

**Dissertação de Mestrado**  
apresentada à UTFPR

por

**Carolina Paula de Almeida**

---

**Aplicação de Sistemas Imunológicos Artificiais para a Predição da  
Estrutura de Proteínas**

---

Banca Examinadora:

Presidente e Orientadora:

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Myriam Regattieri Delgado

UTFPR

Examinadores:

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Elizabeth Ferreira Gouvêa Goldberg

UFRN

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Angelita Maria De Ré

UNICENTRO

Prof. Dr. Leandro dos Santos Coelho

PUC-PR

Curitiba, fevereiro de 2007.

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Carolina Paula de Almeida

**Aplicação de Sistemas Imunológicos Artificiais para a  
Predição da Estrutura de Proteínas**

**Orientador:**

**Prof<sup>a</sup>. Dr<sup>a</sup>. Myriam Regattieri Delgado**

*Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial (CPGEI) da Universidade Tecnológica Federal do Paraná (UTFPR), na área de concentração de Informática Industrial, como requisito parcial à obtenção do grau de Mestre em Ciências.*

Curitiba

2007

Ficha catalográfica elaborada pela Biblioteca da UTFPR – Campus Curitiba

Almeida, Carolina Paula de

Aplicação de sistemas imunológicos artificiais para a predição da estrutura de proteínas / Carolina Paula de Almeida. Curitiba. UTFPR, 2007

XVIII, 116 f. : il. ; 30 cm

Orientadora Prof.<sup>a</sup> Dr.<sup>a</sup>. Myriam Regattieri Delgado

Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Curso de Pós-Graduação em Engenharia Elétrica e Informática Industrial. Curitiba, 2007

Bibliografia F. 104 – 110

1. Sistemas difusos. 2. Sistema imunológico – Inteligência artificial. 3. Engenharia biomédica. I. Delgado, Myriam Regattieri, orient. II. Universidade Tecnológica Federal do Paraná. Curso de Pós-Graduação em Engenharia Elétrica e Informática Industrial. III. Título

CDD: 511.313

Dedico esta dissertação à minha  
mãe e à minha avó. Tenho a certeza de que mesmo  
distantes estamos sempre unidas pelo amor.

# Agradecimentos

A Deus por dar-me tantas oportunidades de aprender a viver e principalmente amar.

À minha amiga e orientadora Myriam, pela confiança e dedicação depositada em mim e neste trabalho, pela qual tenho muita estima e admiração.

Não tenho palavras para agradecer ao meu amigo, colega de trabalho e namorado Richard, sem sua paciência, seu companheirismo e seu amor este trabalho não teria chagado ao fim.

À minha mãe e minha avó, porque não consigo avançar sem me lembrar que devo a elas muito do que sou, que sou parte delas e que elas são parte de mim.

Às minhas grandes amigas Nátalli, Núbbya e Fátima por me ouvirem tantas e tantas vezes, por confiarem em mim, pela nossa amizade que supera as circunstâncias.

À CAPES pelo período em que foi concedida a bolsa.

---

# *Lista de Algoritmos*

---

---

1	Forma básica do algoritmo Clonalg . . . . .	p. 27
2	Forma básica do algoritmo opt-Ainet . . . . .	p. 29
3	Forma básica do algoritmo copt-Ainet . . . . .	p. 31
4	Busca Tabu . . . . .	p. 49
5	Algoritmo Clonalg adotado . . . . .	p. 53
6	Hipermutação . . . . .	p. 57
7	Hipermacromutação . . . . .	p. 57
8	Maturação Fraca da Afinidade . . . . .	p. 59
9	Maturação Intensiva da Afinidade . . . . .	p. 62
10	Algoritmo de Rede Imunológica adotado . . . . .	p. 68

# *Sumário*

---

---

---

Lista de Figuras	p. x
Lista de Tabelas	p. xiii
Lista de Algoritmos	p. xv
Lista de Abreviaturas e Siglas	p. xvi
Resumo	p. xvii
Abstract	p. xviii
<b>1 Introdução</b>	<b>p. 1</b>
1.1 Motivações . . . . .	p. 2
1.2 Objetivos . . . . .	p. 3
1.3 Organização da Dissertação . . . . .	p. 4
<b>2 Predição da Estrutura de Proteínas</b>	<b>p. 5</b>

