbico tende a ser o aminoácido.

Aminoácido	WW	KD	WH	MF	ZL
Trp	1,85	-0,9	-0,3	0,38	1,53
Phe	1,13	2,8	0,32	2,2	1,98
Lys	0,99	-3,9	-2,71	-5,39	-3,46
Tyr	0,94	-1,3	-0,68	1,09	0,49
Leu	0,56	3,8	0,55	1,81	1,82
Ile	0,31	4,5	0,6	1,56	1,97
Cys	0,24	2,5	0,13	-0,49	-0,3
Met	0,23	1,9	0,1	0,76	1,4
Gly	-0,01	-0,4	-0,74	-1,72	-0,19
Val	-0,07	4,2	0,31	0,78	1,46
Ser	-0,13	-0,8	-0,84	-1,83	-0,53
Thr	-0,14	-0,7	-0,52	-1,78	-0,32
Ala	-0,17	1,8	-0,11	0,0	0,38
His	-0,17	-3,2	-2,06	-4,76	-1,44
Asn	-0,42	-3,5	-2,68	-1,64	-2,9
Pro	-0,45	-1,6	-2,23	1,52	-1,44
Gln	-0,58	-3,5	-3,49	-3,95	-3,27
Arg	-0,81	-4,5	-2,58	-3,71	-2,57
Asp	-1,23	-3,5	-2,36	-3,01	-1,84
Glu	-2,02	-3,5	-2,05	-3,47	-1,62

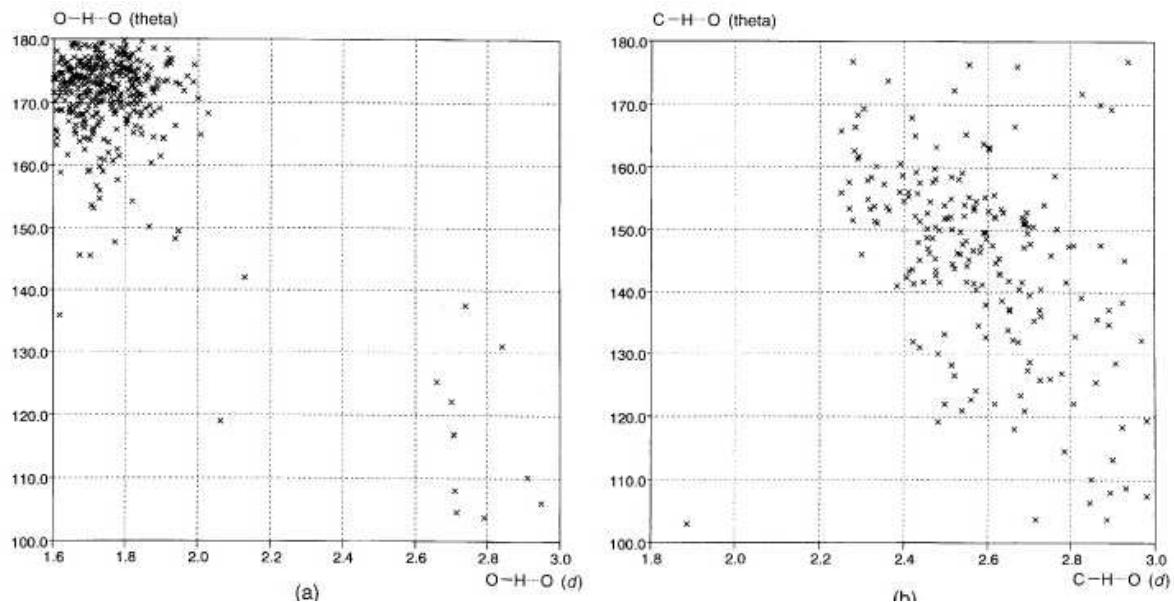
Fontes: WW - Wimley e White (1996); KD - Kyte e Doolittle (1982); MF - Moon e Fleming (2011); WH - White *et al.* (2005); ZL - Zhao e London (2006).

### 3.1.6 *energy-graph.py*

Foi desenvolvido um *script* em Python, *energy-graph*, tendo por função a construção de gráficos de dispersão que mostram a relação entre ângulo, distância e energia das LHs formadas, a partir dos dados gerados pela ferramenta *g\_ener\_hb*. A importância desse tipo de visualização reside na possibilidade de representar

graficamente e de forma geral, a natureza das LHs presentes na estrutura (Figura 5), o que, por sua vez, permite que se observe a tendência seguida pelas LHs com relação aos seus ângulos, distâncias e energias. Para a construção dos gráficos, foi utilizada a biblioteca *Matplotlib*.

Figura 5 -- Representação da relação ângulo  $\theta$  com a distância das LHs, tendo Oxigênio como átomo doador (a) e Carbono como doador (b).



Fonte: Desiraju e Steiner (2001).

### 3.2 Desenvolvimento da plataforma

Uma vez feita a automatização do processamento através do *hb-analyser.py*, iniciou-se o desenvolvimento da plataforma responsável por receber e armazenar os parâmetros e arquivos necessários à análise, executar o script de processamento repassando os parâmetros e arquivos recebidos, e disponibilizar os resultados gerados para o usuário.

Com a crescente demanda por serviços oferecidos na internet, surgiram diversas ferramentas com o objetivo de facilitar o desenvolvimento destas soluções, com foco tanto em *front-end*, que é a parte do sistema acessível ao usuário (interface), como em *back-end*, que é o lado do servidor, onde ocorre todo o processamento.

maior parte do processamento. Aprofunda-se a seguir cada uma dos componentes, exemplificando as tecnologias disponíveis e as utilizadas no desenvolvimento deste projeto.

