

CARLOS ALBERTO MAZIERO

UM MODELO PARA DISSERTAÇÕES E TESES  
(ESCREVI UM TÍTULO MAIS LONGO PARA VER COMO SE COMPORTA A  
QUEBRA DE LINHAS E O ESPAÇAMENTO ENTRE ELAS)

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática no Programa de Pós-Graduação em Informática, setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Aurora Trinidad Ramirez Pozo.

Co-orientador: Roberto Santana .

CURITIBA PR  
2016

# Resumo

O resumo deve conter no máximo 500 palavras, devendo ser justificado na largura da página e escrito em um único parágrafo<sup>1</sup> com um afastamento de 1,27 cm na primeira linha. O espaçamento entre linhas deve ser de 1,5 linhas. O resumo deve ser informativo, ou seja, é a condensação do conteúdo e expõe finalidades, metodologia, resultados e conclusões.

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

**Palavras-chave:** palavra-chave 1, palavra-chave 2, palavra-chave 3.

---

<sup>1</sup>E também não deve ter notas de rodapé; em outras palavras, não siga este exemplo... ;-)

# Abstract

The abstract should be the English translation of the “resumo”, no more, no less.

**Keywords:** keyword 1, keyword 2, keyword 3.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Referencial Teórico</b>	<b>3</b>
2.1	PDP - Problema de Dobramento de Proteínas . . . . .	3
2.2	Dobramento de proteínas . . . . .	3
2.3	Modelos de Representação de Proteínas . . . . .	5
2.3.1	Modelos Discretos . . . . .	6
2.3.2	Considerações Finais . . . . .	9
2.4	AEs - Algoritmos Evolucionários . . . . .	9
2.4.1	NSGAI - Non-dominated sorting Genetic Algorithm II . . . . .	10
2.4.2	IBEA (Indicator-Based Evolutionary Algorithm) . . . . .	10
2.5	Exemplo de figura . . . . .	12
2.6	Exemplo de tabela . . . . .	12
2.7	Exemplo de equação . . . . .	13
2.8	Exemplos de código-fonte . . . . .	13
2.9	Conclusão . . . . .	14
	<b>Referências Bibliográficas</b>	<b>15</b>
<b>A</b>	<b>Exemplo de anexo</b>	<b>16</b>
A.1	Uma Seção . . . . .	16
A.1.1	Uma sub-Seção . . . . .	17

# Lista de Figuras

2.1	Comunicação inter-processos. . . . .	12
-----	--------------------------------------	----

# Lista de Tabelas

2.1	Os 16 modelos centrais do $U\text{CON}_{ABC}$ . . . . .	13
-----	---	----

# Lista de Acrônimos

DINF	Departamento de Informática
PPGINF	Programa de Pós-Graduação em Informática
UFPR	Universidade Federal do Paraná

# Lista de Símbolos

$\alpha$	alfa, primeira letra do alfabeto grego
$\beta$	beta, segunda letra do alfabeto grego
$\gamma$	gama, terceira letra do alfabeto grego
$\omega$	ômega, última letra do alfabeto grego
$\pi$	pi
$\tau$	Tempo de resposta do sistema
$\theta$	Ângulo de incidência do raio luminoso



# Capítulo 1

## Introdução

Este modelo foi proposto com o intuito de padronizar e simplificar as monografias, dissertações e teses produzidas no Departamento de Informática da UFPR. Ele foi vagamente inspirado nas normas da ABNT (conforme indicado em [UFPR, 2015]), mas não as segue *ipsis litteris*. Várias alterações foram feitas com o objetivo de melhorar sua estética e tornar o texto mais legível para trabalhos na área de informática. A versão atualizada deste modelo está disponível em [Maziero, 2015].

Este modelo está baseado em uma classe específica `ppginf.cls`, que aceita várias opções de compilação. A versão do documento pode ser:

- `defesa`: é gerado um documento em espaço 1,5, frente simples e sem as páginas iniciais adicionais; é uma versão adequada para receber as anotações dos membros da banca de defesa.
- `final`: é gerado um documento em espaço simples, frente/verso, com páginas iniciais (capa, ficha catalográfica, folha de aprovação, agradecimentos, etc). É uma versão bem mais compacta, mais ecológica e ideal para a impressão definitiva.

Para obter os melhores resultados, compile este modelo usando a seguinte sequência de passos:

```
pdflatex main          // compilação inicial
bibtex main            // processa referências bibliográficas
pdflatex main          // compilação final
```

ou

```
make                  // faz tudo...
```

Os principais itens considerados na formatação deste documento foram:

- Papel em formato A4, com margens de 20 mm à direita e embaixo, 30 mm nos demais lados. Não devem ser usados cabeçalhos ou rodapés além dos que estão aqui propostos.