<b>2.1</b>	<b>Modelo Hidrofóbico-Polar Tridimensional . . . . .</b>	<b>p. 11</b>
<b>2.2</b>	<b>Abordagens Relacionadas para a Predição da Estrutura de Proteínas no Modelo HP . . . . .</b>	<b>p. 15</b>
<b>3</b>	<b>Sistema Imunológico Artificial . . . . .</b>	<b>p. 17</b>
<b>3.1</b>	<b>Princípio da Seleção Clonal . . . . .</b>	<b>p. 19</b>
<b>3.2</b>	<b>Teoria da Rede Imune . . . . .</b>	<b>p. 22</b>
<b>3.3</b>	<b>Sistema Imunológico Artificial . . . . .</b>	<b>p. 24</b>
	3.3.1 Espaço de Formas . . . . .	p. 24
	3.3.2 Bibliotecas Gênicas . . . . .	p. 25
	3.3.3 Algoritmos . . . . .	p. 26
	Clonalg . . . . .	p. 26
	Opt-Ainet . . . . .	p. 28
	Copt-Ainet . . . . .	p. 30
<b>3.4</b>	<b>Aplicações . . . . .</b>	<b>p. 32</b>
<b>4</b>	<b>Técnicas Auxiliares . . . . .</b>	<b>p. 35</b>
<b>4.1</b>	<b>Sistema de Inferência Fuzzy . . . . .</b>	<b>p. 36</b>
	4.1.1 Introdução aos Conjuntos Fuzzy . . . . .	p. 36
	Funções de Pertinência . . . . .	p. 37
	4.1.2 Base de Dados . . . . .	p. 40
	4.1.3 Base de Regras . . . . .	p. 41
	4.1.4 Mecanismos de Raciocínio . . . . .	p. 42
	Inferência Escalonada . . . . .	p. 43

Inferência Escalonada no Modelo Mamdani . . . . .	p. 45
<b>4.2 Busca Tabu . . . . .</b>	<b>p. 47</b>
4.2.1 Vizinhaça . . . . .	p. 47
4.2.2 Memória Adaptativa . . . . .	p. 48
4.2.3 Critérios de Aspiração . . . . .	p. 49
4.2.4 Algoritmo da Busca Tabu . . . . .	p. 49
<b>5 Metodologia Proposta . . . . .</b>	<b>p. 51</b>
<b>5.1 Clonalgs . . . . .</b>	<b>p. 53</b>
5.1.1 Descrição Geral do Clonalg . . . . .	p. 53
5.1.2 Operadores de Mutação . . . . .	p. 56
Hipermutação . . . . .	p. 56
Hipermacromutação . . . . .	p. 57
Primeira Mutação Construtiva . . . . .	p. 58
5.1.3 Estágios de Maturação da Afinidade . . . . .	p. 58
Maturação Fraca da Afinidade . . . . .	p. 58
Maturação Intensiva da Afinidade . . . . .	p. 60
5.1.4 Operadores de <i>Aging</i> . . . . .	p. 62
Operador de <i>Aging</i> Puro . . . . .	p. 63
Operador de <i>Aging Fuzzy</i> . . . . .	p. 64
5.1.5 Diferenças com Relação à Literatura . . . . .	p. 67
<b>5.2 Redes Imunes . . . . .</b>	<b>p. 68</b>

5.2.1 Recombinação . . . . .	p. 70
5.2.2 Diferenças com Relação à Literatura . . . . .	p. 71
<b>6 Simulações e Resultados</b>	p. 72
<b>6.1 Modelos Implementados</b> . . . . .	p. 73
6.1.1 Clonalgs . . . . .	p. 73
6.1.2 Redes Imunes . . . . .	p. 75
<b>6.2 Ajuste dos Parâmetros</b> . . . . .	p. 78
6.2.1 Ajuste dos Parâmetros dos Modelos Baseados no Clonalg . . . . .	p. 78
6.2.2 Ajuste dos Parâmetros dos Modelos Baseados na Teoria da Rede Imune . . . . .	p. 81
<b>6.3 Resultados dos Modelos Baseados no Clonalg</b> . . . . .	p. 82
6.3.1 Comparação Entre os Três Melhores Modelos de Clonalg . . . . .	p. 87
<b>6.4 Resultados dos Modelos Baseados na Rede Imune</b> . . . . .	p. 89
<b>6.5 Clonalg <i>versus</i> Rede Imune</b> . . . . .	p. 94
<b>6.6 Comparação Entre o Modelo Proposto e Trabalhos da Literatura</b> . . . . .	p. 95
<b>7 Conclusões e Trabalhos Futuros</b>	p. 97
7.1 Conclusões . . . . .	p. 97
7.2 Trabalhos Futuros . . . . .	p. 99
<b>Glossário</b>	p. 101
<b>Referências Bibliográficas</b>	p. 104

