

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO

DEPARTAMENTO DE SISTEMAS DE COMPUTAÇÃO

PROJETO DE PESQUISA FAPESP

**Predição de Estruturas de Proteínas com Algoritmos Paralelos Ortogonais às
Plataformas Paralelas**

LINHA DE FOMENTO: BOLSA NO PAÍS - REGULAR - INICIAÇÃO CIENTÍFICA

Candidato:

Orientador:

Matheus Henrique Junqueira Saldanha

Paulo Sérgio Lopes de Souza

26 de janeiro de 2019

Resumo

Resumo do projeto. Deve haver no máximo 20 linhas como especificado em <http://www.fapesp.br/248#10>.

Sumário

1	Introdução	1
2	Problema da Predição de Estruturas de Proteínas	1
3	Arquiteturas e Modelos de Programação Paralela	2
4	Trabalhos Relacionados	3
5	Objetivos	3
6	Metodologia	3
6.1	Atividades a Serem Desenvolvidas	3
6.2	Cronograma	4
6.3	Desafios Científicos e Técnicos	4
6.4	Material a Ser Utilizado	4
6.5	Forma de Análise dos Resultados	4
6.6	Resultados Esperados	5
6.7	Exequibilidade	5

1 Introdução

A FAPESP não estabelece normas rígidas quanto à formatação do projeto de iniciação científica. Ela especifica, no entanto, que deve haver no máximo 20 páginas escritas com espaçamento duplo. Para outras modalidades de bolsa, ela também expressa preferência por fonte *Times New Roman*, ou semelhantes, em tamanho 12 pontos¹.

2 Problema da Predição de Estruturas de Proteínas

Todas as informações dadas a seguir se referem ao que foi adotado no projeto submetido e aceito. Legendas de figuras ficam abaixo da figura (veja a Figura 1). Legendas de tabelas ficam acima (veja Tabela 1).

Figuras podem ser justapostas, e cada figura recebe sua própria legenda, como visto nas Figuras 1 e 2. No entanto, posicioná-las corretamente não é simples, então cuidado deve ser tomado. Em especial, atenção com os caracteres “%” posicionados no final de certas linhas no arquivo latex deste projeto, pois eles são importantes para o correto funcionamento do ambiente *minipage*.

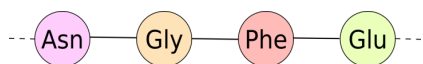


Figura 1: Uma sequência dos aminoácidos: asparagina, glicina, fenilalanina e ácido glutâmico.

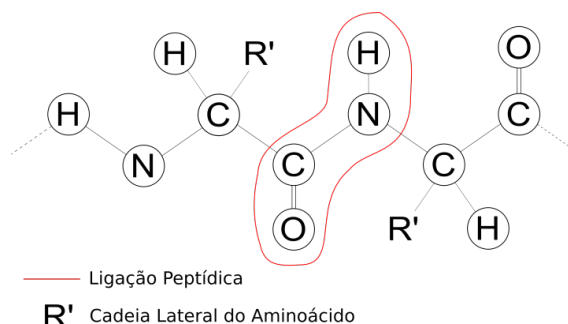


Figura 2: Dois aminoácidos ligados por uma ligação peptídica.

Figuras retiradas de fontes externas devem ter sua fonte citada na legenda da mesma; caso contrário, poderia configurar plágio. Na Figura 3, a fonte é citada e uma URL é fornecida via nota de rodapé.

Quando precisar de um pedaço qualquer de texto para preencher qualquer lugar temporariamente, pode-se usar o comando *lipsum*, que resulta no parágrafo a seguir.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac,

¹Veja, por exemplo, <http://www.fapesp.br/10408>

²<https://www.ebi.ac.uk/pdbe/entry/pdb/1a3n>

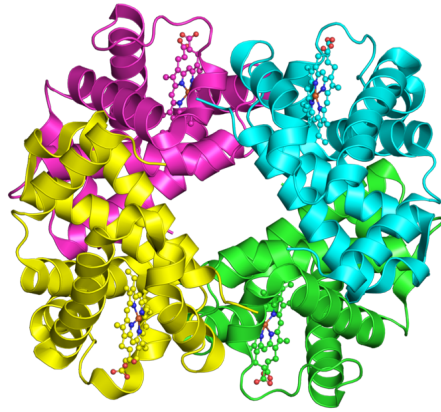


Figura 3: Estrutura tridimensional da hemoglobina, obtida de *Protein Data Bank in Europe* ².

adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Caso tenha algum parágrafo que precisa ser revisado futuramente, pode-se usar o comando *rev*. Por exemplo, **este texto precisa ser revisado**.

3 Arquiteturas e Modelos de Programação Paralela

Citações são essenciais para qualquer projeto de pesquisa. Por exemplo, no contexto de arquiteturas paralelas, uma boa referência é [Tanenbaum, 2016]. Para poder citar utilizando o comando *cite*, preencha o arquivo *bibliography.bib* com o devido registro no formato bibtex, que pode ser facilmente obtido em plataformas como o *Google Scholar*.