### **3.2.1 Back-end**

O *back-end*, como já dito, é a parte da plataforma responsável pelo processamento das requisições feitas pelo usuário, assim como armazenamento de dados. A criação do *back-end* costuma ser mediada por um *framework*, uma ferramenta que tem por objetivo principal direcionar o desenvolvimento, dando celeridade e conferindo certa padronização ao projeto. A escolha do *framework* normalmente se dá com base na linguagem de programação que se pretende utilizar como principal. PHP, Java, C# e Python são exemplos de linguagens frequentemente utilizadas na construção do *back-end* de plataformas web.

#### **3.2.1.1 A linguagem Python**

No desenvolvimento desta plataforma utilizou-se a linguagem Python. Esta linguagem foi apresentada em 1991 por Guido van Rossum, seu criador, sendo tão antiga quanto o projeto precursor da linguagem Java, o *Green Project*, contudo, só nos últimos anos têm se popularizado. O Python é uma linguagem interpretada com o objetivo principal de ser simples e legível, diminuindo assim, o tempo necessário pra aprender a estruturação da sintaxe, de modo que o tempo não gasto pode ser usado no desenvolvimento, contribuindo para a produtividade do desenvolvedor. É hoje uma das linguagens mais utilizadas no meio científico, com aplicações em aprendizagem de máquina, processamento e visualização de dados, e até mesmo na simulação de macromoléculas. Isto se dá em grande medida pela disponibilidade crescente de bibliotecas que auxiliam o desenvolvimento.

Diversos *frameworks* dão suporte ao desenvolvimento web com Python. *Tornado*, *Flask*, *Bottle*, *CherryPy* são exemplos de *frameworks non-full-stack*, ou seja, que focam no *back-end*, não auxiliando diretamente o desenvolvimento da interface gráfica (*front-end*). Já *Web2Py*, *CubicWeb*, *Turbo Gears* e *Django* são classificados como *frameworks full-stack*. Mais robustos, possuem mais recursos, contribuindo também para o desenvolvimento do *front-end*. Para este projeto, o

*framework Django* foi escolhido, por conta da sua estabilidade, escalabilidade e eficiência.

### **3.2.1.2 Framework Django**

O *framework Django*, mantido pela *Django Software Foundation* (DSF). É atualmente utilizado por grandes empresas como *Instagram*, *Mozilla*, *National Geographic* e *Pinterest*, sendo considerado o principal *framework Python* para desenvolvimento web. É composto de uma série de recursos que visam dar celeridade ao desenvolvimento de plataformas web, garantindo também a segurança e a estabilidade. O *Django* é estruturado segundo o padrão *Model-View-Template* (MVT). Neste modelo o armazenamento (*Model*), o processamento (*View*) e a interface (*Template*) são definidos separadamente, mas comunicam-se entre si.

O *Django* simplifica a criação do Banco de Dados (BD) através da utilização do sistema de *Models*. Neste sistema, as tabelas do BD são estabelecidas com base em classes do Python, que definem quais e de que tipo serão as colunas destas tabelas. Estas classes também servem de canal de comunicação com o BD, além de poderem ser usadas como base para criação automática de formulários. O *Django* cria por padrão um BD relacional tendo SQLite3 como gerenciador. Sendo simples e eficiente, exigindo poucas etapas de configuração, embora não possuindo recursos sofisticados (gerenciamento de usuários, diferentes níveis de segurança, campos mais específicos), o SQLite3 cumpre o papel de guardar os dados de cadastro e os resultantes do processamento de forma rápida e estável.

Já as *Views*, podem ser consideradas como o centro do processamento. Elas são as responsáveis por receber, processar e responder às requisições vindas do navegador. A comunicação entre o navegador e o servidor se dá através do protocolo HTTP (*Hypertext Transfer Protocol*), que é um método através do qual o navegador envia e recebe dados. Há dois tipos principais de requisição HTTP recebidas pelas *Views*: GET (requisição de dados) e POST (submissão de dados). As *Views* são funções associadas a endereços URL, de modo que, quando uma URL é acessada no navegador a *View* ligada a ela é acionada, recebendo uma requisição do tipo GET, e retornando como resposta os dados necessários para a renderização da página pelo navegador (*Template*). Já quando um formulário é submetido ao navegador, por exemplo, a *View* ligada a ele recebe uma requisição

do tipo POST, sendo responsável por tratar os dados recebidos e, através dos *Models*, enviá-los para o BD.

Os *Templates* constituem o *front-end*, ou seja, a interface gráfica da aplicação web. Embora os *Templates* sejam gerenciados pelo *Django*, sua construção é diferente dos demais componentes. A seguir são descritas as tecnologias utilizadas na construção das telas.

### **3.2.2 *Front-end***

Diversas tecnologias podem ser empregadas na criação do *front-end*, inclusive *frameworks* focados nesse tipo de desenvolvimento. Tendo em vista, a não necessidade da construção de muitas telas, e a simplicidade das que deveriam ser feitas, optou-se por se utilizar apenas as tecnologias fundamentais, algumas vezes auxiliadas por bibliotecas. Segue uma introdução a estas ferramentas:

#### **3.2.2.1 HTML**

*Hypertext Markup Language* (HTML) é a linguagem de marcação, criada por Tim Berners-Lee, Robert Cailliau em 1989, está na sua 5<sup>a</sup> versão, sendo usada como padrão nos navegadores. Com estrutura simples, é utilizada para estabelecer a posição de determinados elementos na página web (divisões, textos, links, formulários, botões, imagens, etc.), não sendo capaz de fazer estilizações complexas.