- O texto principal do documento escrito em 12 pontos. O fonte principal do texto pode ser selecionado no arquivo `packages.tex`.
- Código-fonte, listagens e textos similares são formatados em fonte Courier 12 ou 10 pontos.
- O espaçamento padrão entre linhas é 1,5 linhas (1 linha na versão final). Não inserir espaços adicionais entre parágrafos normais. Figuras, tabelas, listagens e listas de itens devem ter um espaço adicional antes e após os mesmos.
- As páginas iniciais não são numeradas.
- O corpo do texto é numerado com algarismos arábicos (1, 2, 3, ...) a partir da introdução, até o final do documento. Os números de página devem estar situados no alto à direita (páginas direitas) ou à esquerda (páginas esquerdas).
- Expressões em inglês, grego, latim ou outras línguas devem ser enfatizadas em itálico, como *sui generis* ou *scheduling* (use o comando `\emph{...}`).
- Para reforçar algo, deve-se usar somente **negrito**. Sublinhado ou MAIÚSCULAS não devem ser usados como forma de ênfase!
- As notas de rodapé também têm um modelo<sup>1</sup>. Notas de rodapé servem para fazer algum comentário paralelo; não as use para colocar URLs, referências bibliográficas ou significado de siglas.

Felizmente o  $\text{\LaTeX}$  resolve a maior parte dessas questões!

---

<sup>1</sup>As notas de rodapé devem ser escritas em tamanho 10 pt, numeradas em arábico.

# Capítulo 2

## Referencial Teórico

### 2.1 PDP - Problema de Dobramento de Proteínas

Proteínas são estruturas básicas, essenciais para vida e possuem incontáveis funções biológicas. Proteínas são sintetizadas pelos ribossomos seguindo um formato provido pelo mensageiro RNA (mRNA). Durante a síntese, as proteínas dobras (enovelam) em uma estrutura tridimensional única, conhecida como conformação nativa. Este processo é chamado: dobramento de proteínas (*protein folding*). A função biológica de uma proteína depende da sua estrutura tridimensional.

As proteínas são polímeros compostos por sequências de aminoácidos (também chamados de resíduos) conectados linearmente por ligações peptídicas. Cada aminoácido é composto por um átomo central de carbono ( $C\alpha$ ) conectado a um átomo de hidrogênio, um grupo amina, um grupo carboxila e uma cadeia lateral (*side-chain*) a qual confere a cada aminoácido uma função distinta. Uma ligação peptídica é formada por dois aminoácidos quando o grupo carboxila de uma molécula reage com o grupo amina da outra. Este processo de agregação de aminoácidos é conhecido como desidratação pois libera uma molécula de água ( $H_2O$ ) [?]. Proteínas podem ser chamadas de cadeias polipeptídicas. Todos os aminoácidos tem o mesmo *backbone* e se diferem dos outros apenas pela *side-chain*, a qual pode ser um simples átomo de hidrogênio ou até um grupo heterocíclico complexo. A *side-chain* define as propriedades físicas e químicas dos aminoácidos de uma proteína [?].

### 2.2 Dobramento de proteínas

É o processo em que cada cadeia polipeptídica é transformada em uma estrutura compacta que realiza alguma função biológica [?]. Estas funções incluem controle e regulação de processos químicos essenciais para os organismos vivos [?]. A estrutura tridimensional mais estável é chamada de conformação nativa e é a qual permite que a proteína exerça corretamente sua função biológica [?, ?].

Experimentos conduzidos por Anfinsen et al. [?, ?, ?], mostraram que as proteínas possuem apenas uma conformação nativa e que as informações essenciais que codificam a estrutura estão contidas na sequência de aminoácidos. A conformação tridimensional nativa é dada pela estrutura primária (sequência de aminoácidos) de uma proteína.

Muitas proteínas podem desnaturar por modificações no ambiente em que estão inseridas, conforme demonstrado por [?, ?, ?]. Durante o processo de desnaturação as proteínas perdem sua forma nativa (desdobram) e, conseqüentemente, perdem sua função. O exemplo mais conhecido de desnaturação proteica é o da clara do ovo. A clara do ovo é composta por água e albumina. A albumina é uma proteína polar, portanto solúvel em água. Ao fritar ou cozinhar o ovo, eleva-se a temperatura, levando à desnaturação da albumina que, mesmo ao retornar à temperatura original, não consegue voltar à sua conformação nativa. Além de se desdobrarem é possível que ocorram erros de dobramento na formação das proteínas causando com que a proteína não exerça sua função biológica corretamente. Estudos tentam identificar causas para os erros de dobramento das proteínas pois muitas enfermidades são causadas pelo mal dobramento de proteínas como, por exemplo, mal de Alzheimer [?, ?], alguns tipos de câncer [?, ?, ?], fibrose cística [?], arteriosclerose [?], mal de Parkinson [?], entre outras. Portanto, entender como o processo de dobramento de proteínas ocorre é de fundamental importância. Um dos objetivos comuns das ciências biológicas é caracterizar funcionalmente sequências de proteínas através da resolução de suas conformações nativas [?]. Varias áreas da ciência, tais como Biologia, Medicina, Química Orgânica, realizam diferentes estudos das proteínas. Muitos destes estudos são voltados para o processo de dobramento das proteínas que pode sofrer alterações: tanto em como a conformação estará disposta no espaço, como ela estará agrupada e sobre sua má formação. Isto é muito relevante para estudos que visam à produção de medicamentos, suplementos alimentares, técnicas que manipulam o DNA, ou para formação de novos compostos proteicos sintéticos em laboratório [?]. É importante mencionar que apesar do avanço na grande quantidade de proteínas que se tem conhecimento por conta de projetos de sequenciamento genômico, apenas uma pequena fração de estruturas tridimensionais é conhecida.