Ressalta-se que a FAPESP não estabelece normas rígidas de formatação do projeto. Existem muitos pacotes latex interessantes que podem ser utilizados para embelezar o documento. Verifique, por exemplo, o pacote *hyperref*, que pode ser utilizado para tornar links e referências internas (para figuras, tabelas, seções *etc*) “clicáveis” e coloridos. Neste modelo, limitou-se a utilizar pacotes que foram utilizados no projeto submetido e aceito.

4 Trabalhos Relacionados

5 Objetivos

6 Metodologia

A partir deste ponto, o conteúdo do projeto submetido e aceito é replicado, para ser tomado como base para a escrita da seção de metodologia.

6.1 Atividades a Serem Desenvolvidas

A execução do projeto pode ser dividida nas seguintes etapas:

1. **Seleção dos algoritmos**, quando algoritmos paralelos de predição de estruturas de proteínas, encontrados na literatura, serão selecionados, com suporte da engenharia reversa sobre os mesmos. Espera-se nesta etapa levantar não apenas quais algoritmos considerar para o desenvolvimento do projeto, mas também determinar seus principais requisitos com vistas a implementações paralelas em diferentes arquiteturas;
2. **Reengenharia dos algoritmos selecionados**, quando a metodologia de projeto PCAM será aplicada aos algoritmos selecionados, instanciando-os aos modelos de programação considerados no projeto;
3. **Implementação dos algoritmos**, na qual os algoritmos paralelos serão implementados e documentados, ao mesmo tempo que serão executados os testes de unidade durante a implementação para favorecer a confiabilidade do código final, diminuindo a possibilidade de defeitos não revelados nos códigos desenvolvidos;
4. **Projeto e execução de experimentos**, quando experimentos serão projetados e executados visando validar os algoritmos implementados e obter os dados necessários para avaliar o desempenho dos mesmos (vide Seção 6.5 para maiores detalhes);
5. **Finalização da documentação**, na qual será elaborada a documentação final do conjunto de algoritmos analisados e implementados, a ser disponibilizada publicamente para uso de outros

pesquisadores. A disponibilização da documentação (em conjunto com os algoritmos implementados) ocorrerá na página do Laboratório de Sistemas Distribuídos e Programação Concorrente (LaSDPC) do ICMC/USP;

6. **Redação de artigos científicos**, quando serão escritos artigos científicos descrevendo os resultados obtidos neste projeto de Iniciação Científica. Espera-se submeter artigos para o Simpósio em Sistemas Computacionais de Alto Desempenho (WSCAD) em 2018;
7. **Redação dos Relatórios Científicos**, quando serão escritos os relatórios exigidos pelas normas da FAPESP.

6.2 Cronograma

Com base nas tarefas enumeradas na Seção 6.1, é mostrado na Tabela 1 o cronograma a ser executado durante a realização deste projeto.

Tabela 1: Cronograma das atividades.

Fases	Meses											
	1	2	3	4	5	6	7	8	9	10	11	12
1. Seleção algoritmos	x	x										
2. Reengenharia algoritmos		x	x									
3. Implementação algoritmos		x	x	x	x	x	x					
4. Experimentos			x	x	x	x	x	x	x	x		
5. Documentação								x	x	x		
6. Artigos									x	x	x	
7. Relatórios					x	x					x	x

6.3 Desafios Científicos e Técnicos

6.4 Material a Ser Utilizado

6.5 Forma de Análise dos Resultados

1. eficiência e *speedup* em relação à versão sequencial do mesmo algoritmo, variando-se a carga de trabalho (proteínas a serem preditas) e o tamanho da arquitetura sendo usada para a execução paralela;

2. comparação da eficiência e *speedups* obtidos pelos diferentes algoritmos, também variando-se as proteínas e as arquiteturas;
 3. menores energias calculadas, para cada dado de entrada e arquitetura utilizada.
1. devem ser implementados entre 4 e 8 algoritmos paralelos de predição de PSP. Este número foi estipulado em função dos algoritmos já encontrados na literatura em artigos científicos. Caso novos artigos sejam encontrados, seus algoritmos também serão considerados. Outrossim, acredita-se que a implementação dos algoritmos já descobertos seja possível; porém, isso será analisado em maior profundidade, de fato, durante o desenvolvimento do projeto, abrindo-se a possibilidade para que nem todos os algoritmos venham a ser codificados (por falta de informações e/ou retorno de seus autores, caso necessário);
 2. os algoritmos serão, tanto quanto possível, projetados visando execução em diferentes arquiteturas paralelas e codificação com diferentes modelos de programação.

6.6 Resultados Esperados

6.7 Exequibilidade

Referências

[Tanenbaum, 2016] Tanenbaum, A. S. (2016). *Structured computer organization*. Pearson Education India.