#### **3.2.2.2 CSS**

Tendo seus parâmetros estabelecidos pelo *World Wide Web Consortium* (W3C), o *Cascading Style Sheets* (CSS) é utilizado com a finalidade de estilizar os elementos dispostos pelo HTML. Através deste recurso é possível mudar a cor, tamanho, forma, fonte, dentre outros atributos dos elementos, permitindo, desta forma, a construção de uma interface amigável para o usuário.

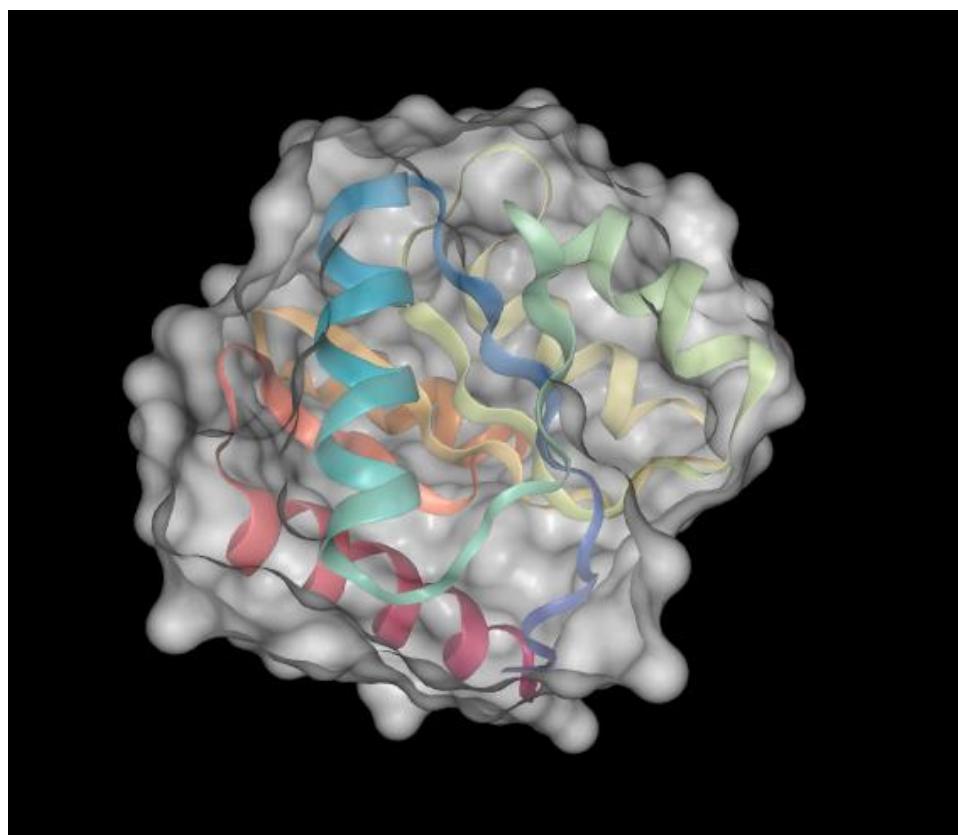
#### **3.2.2.3 JavaScript**

JavaScript (JS) é uma linguagem de programação Turing Completa, podendo armazenar variáveis de diferentes tipos (strings, inteiros, reais, listas, dicionários,

etc.), possuindo estruturas condicionais e laços de repetição, tendo, além de outros atributos, a capacidade de interagir nativamente com o *Document Object Model* (DOM), sobre o qual são estruturadas as páginas web, podendo conferir, desta forma dinamicidade às páginas antes estáticas, além de possibilitar o pré-processamento de dados antes do seu envio para o servidor.

Duas bibliotecas JS foram utilizadas neste trabalho. A primeira, JQuery, auxilia na interação com os elementos do DOM, e na comunicação com o servidor via métodos AJAX (*Asynchronous JavaScript And XML*). O NGL foi a segunda biblioteca utilizada, sendo responsável por permitir a visualização da estrutura tridimensional moléculas no navegador, inclusive proteínas. No caso de proteínas, o NGL disponibiliza diferentes formas de visualização, como: palitos, *cartoon*, superfície, dentre outros.

Figura 6 – Visualização da estrutura tridimensional de uma proteína com a biblioteca NGL.



Fonte: Elaborado pelo autor.