A cristalografia de raios-X e espectroscopia de RNM são os métodos experimentais mais poderosos para o estudo da estruturas de proteínas [?] [?]. Entretanto estes métodos são altamente custosos tanto em esforços computacionais, de tempo e financeiros, e estão disponíveis apenas para algumas instituições.

Embora o conceito de dobramento de proteínas tenha surgido da área de biologia molecular, este problema é um problema interdisciplinar, o qual requer apoio de muitas áreas do conhecimento, e é considerado como um dos desafios atuais mais importantes da biologia e bioinformática [?].

Na biologia computacional existem dois problemas que tratam sobre o dobramento de proteínas. São eles: problema de predição estrutura de proteínas (ou PSP - *Protein Structure Prediction*), que trata de predizer a estrutura tridimensional (conformação) a partir de sua sequência (estrutura primária); e o problema de dobramento de proteínas (PDP ou PFP - *Protein*

*Folding Problem*), o qual trata da determinação dos passos/eventos que conduzem o dobramento a partir da estrutura primária até a conformação nativa [?]. Porém, na literatura, é encontrado ambos os termos sendo utilizados sem nenhuma distinção, normalmente se referindo apenas ao primeiro problema [?].

A ciência da computação desempenha um papel importante nisto, propondo e desenvolvendo modelos e soluções computacionais para o estudo de ambos os problemas PSP e PDP [?]. Muitas estratégias computacionais descrevem modelos de predição de estruturas de proteínas com diferentes níveis de detalhamento e complexidade mas que permitem uma representação fidedigna da estrutura tridimensional, sem perda de viabilidade computacional [?]. Dessa maneira evita-se a obrigatoriedade do uso de métodos caros, aumentando a competitividade e auxiliando na consolidação de centros de pesquisa que desenvolvem estudos nesta área [?].

Acredita-se que a conformação nativa de uma proteína é a sua estrutura mais estável, adquirindo um estado de energia livre mínima, o que gerou a chamada hipótese da termodinâmica [?]. Os modelos de predição de estruturas normalmente são baseados nas leis da termodinâmica onde o problema é modelado como um problema de minimização da energia livre a respeito das possíveis conformações que uma proteína pode assumir [?]. A minimização da energia livre é assumida como o principal fator para a formação da estrutura da proteína. Portanto, a conformação nativa de uma proteína é dada por aquela que possui o menor valor de energia livre.

Segundo [?], um modelo computacional deve possuir algumas características:

- Um conjunto de entidades que representam os átomos e as ligações entre eles.
- Regras que definem as possíveis conformações.
- Uma função que seja computacionalmente factível, para calcular a energia livre das possíveis conformações.

A próxima subseção irá discorrer sobre alguns modelos de representação de estruturas de proteínas.

## 2.3 Modelos de Representação de Proteínas

Em suma existem duas classes de modelos de representação de estruturas de proteínas: analítico (também conhecido como *all atom*) e discreto (chamado também de *coarse-grained*). Os modelos analíticos possuem uma descrição detalhada da estrutura tridimensional incluindo informações de todos os átomos que constituem uma proteína. Já os modelos discretos descrevem as proteínas com um nível bastante reduzido de detalhes. Os modelos discretos recentemente ganharam maior interesse, por conta de dois fatores: a simulação de modelos analíticos nem sempre é computacionalmente possível e modelos discretos possibilitam simulações biologicamente relevantes com melhor aproveitamento computacional [?]. Embora simulações utilizando

modelos discretos ainda não possam ser consideradas tão preditivas quanto simulações analíticas, avanços notáveis têm sido alcançados, referente ao uso de metodologias mais rigorosas e criação de algoritmos para melhor explorar o espaço de busca [?]. Esta proposta irá descrever apenas os modelos discretos pois visa a utilização do modelo hidrofóbico polar (HP) 2D.