**Apêndice A – Aminoácidos**

p. 111

**Apêndice B – Estruturas das Proteínas**

p. 112

---

## *Lista de Figuras*

---

---

2.1	Estrutura básica de um aminoácido . . . . .	p. 6
2.2	Tipos de ligações não-covalentes . . . . .	p. 6
2.3	Forças hidrofóbicas atuando sobre uma proteína . . . . .	p. 7
2.4	Estrutura primária de uma proteína . . . . .	p. 8
2.5	Padrões estruturais que compõem a estrutura secundária das proteínas . . . . .	p. 8
2.6	Estrutura terciária de uma proteína . . . . .	p. 9
2.7	Estrutura quaternária de uma proteína . . . . .	p. 9
2.8	Representação de dobramento de proteínas em lattices . . . . .	p. 12
2.9	Representação de dobramento de proteínas em modelos livres . . . . .	p. 12
2.10	Representação de dobramento de proteína no modelo 3D HP . . . . .	p. 14
2.11	Tipos de representação do caminho no <i>lattice</i> . . . . .	p. 15
3.1	Princípio da seleção clonal . . . . .	p. 20
3.2	Memória imunológica . . . . .	p. 21
3.3	Maturação da afinidade . . . . .	p. 21

3.4	Princípio da rede imune . . . . .	p. 23
3.5	Reconhecimento por complementaridade . . . . .	p. 24
3.6	Processo de construção de um anticorpo utilizando biblioteca gênica	p. 26
4.1	Funções de pertinência em conjuntos clássicos e <i>fuzzy</i> . . . . .	p. 38
4.2	Exemplos de formatos das funções de pertinência . . . . .	p. 39
4.3	Variável lingüística . . . . .	p. 40
4.4	Inferência no modelo Mamdani . . . . .	p. 46
5.1	Exemplo de estrutura predita pelos modelos propostos. . . . .	p. 52
5.2	Exemplo de vizinhança para uma proteína com 6 resíduos. . . . .	p. 61
5.3	Exemplo de uma lista tabu de tamanho 7. . . . .	p. 61
5.4	Funções de pertinência . . . . .	p. 65
5.5	Geração de um anticorpo de tamanho 10 através de recombinação.	p. 70
6.1	Gráfico representando a diferença percentual da média dos modelos baseados no Clonalg e do ótimo conhecido na literatura. . . . .	p. 85
6.2	Gráfico do tempo que cada modelo baseado no Clonalg leva para uma execução. . . . .	p. 85
6.3	Comportamento dos três melhores modelos baseados no Clonalg: ClonalgII <sub>inf</sub> , ClonalgIII <sub>inf</sub> e ClonalgIII <sub>inf-BT</sub> . . . . .	p. 88
6.4	Gráfico representando a diferença percentual da média dos modelos baseados na teoria da Rede Imune e do ótimo conhecido na literatura	p. 92
6.5	Gráfico do tempo que cada modelo baseado na teoria da Rede Imune leva para uma execução . . . . .	p. 92
6.6	Relação entre as melhores afinidades encontradas pelos algoritmos Bactacking-EA e Aging-SIA e os piores valores de afinidade encontrados pelo ClonalgIII <sub>inf-BT</sub> . . . . .	p. 96

B.1	Estrutura da proteína de tamanho 20. . . . .	p. 112
B.2	Estrutura da proteína de tamanho 24. . . . .	p. 113
B.3	Estrutura da proteína de tamanho 25. . . . .	p. 113
B.4	Estrutura da proteína de tamanho 36. . . . .	p. 114
B.5	Estrutura da proteína de tamanho 48 . . . . .	p. 114
B.6	Estrutura da proteína de tamanho 50 . . . . .	p. 115
B.7	Estrutura da proteína de tamanho 60 . . . . .	p. 116