### 3.3 Configuração do servidor

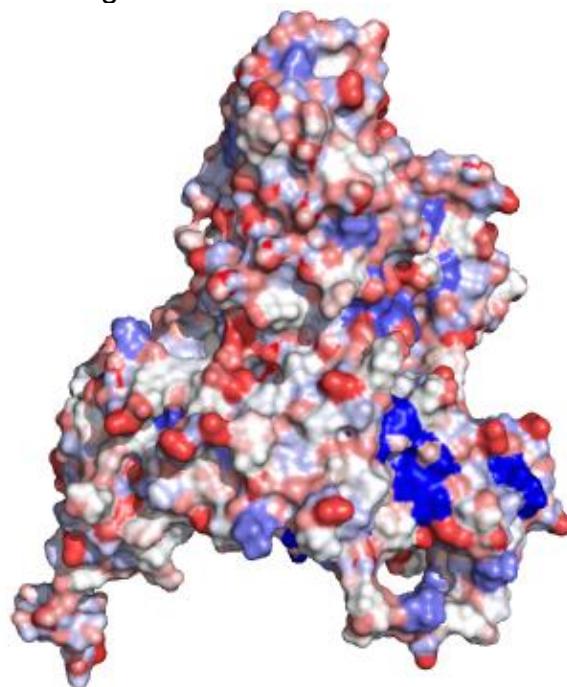
Uma vez construída a plataforma, a fase seguinte foi a instalação desta em um servidor, objetivando-se, assim, permitir o acesso aos usuários. Para isso, foi utilizada uma máquina com *Ubuntu Server 16.04 LTS* como sistema operacional. Para a configuração e estabelecimento da plataforma, foi necessária a utilização do *Nginx* como servidor HTTP. O *Nginx* deve ser o primeiro a detectar uma requisição vinda da internet. Se a requisição for por um arquivo estático (CSS, JS ou imagens), responde ao navegador automaticamente. Se a requisição for direcionada a uma *View*, o *Nginx* redireciona a requisição para o *Gunicorn* que é um servidor HTTP WSGI (*Web Server Gateway Interface*) para *frameworks Python* com a função de media a interação do servidor (*gateway*) com a aplicação (*framework*). Para garantir que o Gunicorn, Nginx, e Django permanecerão em funcionamento sempre que o servidor estiver ligado, utilizou-se o *Supervisor*, uma aplicação responsável por monitorar os processos dos programas indicados, garantindo que estarão sempre em funcionamento.

## 4 RESULTADOS

### 4.1 Processamento

Visando demonstrar o funcionamento do script `pdb-color.py` aliado aos dados gerados pelo `g_ener_hb`, foi feita a análise de um arquivo de trajetória oriundo da simulação de um sistema contendo scFv (Fragmento variável de anticorpo de cadeia simples) e o antígeno CD19 em água.

Figura 7 -- Perfil de hidropatia observado após obter os resultados da utilização dos scripts para as ligações de hidrogênio intermoleculares.



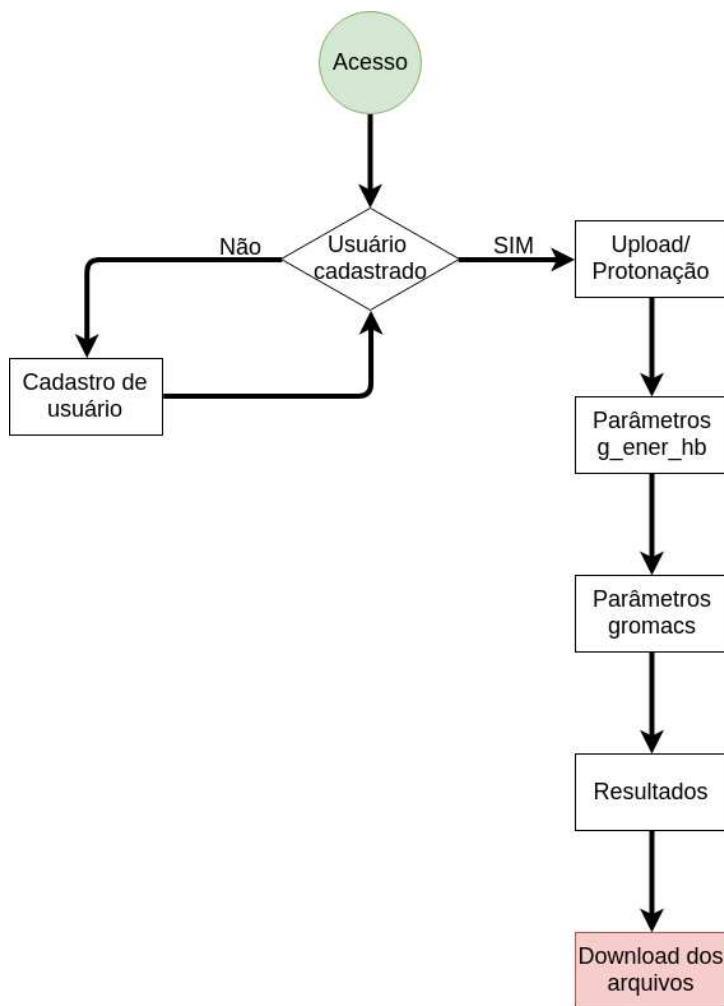
Fonte: Elaborado pelo autor.

Na figura 7 gerada com o *software* de visualização molecular *Pymol*, observou-se como se dá o perfil de formação das LHs intermoleculares, ou seja, LHs formadas entre os resíduos das proteínas com as moléculas de água. Como descrito na seção 3.1.5, as estruturas estão coloridas de acordo com a energia das LHs formadas e o perfil de hidropatia dos resíduos, onde as regiões em vermelho representam os resíduos que formam LHs com as moléculas de solvente, que nesse caso é a água, durante a trajetória. As regiões em branco representam resíduos que não formam LHs, e, por fim, as regiões em azul são referentes a resíduos que não formam LHs e que possuem alto índice de hidrofobicidade, coloridas com base na escala de hidropatia estabelecida por Wimley e White (1996).

## 4.2 Telas

Sendo um dos objetivos iniciais, a criação e implementação de uma interface intuitiva para entrada e visualização dos dados, seis telas principais foram construídas para a plataforma (figuras 9 a 14). O modo de navegação é demonstrado na figura 8.