### 2.3.1 Modelos Discretos

Os modelos computacionais mais simples são os conhecidos como modelos de grade (*lattice models*). Estes modelos consideram as estruturas de proteínas como um colar de esferas posicionado em uma grade. O grau de liberdade dos movimentos é restrito à estrutura da grade, que pode ser 2D (plano) ou 3D (espacial). Conformações válidas são aquelas que os aminoácidos adjacentes na sequência também são adjacentes na grade e cada aminoácido ocupe uma posição distinta na grade. Muitos modelos de grade têm sido propostos e aplicados ao PDP. Os modelos 2D-HP e 3D-HP são exemplos de modelos de grade.

#### Modelo Hidrofóbico-Polar HP

No modelo HP os aminoácidos são classificados em 2 tipos: Hidrofílicos (Polar) e Hidrofóbico. Consequentemente, uma proteína é representada por uma *string* de caracteres definida por um alfabeto binário  $\{H, P\}$ . Este modelo considera que as interações entre aminoácidos hidrofóbicos (H) representam a contribuição mais importante para a energia livre de uma proteína. Portanto existe uma relação inversamente proporcional: quanto maior for a quantidade de interações hidrofóbicas (H-H), menor será a energia livre de uma proteína. Uma interação hidrofóbica (também conhecida como contato hidrofóbico) é definida como um par de aminoácidos do tipo H-H que não sejam consecutivos na sequência mas sejam adjacentes na grade. Como dito anteriormente, uma conformação é dita válida quando nenhuma posição da grade é ocupada por mais que um aminoácido. Conformações inválidas possuem colisões entre os aminoácidos. Dada uma conformação válida para o modelo HP e  $n$  o número de interações hidrofóbicas, a energia da conformação pode ser facilmente calculada utilizando a equação 2.1:

$$E(c) = n \cdot (-1) \quad (2.1)$$

Quando uma proteína é dobrada na sua conformação nativa, os aminoácidos hidrofóbicos tendem a se agrupar no interior da estrutura, protegidos por aminoácidos polares posicionados no exterior. Dessa maneira, um núcleo hidrofóbico é formado em proteínas dobradas [?]. Embora simples, a estratégia computacional de buscar uma solução para o PDP utilizando modelo HP é considerada como um problema *NP*-completo [?, ?, ?]. O espaço de busca do modelo HP possui algumas características mencionadas na literatura [?, ?, ?, ?, ?] :

- Elevada degenerescência.

- Espaço de busca multimodal.
- Muitas regiões com conformações inválidas.

A figura ?? apresenta um exemplo para os modelos HP (2D e 3D). Os pontos pretos são aminoácidos do tipo H e os brancos são aminoácidos do tipo P. As linhas pontilhadas representam as interações hidrofóbicas.

Diversos trabalhos aplicam algoritmos evolucionários e bio-inspirados ao problema de dobramento de proteínas utilizando o modelo HP. Uma decisão comum a todos trabalhos que utilizam o modelo HP é a de como representar as variáveis de entrada. Na literatura é possível encontrar basicamente três representações [?, ?]:

- Coordenadas cartesianas: Este método representa a posição de cada aminoácido utilizando suas coordenadas espaciais (x,y) no plano cartesiano 2D ou (x,y,z) no plano cartesiano 3D. Geralmente, sua utilização não é adequada para algoritmos baseados em população, pois estruturas idênticas ou semelhantes podem ter coordenadas totalmente diferentes [?];
- Coordenadas internas: Nesta representação as conformações são representadas por conjuntos de movimentos que ditam como a estrutura final irá se parecer. Esta representação é a mais utilizada em abordagens com algoritmos evolucionários para o PDP [?]. Existem duas possibilidades de se representar conformações utilizando coordenadas internas:
  - Coordenadas absolutas: Este tipo de coordenada é baseado na orientação do eixo da grade onde a conformação esta contida. No caso de uma grade bidimensional os possíveis movimentos são: {N, S, L, O} ou norte,sul,leste e oeste. Já em uma grade 3D os possíveis movimentos são: {N, S, L, O, F, T} que correspondem aos mesmos movimentos no caso 2D porém com dois movimentos a mais: para frente e para trás.
  - Coordenadas relativas: Este tipo de representação define a posição de cada aminoácido da cadeia em relação ao movimento do seu predecessor. O conjunto de movimentos possíveis para a grade 2D é definido por {F, E, D}, que correspondem aos movimentos: frente (continuar no mesmo sentido do aminoácido anterior), à esquerda e à direita. Em um cubo 3D, os possíveis movimentos são {F, E, D, C, B, }, possuindo dois movimentos a mais: para cima e para baixo.
- Matriz de distâncias: descreve a conformação de uma proteína através de uma matriz quadrada que representa a distância entre os aminoácidos. Este tipo de representação é raramente utilizado na literatura [?].