# *Lista de Tabelas*

---

---

---

5.1	Regras fuzzy utilizadas no sistema de inferência fuzzy . . . . .	p. 66
5.2	Diferenças entre os modelos de Clonalg . . . . .	p. 67
5.3	Diferenças entre os modelos de Rede . . . . .	p. 71
6.1	Seqüências do <i>benchmark</i> para o modelo 3D HP utilizadas neste trabalho com os melhores valores de energia conhecidos - $E^*$ . . .	p. 73
6.2	Características dos algoritmos propostos baseados no Clonalg. . .	p. 75
6.3	Características dos algoritmos propostos baseados na rede imune.	p. 78
6.4	Comparação dos resultados obtidos pelos modelos baseados no algoritmo Clonalg . . . . .	p. 83
6.5	Hipóteses de que as distribuições sejam diferentes (h) com seus respectivos p-valores obtidos nas simulações dos modelos de Clonalg	p. 84
6.6	Comparação entre os modelos de rede implementados . . . . .	p. 90
6.7	Hipóteses de que as distribuições sejam diferentes (h) com seus respectivos p-valores obtidos nas simulações dos modelos de Rede Imune . . . . .	p. 91
6.8	Comparação entre ClonalgIII <sub>inf-BT</sub> e a RedeI <sub>inf-BT</sub> . . . . .	p. 94

6.9	Comparação entre $\text{ClonalgIII}_{inf-BT}$ , Backtracking-EA e Aging-SIA	p. 95
A.1	Tabela de aminoácidos. . . . .	p. 111





# Lista de Abreviaturas e Siglas

PPEP	Problema da Predição da Estrutura de Proteínas	p. 1
SIA	Sistema Imunológico Artificial	p. 2
HP	Hidrofóbico-Polar	p. 3
NMR	<i>Nuclear Magnetic Resonance</i>	p. 10
SI	Sistema Imunológico Biológico	p. 17
Clonalg	<i>Clonal Selection Algorithm</i>	p. 26
Opt-ainet	<i>Artificial Immune Network for Optimization</i>	p. 28
Copt-ainet	<i>Artificial Immune Network for Combinatorial Optimization</i>	p. 30
SIF	Sistema de Inferência <i>Fuzzy</i>	p. 35
BT	Busca Tabu	p. 35
J2SE	<i>Java 2 Standard Edition</i>	p. 52

# Resumo

---

---

---

Este trabalho apresenta modelos computacionais baseados em Sistemas Imunológicos Artificiais para resolver o problema da Predição da Estrutura de Proteínas no modelo Tri-dimensional Hidrofóbico-Polar. A Predição da Estrutura de Proteínas consiste em encontrar o arranjo espacial dos aminoácidos de uma proteína que possua energia mínima. A metodologia proposta tem foco nos Sistemas Imunológicos Artificiais mas utiliza como técnicas auxiliares os Sistemas de Inferência Fuzzy e a Busca Tabu. São aplicados dois tipos de algoritmos imunológicos: o algoritmo Clonalg e a Rede Imune. Em ambos os casos são comparados métodos baseados em penalidade com métodos que trabalham somente com soluções factíveis. Em alguns modelos implementados, um operador de Aging (que pode ser Fuzzy ou Puro) é usado para decidir quais anticorpos serão eliminados da população antes da etapa de seleção. Além disso, são implementados dois estágios de maturação da afinidade: fraco e intensivo, sendo o primeiro baseado nos operadores de hipermutação e hiper-macromutação e o segundo implementado através da Busca Tabu. Para validação dos resultados os algoritmos foram aplicados a algumas instâncias do *benchmark* Tortilla. Os resultados mostram que os modelos baseados no Clonalg superam os baseados na Rede Imune. Os melhores resultados foram obtidos com a modelo de Clonalg que uniu as três técnicas supracitadas e são comparáveis àqueles encontrados na literatura. Em alguns casos inclusive, novos valores de mínimo de energia foram encontrados pela metodologia proposta.

*Palavras-chave:* Sistemas Imunológicos Artificiais, Sistema de Inferência *Fuzzy*, Busca Tabu, Predição da Estrutura de Proteínas.

# *Abstract*

---

---

---

This work presents hybrid Immune-based Systems to solve the Protein Folding Problem for the three-dimensional hydrophobic-Polar model (3D HP). The Protein Structure Prediction consists of finding the spacial arrangement of a protein's amino acids that has a minimal energy. The proposed methodology is focused on Artificial Immune Systems supported by but uses Fuzzy Inference Systems and Tabu Search. Two kinds of algorithms are applied: the Clonal algorithm and the Immune Network algorithm. In both cases penalty based approaches are compared with approaches that allows only feasible antibodies. In some of the implemented models, an aging operator (that can be fuzzy or pure) is used to decide which antibodies will be eliminated of the population before the selection stage. Moreover, two affinity maturation stages were implemented: one weak and one intensive. The first is based on the hypermutation and hypermacromutation operators while the second is based on the tabu search metaheuristic. To validate the results, the models were applied to some instances of the Tortilla benchmark. The results show that the models based on the Clonal algorithm are superior to the models based on the immune network theory. The best results were achieved with the model that used the three technics previously commented. These results are comparable to those reported in the literature. In some cases new energy values were found by the proposed methodology.