Figura 8 -- Processo de utilização da plataforma.



Fonte: Elaborado pelo autor.

A utilização da plataforma inicia-se com o login do usuário (figura 9), que é redirecionado para o cadastro (Figura 10), caso ainda não esteja cadastrado.

Figura 9 -- Tela de login.

The screenshot shows a login form with a light gray background. At the top right is a 'Login' button. Below it is a white rectangular input field divided into two sections: 'Username:' above a text input box and 'Password:' above another text input box. A blue 'Login' button is centered below these fields. At the bottom of the form, a small link reads 'Do not have an account? [Register](#)'.

Fonte: Elaborado pelo autor.

Na tela de cadastro (figura 10), são requeridas algumas informações pessoais do usuário (nome, email e instituição), assim como os dados que serão usado para acesso à plataforma (usuário e senha).

Figura 10 --Tela de cadastro de usuário.

The screenshot shows a user registration form with a light gray background. At the top right is a 'Login' button. Below it is a white rectangular input field containing six stacked text input boxes labeled 'Username', 'Name', 'E-mail', 'Institution', 'Password', and 'Confirm password'. A blue 'Submit' button is located at the bottom right of the form.

Fonte: Elaborado pelo autor.

A partir de então o usuário é levado à primeira página da ferramenta (figura 11), onde deverá selecionar o arquivo de entrada (PDB), assim como os estados de protonação dos resíduos de Histidina e N e C-terminal.

Figura 11. Tela de upload e seleção do estado de protonação dos resíduos de Histidina e N e C-terminal.

Residue	Protonation State
N-terminal	NH3+
C-terminal	NH3+

Residue Number	Protonation State
HIS 149	H on ND1 only
HIS 156	H on ND1 only

Fonte: Elaborado pelo autor.

Clicando em “next”, o usuário será encaminhado para a tela de entrada dos parâmetros da ferramenta *g\_ener\_hb* (figura 12), nesta poderá definir os valores desejados para energias e distâncias máximas, ângulos  $\theta$  e  $\beta$  mínimos das LHS.

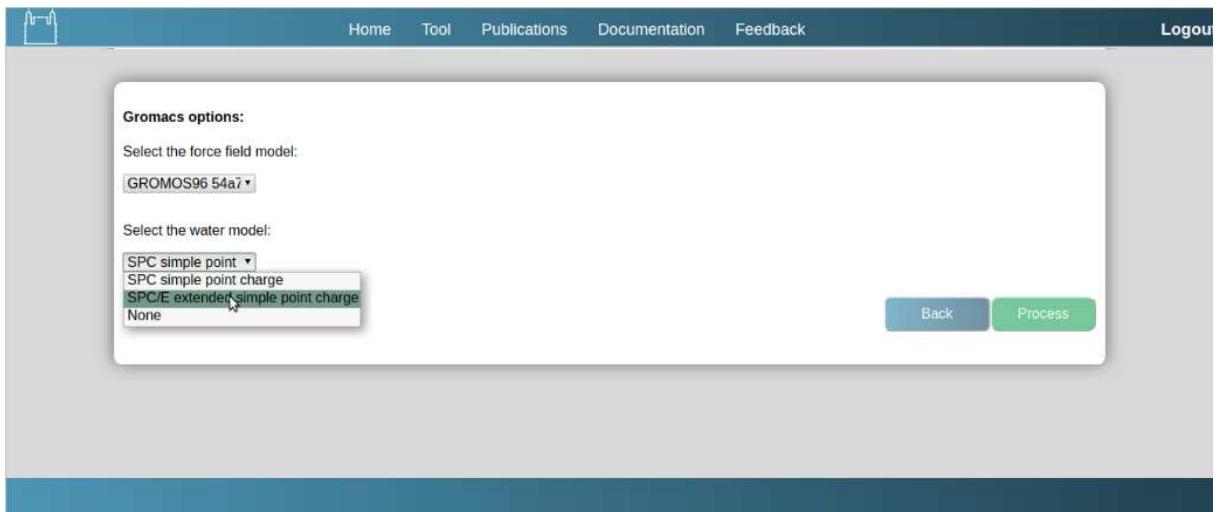
Figura 12 -- Tela de inserção dos parâmetros de funcionamento da ferramenta *g\_ener\_hb*.

In the next section you can set the options for INTRA HBs.	In the next section you can set the options for INTER HBs.
Maximum energy: -0,8	Maximum energy: -0,8
Maximum distance: 0,3	Maximum distance: 0,31
Minimum beta angle: 120	Minimum beta angle: 90
Maximum beta angle: 120	Maximum beta angle: 120
Minimum theta angle: 120	Minimum theta angle: 165
Maximum theta angle: 180	Maximum theta angle: 180

Fonte: Elaborado pelo autor.

Na próxima tela (figura 13) estarão disponíveis os campos para escolha dos parâmetros do GROMACS: campo de força e modelo atômico da água, que deverão ser usados na análise.

Figura 13 -- Tela de escolha do campo de força e modelo atômico da água que deverão ser utilizados pelo GROMACS.