## Outros Modelos

Além do modelo HP, outros modelos simples em grade são utilizados para representar a estrutura de proteínas em outros estudos encontrados na literatura. Por exemplo:

- Modelo PH (*Perturbed Homopolymer*): Proposto por Shakhnovich et al. [?], as reações entre aminoácidos hidrofóbicos não são levadas em consideração, mas as interações entre aminoácidos do mesmo tipo são favorecidas, ou seja, H-H e P-P, desfavorecendo ligações H-P [?].
- Modelo LPE (*Lattice Polymer Embedding*): Modelo proposto por Unger e Moult [?]. A modelagem é feita a partir de uma sequência de aminoácidos,  $A = a_1, \dots, a_n$  atrelada a uma grade cúbica. Cada aminoácido possui um coeficiente de afinidade, definido para cada par  $a_i, a_j (c(a_i, a_j))$ . O objetivo da função de energia é minimizar o produto dos coeficientes pela distância entre os aminoácidos [?].
- Modelo HP-TSSC (*Hydrophobic-Polar Tangent Spheres Side Chain Model*): este modelo proposto por Hart et al. [?] é baseado no modelo HP, porém não utiliza uma grade. Neste modelo a proteína é modelada via um grafo tridimensional, onde a cadeia lateral e o *backbone* de cada aminoácido são esferas de mesmo raio [?].
- Modelo CGE (*Charged Graph Embedding*): Modelo descrito por Ngo et al. [?]. Neste modelo, uma carga (*charge*) é atribuída a cada resíduo. Entretanto, as conformações permitidas não são realistas [?].
- Modelo HPNX: modelo proposto por Bornberg-Bauer [?]. Divide os 20 aminoácidos em 3 classes: hidrofóbicos (H), positivos (P), negativos (N) e neutros (X). Este modelo, assim como o modelo HP, utiliza uma grade. Interações entre aminoácidos hidrofóbicos (H-H) representam interações de atração e diminuem a energia da conformação em 4,0, as interações entre positivos (P-P) e negativos (N-N) representam interações de repulsão e aumentam a energia livre em 1,0 e as interações entre N e P decrescem a energia em 1,0. O objetivo também consiste em minimizar a energia livre. Da mesma maneira que o modelo HP, quanto mais interações hidrofóbicas melhor será o dobramento. Porém este modelo não desconsidera o valor das demais interações [?].
- Modelo HP-helicoidal (Helical-HP): este modelo proposto por Thomas e Dill [?] considera apenas uma grade bidimensional e inclui dois tipos de interação: interações não-locais através de energia de contatos hidrofóbicos e interações locais representadas por uma tendência à formação de  $\alpha$ -hélices (chamada de propensão hélica) [?].
- Modelo *off-lattice* AB: este modelo proposto por Stillinger et al. [?] divide os aminoácidos em duas classes de acordo com sua polaridade: Hidrofóbicos (A) e Hidrofílicos (ou polares B). Inicialmente, este modelo foi aplicado em duas dimensões (2D AB *off-lattice*) e posteriormente aplicado para três dimensões (3D AB *off-lattice*). Os aminoácidos não consecutivos interagem através de um potencial modificado de Lennard-Jones. Os ângulos de torção entre ligações sucessivas também contribuem no cálculo da função de energia [?].



### 2.3.2 Considerações Finais

Neste seção o problema de dobramento de proteínas foi apresentado e sua importância para biológica computacional, química orgânica e medicina. Também foi mencionado que existem diversos modelos para representar estruturas de proteínas. Cada modelo tem suas peculiaridades e considera interações diferentes. Não existe um modelo que represente de maneira real o dobramento de proteínas, pois se trata de um processo ainda não completamente compreendido pelos cientistas e pesquisadores. Os modelos propostos tem diferentes níveis de detalhe e complexidade. O modelo mais simples é o HP mas apesar da sua simplicidade se apresenta como um problema *NP*-completo. O modelo HP será utilizado nesta proposta por conta de sua simplicidade de implementação, assim como o baixo custo computacional para realizar simulações do cálculo de energia. Utilizando o modelo HP diferentes maneiras de representar as variáveis existem. Nesta dissertação será utilizada a representação relativa pois é mencionado na literatura que esta tem uma maior capacidade de guiar algoritmos de busca a melhores resultados.

## 2.4 AEs - Algoritmos Evolucionários

Um algoritmo evolucionário (AE) é uma técnica de busca, altamente paralela, inspirada na teoria da seleção natural e reprodução genética de Charles Darwin. De acordo com a teoria de Darwin, a seleção natural irá favorecer os indivíduos que forem mais aptos, dessa maneira, estes indivíduos tem uma maior probabilidade de reprodução. Indivíduos com mais descendentes tem uma chance maior de perpetuarem seus códigos genéticos nas gerações futuras. O código genético é a identidade de cada indivíduo e é representado por cromossomos. Estes princípios são utilizados na implementação de algoritmos computacionais, que procuram por soluções melhores para um dado problema, evoluindo uma população de soluções codificadas em cromossomos artificiais – estruturas de dados utilizadas para representar soluções factíveis para um dado problema [?].

De maneira geral, problemas reais de otimização possuem múltiplos objetivos a serem minimizados/maximizados e estão presentes em muitas áreas do conhecimento. Para otimizar problemas multiobjetivos, dois ou mais objetivos são considerados os quais geralmente são conflitantes. Para estes problemas é impossível encontrar uma única solução ótima. Um conjunto de soluções é encontrado avaliando a dominância de Pareto [?] entre as soluções. O objetivo principal é encontrar o conjunto de soluções que sejam não dominadas entre si. Uma solução domina outra, se e somente se, for melhor em pelos um dos objetivos, sem ser pior em qualquer outro. Este conjunto de soluções constitui a fronteira de Pareto. Encontrar a fronteira real de Pareto é um problema NP-Completo [?], dessa maneira, o objetivo é encontrar uma boa aproximação da fronteira real.

Algoritmos Evolucionários Multi-Objetivos (AEMOs) são extensões de AEs para problemas multi-objetivos os quais aplicam conceitos da dominância de Pareto para criar diferentes estratégias para evoluir e manter a diversidade das soluções. Nesta dissertação foram explorados dois AEMOs: NSGAII [?] and IBEA [?].

### 2.4.1 NSGAII - Non-dominated sorting Genetic Algorithm II

O algoritmo 1 apresenta o pseudo código do NSGAII. O algoritmo recebe como parâmetro  $N$  o tamanho da população e  $T$  o número máximo de avaliações. O algoritmo inicia criando uma população com tamanho  $N$  chamada  $P_0$ . A população  $P_0$  é classificada de acordo com aptidão e a relação de não dominância. A população  $P_0$  é submetida ao operador de seleção: torneio binário para selecionar duas soluções que serão utilizadas para gerar descendentes. Operadores de cruzamento e mutação são aplicados na soluções selecionadas gerando duas soluções distintas descendentes. Ao fim do processo de reprodução as soluções descendentes são avaliadas e adicionadas a população chamada  $Q_0$ .

Após esta etapa,  $P_0$  e  $Q_0$  são adicionadas em uma população auxiliar chamada  $R$ . Utilizando o conceito de não dominância,  $R$  é ordenada criando fronteiras, onde cada solução da primeira fronteira não é dominada por nenhuma outra solução, já soluções da segunda fronteira são dominadas apenas por soluções contidas na primeira fronteira, e assim por diante. Para cada fronteira, as soluções são avaliadas utilizando um mecanismo de *Crowding-Distance* as soluções com maiores valores irão ser adicionadas na população da próxima geração chamada  $P_t$  onde  $t$  é a avaliação corrente.

Após criar e preencher  $P_t$  com as soluções não dominadas de todas as fronteiras, a população  $P_t$  é avaliada e então passa para um novo torneio binário e reprodução, dessa maneira, iniciando uma novo ciclo do algoritmo.

### 2.4.2 IBEA (Indicator-Based Evolutionary Algorithm)

No contexto de otimização multiobjetiva, otimizar consiste em tentar encontrar a fronteira com uma boa aproximação da fronteira real de Pareto. Entretanto, não existe uma definição geral para "uma boa aproximação". Consequentemente, indicadores de qualidade vem sendo utilizados para avaliar a qualidade da aproximação de fronteiras. O indicador *hypervolume* é um exemplo de indicador utilizado para avaliação e comparação das fronteiras.

No algoritmo IBEA, indicadores de qualidade são utilizados para avaliar o conjunto de soluções não dominadas [?]. Para utilizar o IBEA, é necessário definir qual indicador será utilizado para associar cada solução a um valor scalar. Um dos indicadores bastante utilizados

**Algoritmo 1 NSGAII**


---

```

1:  $N \leftarrow \text{Population Size}$ 
2:  $T \leftarrow \text{Max evaluations}$ 
3:  $P_0 \leftarrow \text{CreatePopulation}(N)$ ;
4:  $\text{CalculateFitness}(P_0)$ ;
5:  $\text{FastNonDominatedSort}(P_0)$ ;
6:  $Q_0 \leftarrow 0$ 
7: while  $Q_0 < N$  do
8:    $\text{Parents} \leftarrow \text{BinaryTournament}(P_0)$ ;
9:    $\text{Offspring} \leftarrow \text{CrossoverMutation}(\text{Parents})$ ;
10:   $Q_0 \leftarrow \text{Offspring}$ 
11: end while
12:  $\text{CalculateFitness}(Q_0)$ ;
13:  $t \leftarrow 0$ 
14: while  $t < T$  do
15:   $R_t \leftarrow P_t \cup Q_t$ ;
16:   $\text{Fronts} \leftarrow \text{FastNonDominatedSort}(R_t)$ ;
17:   $P_{t+1} \leftarrow 0$ 
18:   $i \leftarrow 0$ 
19:  while  $P_{t+1} + \text{Front}_i < N$  do
20:     $\text{CrowdingDistanceAssignment}(\text{Front}_i)$ ;
21:     $P_{t+1} \leftarrow P_{t+1} \cup \text{Front}_i$ 
22:     $i \leftarrow i + 1$ 
23:  end while
24:   $\text{CrowdingDistanceSort}(\text{Front}_i)$ ;
25:   $P_{t+1} \leftarrow P_{t+1} \cup \text{Front}_i[1 : (N - P_{t+1})]$ 
26:   $\text{Parents} \leftarrow \text{BinaryTournament}(P_{t+1})$ ;
27:   $Q_{t+1} \leftarrow \text{CrossoverMutation}(\text{Parents})$ ;
28:   $t \leftarrow t + 1$ 
29: end while
30: return  $P \leftarrow \text{Set of non-dominated solutions.}$ 