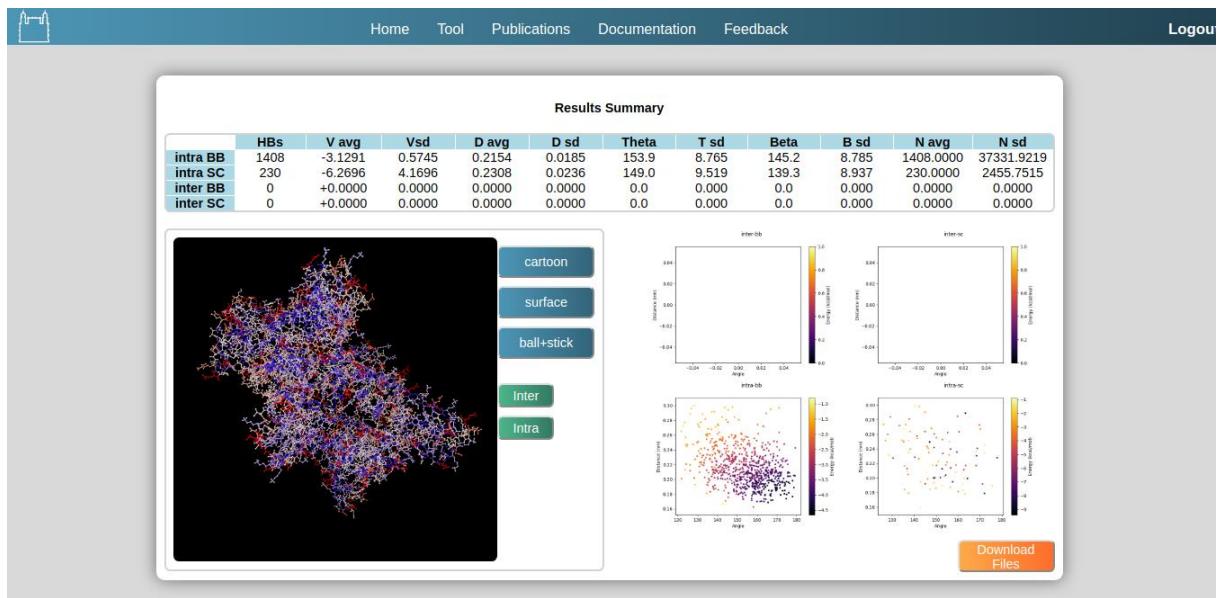


Fonte: Elaborado pelo autor.

Por fim, após o processamento, o usuário deve ser apresentado à tela de resultados (figura 14), onde será mostrado o resumo dos valores referentes às LHs formadas. Ainda na tela de resultados, a estrutura tridimensional da proteína poderá ser visualizada, estando colorida de acordo com os valores inseridos na coluna de *b-factor* do arquivo PDB pelo script *pdb-color.py* (seção 3.1.5). Ainda na tela de resultados, são disponibilizados gráficos de dispersão, nos quais são dispostos pontos representantes das LHs, de acordo com seu ângulo  $\theta$  e comprimento. Os pontos são coloridos com base na energia da LH na qual estão envolvidos. É disponibilizada ao usuário a possibilidade de fazer o download de todos os dados gerados pela ferramenta *g\_ener\_hb*, através do botão “Download”.

As páginas geradas são de entendimento intuitivo e simples, facilitando a utilização pelo usuário. O sistema de requisições a partir da construção de *views* dedicadas e da utilização do jQuery visando a dinamicidade, funcionou corretamente. Deste modo, as escolhas dos estados de protonação dos resíduos de Histidina, do campo de força e do modelo atômico da água, apresentam funcionamento responsivo.

Figura 14 -- Tela de resultados.



Fonte: Elaborado pelo autor.

A integração das *Views* com o banco de dados se deu eficientemente, de modo que os dados inseridos pelo usuário foram armazenados sem perdas, e puderam ser usados no processamento. Da mesma forma, a automatização da execução do GROMACS e do *g\_ener\_hb* ocorreu de forma eficaz, analisando os dados de entrada e retornando os resultados.

Por fim, é importante observar as diferenças existentes entre as figura 2 e 8, que representam o processo executado automaticamente e o trajeto do usuário pela plataforma até a obtenção dos resultados, respectivamente. Onde, na figura 2 observa-se um maior número de passos, quando comparada com a figura 8. A simplificação alcançada se torna ainda mais evidente ao se considerar que no caso da execução deste processo em uma máquina local, o usuário teria que, além de fazer a instalação e configuração do GROMACS e *g\_ener\_hb*, gerenciar um número considerável de arquivos gerados durante o processamento.

## 5 DISCUSSÃO

A ferramenta *g\_ener\_hb* tornou possível a análise energética das LHs formadas em sistema moleculares, sendo capaz de atribuir a uma determinada LH, uma energia de ligação. Adicionando assim, uma nova dimensão ao estudo das LHs através de métodos de Dinâmica Molecular. Seu funcionamento se dá por linha de comando, recebendo arquivos gerados por ferramentas do pacote GROMACS e gerando como resultado listas extensas das LHs formadas, seus ângulos, distâncias e energias.

Embora eficientes, a instalação do GROMACS e *g\_ener\_hb* não é simples, envolvendo o processo de compilação do pacote GROMACS e resolução de eventuais erros por falta de dependências, e outras variáveis que demandariam tempo considerável para serem resolvidos. Vale ressaltar que ambas as ferramentas são compatíveis apenas com o sistema operacional Linux e que uma série de passos é requerida para se chegar a um resultado satisfatório. Para utilização do GROMACS instalado no próprio computador, o cliente teria de entender, pelo menos de forma básica, cada ferramenta utilizada em determinado processamento. Desta forma, tanto o processo de instalação quanto utilização demanda investimento de tempo por parte do usuário.