```

---

em conjunto com IBEA é o *hypervolume* por conta da sua capacidade de avaliar a convergência e a diversidade do espaço de busca ao mesmo [?].

$$F(x_i) = \sum_{x_j \in (P - x_i)} -e^{\frac{-I_{Hy}(x_j, x_i)}{k}} \quad (2.2)$$

A equação de *fitness* do IBEA é apresentada pela equação 2.2 e é utilizada para calcular a contribuição de uma dada solução para o valor do indicador referente a população, onde  $k$  é um fator escalar dependente do  $I_{Hy}$ , o qual representa o indicador de qualidade sendo utilizado. O valor  $F(x_i)$  corresponde à perda de qualidade da aproximação, da fronteira real de Pareto, se a solução  $x_i$  for removida da população [?].

## 2.5 Exemplo de figura

A forma sugerida para incluir figuras em um documento  $\text{\LaTeX}$  é importá-las usando o pacote `graphicx`. Como formatos gráficos sugere-se:

- Formatos *raster*, como PNG (*Portable Network Graphics*) ou JPG (*Joint Photographic Experts Group*) para fotografias; procure usar uma resolução de ao menos 150 dpi (*dots per inch*).
- Formatos vetoriais, como PDF (*Portable Document Format*) ou EPS (*Extended PostScript*) para diagramas e gráficos<sup>1</sup>.

A maior parte das ferramentas permite exportar figuras nesses formatos (a figura do exemplo foi produzida com o *Inkscape*, um programa livre multiplataforma). A figura 2.1 mostra um exemplo de inclusão de figura em PDF.

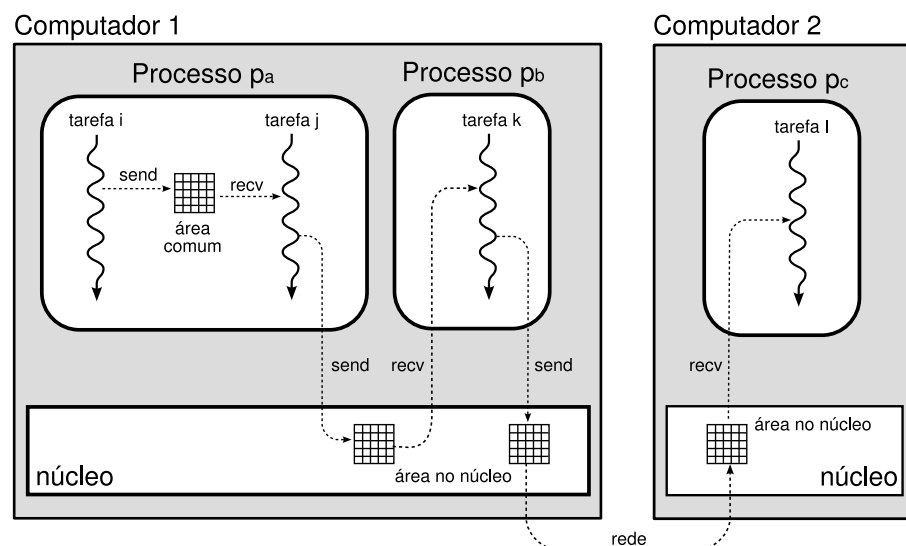


Figura 2.1: Comunicação inter-processos.

Para mais informações consulte [Goossens et al., 1993].

## 2.6 Exemplo de tabela

Tabelas são elementos importantes de um documento. No  $\text{\LaTeX}$  as tabelas podem ser objetos flutuantes (definidas no ambiente `table` e referenciadas por números usando `label` e `ref`) ou objetos fixos simples, criados pelo ambiente `tabular`. A tabela 2.1 é um exemplo de tabela flutuante, cuja posição no texto pode variar em função das quebras de página.