Assim, este trabalho foi desenvolvido com o objetivo principal de criar uma plataforma web, intuitiva e simples, que permitisse a análise das LHs em sistemas moleculares proteína-água através da aplicação do GROMACS e da ferramenta *g\_ener\_hb*, sem a necessidade de instalação. Para isso, diversas outras tecnologias foram aplicadas, tanto na automatização do processamento, quanto na criação de um arcabouço capaz de possibilitar a utilização dessas ferramentas de forma prática e sem a necessidade de qualquer instalação.

A plataforma desenvolvida neste trabalho destaca-se em comparação a outras ferramentas disponíveis com propósito semelhantes, especialmente o HBCS (GURUSARAN *et al.*, 2016), por também ser uma ferramenta online, mas diferindo por não permitir a análise da energia das LHs. Além disso, a ferramenta aqui desenvolvida fornece a visualização dos resultados incorporados à estrutura tridimensional da proteína através da biblioteca NGL, disponibilizando ainda uma série de gráficos produzidos a partir dos resultados gerados na análise.

## 6 CONCLUSÃO

A automatização do processamento foi atingida com êxito através da construção do *script hb-analyser.py* responsável pela integração do GROMACS, *g\_ener\_hb* e *scripts pdb-color.py* e *energy-graph.py*, utilizando sequencialmente os arquivos de saída de um como entrada do outro. Uma interface simples e objetiva foi criada visando a inserção dos dados pelo usuário, assim como a visualização dos resultados. Para isso, diversas ferramentas de *front-end* foram empregadas, de modo a fornecer uma boa experiência de utilização ao usuário. Por fim, ambas as partes, processamento e interface, foram integradas através da utilização do *framework Django*, responsável por gerenciá-las, como também ao Banco de Dados. Por fim, a plataforma foi instalada e configurada em um servidor com o auxílio das ferramentas Gunicorn, Nginx e Supervisor, concluindo-se, assim, a criação de uma plataforma web para análise das Ligações de Hidrogênio em sistemas moleculares de Proteína e Água.

### 6.1 Trabalhos futuros

Diversas implementações são possíveis a partir da plataforma desenvolvida neste trabalho. Abaixo, exemplo de melhorias que poderiam ser integradas:

- Análise automática dos estados de protonação dos resíduos com base no pH através da execução do ProPka, dispensando a necessidade do usuário selecioná-los manualmente.
- Implementação de outros métodos de visualização dos resultados, como grafos, por exemplo.
- Criação de uma área do usuário, em que este possa gerir os seus dados, assim como acessar os resultados de análises já feitas.
- Por fim, permitir a análise, não apenas de estruturas estáticas, mas também arquivos de trajetória, que contenham diversos estados da mesma estrutura em função do tempo, o que poderia tornar a análise das LHS mais fidedigna, principalmente com relação à interação da proteína com a água. Para isso, problemas de incompatibilidade entre versões diferentes do GROMACS teriam de ser solucionados.

## REFERÊNCIAS

- ABRAHAM, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, p. 19–25, 2015.
- ALDER, B. J.; WAINWRIGHT, T. E. Phase Transition for a Hard Sphere System. **The Journal of Chemical Physics**, v. 27, n. 5, p. 1208–1209, 1957.
- ARUNAN, E. *et al.* Defining the hydrogen bond: An account (IUPAC Technical Report). **Pure and Applied Chemistry**, v. 83, n. 8, p. 1619–1636, 2011.
- BAKER, E. N.; HUBBARD, R. E. Hydrogen bonding in globular proteins. **Progress In Biophysics And Molecular Biology**, v. 44, n. 2, p.97-179, 1984.
- BERENDSEN, H. J. C.; VAN DER SPOEL, D.; VAN DRUNEN, R. GROMACS: A message-passing parallel molecular dynamics implementation. **Computer Physics Communications**, v. 91, n. 1–3, p. 43–56, 1995.
- BERMAN, H. M. *et al.* 24.5. The Protein Data Bank, 1999–. **International Tables for Crystallography**, v. F, p. 675–684, 2006.
- BURLEY, S. K. *et al.* Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. **New Comprehensive Biochemistry**. v. 11, p.627–641, 2017.
- CORNELL, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. **Journal of the American Chemical Society**, v. 117, n. 19, p. 5179–5197, 1995.
- CUMMINGS, P. T.; GILMER, J. B. Open-source molecular modeling software in chemical engineering. **Current Opinion in Chemical Engineering**, v. 23, p. 99–105, 2019.
- DESIRAJU, Gautam R.; STEINER, Thomas. **The Weak Hydrogen Bond**. Oxford: Oxford University Press, 2001.
- DUAN, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. **Journal of Computational Chemistry**, v. 24, n. 16, p. 1999–2012, 2003.
- EBEJER, J. P. *et al.* The emerging role of cloud computing in molecular modelling. **Journal of Molecular Graphics and Modelling**, v. 44, p. 177–187, 2013.
- FERSHT, A. R. The hydrogen bond in molecular recognition. **Trends in Biochemical Sciences**, v. 12, n. C, p. 301–304, 1987.
- GARCIA, A. E.; SANBONMATSU, K. Y. -Helical stabilization by side chain shielding of backbone hydrogen bonds. **Proceedings of the National Academy of Sciences**, v. 99, n. 5, p. 2782–2787, 2002.