<sup>1</sup>NUNCA use JPG ou GIF para desenhos vetoriais, pois o resultado final geralmente fica borrado.

Tabela 2.1: Os 16 modelos centrais do UCON<sub>ABC</sub>

	0 (imutável)	1 ( <i>pre-update</i> )	2 ( <i>on-update</i> )	3 ( <i>pos-update</i> )
preA	•	•	—	•
onA	•	•	•	•
preB	•	•	—	•
onB	•	•	•	•
preC	•	—	—	—
onC	•	—	—	—

## 2.7 Exemplo de equação

Equações destacadas devem ser numeradas como mostra a equação 2.3:

$$E = m \times c^2 \quad (2.3)$$

## 2.8 Exemplos de código-fonte

Códigos-fonte podem ser produzidos de forma simples através do ambiente `verbatim`, como mostra este exemplo:

```
#include <stdio.h>

int main (int argc, char *argv[])
{
    int i ;                                // uma variavel local

    for (i=0; i< 100; i++)                 // um laço for
        printf ("i vale %d\n", i) ;       // uma saída na tela
}
```

No entanto, é preferível usar pacotes especializados para a edição ou inclusão de códigos-fonte, como o pacote `listings`. Eis um exemplo de código-fonte escrito com esse pacote:

```
1 #include <stdio.h>
2
3 int main (int argc, char *argv[])
4 {
5     int i ;                                // uma variável local
6
7     for (i=0; i< 100; i++)                 // um laço for
8         printf ("i vale %d\n", i) ; // uma saída na tela
9 }
```

Esse pacote também permite incluir códigos-fonte de arquivos externos. Eis um exemplo:

```
1 #include <stdio.h>
2
3 int main (int argc, char *argv[])
4 {
5     int i ;                // uma variável local
6
7     for (i=0; i< 100; i++)    // um laço for
8         printf ("i vale %d\n", i) ; // uma saída na tela
9 }
```

## 2.9 Conclusão

Todo capítulo (com exceção da introdução e da conclusão) deve encerrar com uma pequena conclusão local, resumindo os tópicos apresentados no capítulo e preparando o leitor para o próximo capítulo (exceto se esse for a conclusão geral). Caso o capítulo tenha apresentado resultados obtidos pelo próprio autor, estes devem ser sucintamente lembrados aqui.

# Referências Bibliográficas

[Goossens et al., 1993] Goossens, M., Mittelbach, F. e Samarin, A. (1993). *The L<sup>A</sup>T<sub>E</sub>X Companion*. Addison-Wesley.

[Maziero, 2015] Maziero, C. (2015). Modelo PPGInf UFPR para teses e dissertações. <http://www.inf.ufpr.br/maziero>. Acessado em 30/11/2015.

[Oetiker et al., 2007] Oetiker, T., Partl, H., Hyna, I. e Schlegl, E. (2007). *The Not So Short Introduction to L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>*. <http://www.ctan.org/tex-archive/info/lshort>.

[UFPR, 2015] UFPR, B. (2015). Manual de normalização de documentos científicos de acordo com as normas da ABNT. Relatório Técnico ISBN 9788584800025, Sistema de Bibliotecas – Universidade Federal do Paraná, Curitiba PR.

# Apêndice A

## Exemplo de anexo

Os apêndices são uma extensão do texto, destacados deste para evitar descontinuidade na sequência lógica ou alongamento excessivo de determinado assunto ou tópico secundário dentro dos capítulos da dissertação ou da tese. São contribuições que servem para esclarecer, complementar, provar ou confirmar as ideias apresentadas no texto dos capítulos e que são importantes para a compreensão dos mesmos.

Todos os apêndices devem vir após as referências bibliográficas e devem ser enumerados por letras maiúsculas (A, B, C, ...).

### A.1 Uma Seção

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam



et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt congue.

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

### A.1.1 Uma sub-Seção

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Phasellus id magna. Duis malesuada interdum arcu. Integer metus. Morbi pulvinar pellentesque mi. Suspendisse sed est eu magna molestie egestas. Quisque mi lorem, pulvinar eget, egestas quis, luctus at, ante. Proin auctor vehicula purus. Fusce ac nisl aliquam ante hendrerit pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi wisi. Etiam arcu mauris, facilisis sed, eleifend non, nonummy ut, pede. Cras ut lacus tempor metus mollis placerat. Vivamus eu tortor vel metus interdum malesuada.

Sed eleifend, eros sit amet faucibus elementum, urna sapien consectetur mauris, quis egestas leo justo non risus. Morbi non felis ac libero vulputate fringilla. Mauris libero eros, lacinia non, sodales quis, dapibus porttitor, pede. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi dapibus mauris condimentum nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Etiam sit amet

erat. Nulla varius. Etiam tincidunt dui vitae turpis. Donec leo. Morbi vulputate convallis est. Integer aliquet. Pellentesque aliquet sodales urna.