GURUSARAN, M. *et al.* Hydrogen Bonds Computing Server ( HBCS ): an online web server to compute hydrogen-bond interactions and their precision. **Journal of Applied Crystallography**, v. 49, n. 2, p. 642–645, 2016.

GUVENCH, O.; MACKERELL, A. D. Comparison of protein force fields for molecular dynamics simulations. **Methods in Molecular Biology**, v. 443, p. 63–88, 2008.

HASHEM, I. A. T. *et al.* The rise of “big data” on cloud computing: Review and open research issues. **Information Systems**, v. 47, p. 98–115, 2015.

HAYES, B. Cloud computing. **Communications of the ACM**, v. 51, n. 7, p. 9, 2008.

HORNAK, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. **Proteins: Structure, Function, and Bioinformatics**, v. 65, n. 3, p. 712–725, 2006.

KAMINSKI, G. A. *et al.* Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. **The Journal of Physical Chemistry B**, v. 105, n. 28, p. 6474–6487, 2001.

KYTE, J.; DOOLITTLE, R. F. A simple method for displaying the hydropathic character of a protein. **Journal of Molecular Biology**, v. 157, n. 1, p. 105–132, 1982.

KOLLMAN, P. A. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. **Accounts of Chemical Research**, v. 29, n. 10, p. 461–469, 1996.

LINDORFF-LARSEN, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. **Proteins: Structure, Function, and Bioinformatics**, v. 78, n. 8, p. NA-NA, 2010.

LOURENZONI, M. R. **Estudo, via simulação molecular, da interação de dois peptídeos da região 115-129 da miotoxina II do veneno da serpente Bothrops asper com membranas celulares.** 2005. Tese (Doutorado em Química) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2005.

MICHAUD-AGRAWAL, N. *et al.* MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. **Journal of Computational Chemistry**, v. 32, n. 10, p. 2319–2327, 2011.

MOON, C. P.; FLEMING, K. G. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 25, p. 10174–7, 2011.

OOSTENBRINK, C.; VILLA, A.; MARK, A. E.; VAN GUNSTEREN, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. **Journal of Computational**

**Chemistry**, v. 25, n. 13, p. 1656–1676, 2004.

PFEIFFER, P. et al. Zur Theorie der Farblacke, II. **Justus Liebig's Annalen der Chemie**, v. 398, n. 2, p. 137–196, 1913.

ROE, D. R.; CHEATHAM, T. E. PTraj and CPPtraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. **Journal of chemical theory and computation**, v. 9, n. 7, p. 3084–95, 2013.

ROSE, A. S. et al. NGL viewer: web-based molecular graphics for large complexes. **Bioinformatics**. v. 34. p. 3755–3758, 2018.

SCHMID, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. **European Biophysics Journal**, v. 40, n. 7, p. 843–856, 2011.

SCHULER, L. D.; DAURA, X.; VAN GUNSTEREN, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. **Journal of Computational Chemistry**, v. 22, n. 11, p. 1205–1218, 2001.

VAN DER SPOEL, D. et al. Gromacs: Fast, flexible, and free. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1701–1718, 2005.

WANG, J.; CIEPLAK, P.; KOLLMAN, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? **Journal of Computational Chemistry**, v. 21, n. 12, p. 1049–1074, 2000.

WERNER, A. Ueber Haupt- und Nebenvalenzen und die Constitution der Ammoniumverbindungen. **Justus Liebigs Annalen der Chemie**, v. 322, n. 3, p. 261–296, 1902.

WHITE, S. H. et al. Recognition of transmembrane helices by the endoplasmic reticulum translocon. **Nature**, v. 433, n. 7024, p. 377–81, 2005.

WIMLEY, W. C.; WHITE, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfacesne interfaces. **Nature Structural Biology**, v. 3, n. 10, p. 842–848, 1996.

ZHAO, G.; LONDON, E. An amino acid “transmembrane tendency” scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. **Protein Science**, v. 15, n. 8, p. 1987–2001, 2006.

## APÊNDICE A – CAMPOS DE FORÇA DISPONÍVEIS PARA USO COM O GROMACS 2016.1

- 1) AMBER03 protein, nucleic AMBER94 (DUAN et al, 2003);
- 2) AMBER94 force field (CORNELL et al., 1995);
- 3) AMBER96 protein, nucleic AMBER94 (KOLLMAN et al., 1996);
- 4) AMBER99 protein, nucleic AMBER94 (WANG et al., 2000);
- 5) AMBER99SB protein, nucleic AMBER94 (HORNAK et al., 2006);
- 6) AMBER99SB-ILDN protein, nucleic AMBER94 (LINDORFF-LARSEN et al., 2010);
- 7) AMBERGS force field (GARCIA; SANBONMATSU, 2002);
- 8) CHARMM27 all-atom force field (CHARM22 plus CMAP for proteins);
- 9) GROMOS96 43a1 force field;
- 10) GROMOS96 43a2 force field (improved alkane dihedrals);
- 11) GROMOS96 45a3 force field (CHULER, L. D.; DAURA, X.; VAN GUNSTEREN, W. F., 2001);
- 12) GROMOS96 53a5 force field (OOSTENBRINK et al., 2004);
- 13) GROMOS96 53a6 force field (OOSTENBRINK et al., 2004);
- 14) GROMOS96 54a7 force field (SCHMID et al., 2011);
- 15) OPLS-AA/L all-atom force field (KAMINSKI, G. A. et al., 2001);