*Keywords:* Artificial Immune Systems, Fuzzy Inference Systems, Tabu Search, Protein Structure Prediction.

## Capítulo

**1***Introdução*

---

---

Atualmente, uma quantidade imensa de dados tem sido gerada por diversos projetos genômicos no mundo todo. Estes dados, quando ainda não interpretados, têm pouca relevância tanto científica quanto comercial. Portanto, é de vital importância que as informações e os conhecimentos que estes dados representam sejam extraídos. Esta extração exige técnicas e recursos computacionais poderosos. Desta necessidade surgiu a área de Bioinformática (BERGERON, 2003), a qual alia conhecimentos de Biologia e Computação para solucionar de forma mais eficiente os problemas decorrentes dos projetos genômicos. A área de Bioinformática é uma área relativamente recente, tendo inúmeros ramos a serem pesquisados. Um dos desafios da área é o Problema da Predição da Estrutura de Proteínas (PPEP) ou Dobramento de Proteínas (DP).

O PPEP consiste em encontrar a estrutura tri-dimensional de proteínas a partir de sua seqüência de aminoácidos. Devido à complexidade do problema, diversos modelos computacionais têm sido aplicados na tentativa de solucioná-lo. Dentre estes modelos, destacam-se aqueles baseados em fenômenos naturais, os quais são denominados modelos de Computação Natural (CN) (DE CASTRO, 2006).

Dentre os vários modelos de CN, os Sistemas Imunológicos Artificiais (SIAs) vêm ganhando destaque nos últimos anos devido ao seu sucesso na solução de problemas que vão desde reconhecimento de caracteres (DE CASTRO, 2001) até otimização dinâmica (JUNQUEIRA et al., 2006).

Neste trabalho são aplicados modelos de SIAs para a Predição da Estrutura de Proteínas.

## 1.1 Motivações

Determinar a estrutura tridimensional de seqüências de proteínas nos genomas representa um dos maiores desafios para a biologia estrutural (ALBERTS et al., 2002). Sabe-se que cada proteína dobra-se em uma única conformação que é energeticamente favorável e especificada por sua seqüência de aminoácidos. Entretanto, evidências recentes sugerem que uma proteína pode dobrar-se em uma estrutura tri-dimensional alternativa como resultado de mutações ou outros motivos ainda não identificados (ALBERTS et al., 2005). Quando uma proteína encontra-se mau dobrada ela não somente perde a sua funcionalidade normal como pode também passar a desempenhar funções que prejudiquem a dinâmica da célula. Existem diversas doenças relacionadas direta ou indiretamente ao mau dobramento de proteínas, tais como: Mal de Alzheimer, Mal de Parkinson, Diabetes tipo II e vários tipos de câncer (ALBERTS et al., 2002).

A Predição da Estrutura de Proteínas ou o dobramento de proteínas é um dos desafios da área de bioinformática, pois ainda não existem informações suficientes para representá-lo formalmente. Portanto, ainda não se encontrou um modelo computacional para a predição da estrutura de proteínas que seja aplicável a proteínas reais com eficiência.

Existem alguns métodos experimentais para a determinação da estrutura tri-dimensional de proteínas (ver Capítulo 2). Porém, estes métodos são muito lentos (demorando até mesmo anos para a determinação da estrutura de uma proteína) e também dispendiosos.

Devido ao fato deste problema ter ainda pontos desconhecidos e dos métodos experimentais não serem eficientes (KOLINSKI; SKOLNICKB, 2004), modelos e

métodos computacionais têm sido aplicados na tentativa de solucionar o Problema da Predição da Estrutura de Proteínas. Neste trabalho, o PPEP é atacado por um método computacional baseado em Sistemas Imunológicos Artificiais e modelado através de um modelo de representação reduzido denominado 3D HP (hidrofóbico-polar tridimensional). Mesmo trabalhando em um modelo reduzido é provado que este problema é NP-Árduo (CRESCENZI et al., 1998; BERGER; LEIGHTON, 1998) e possui um grande grau de epistasia (sinergia), visto que a posição de um aminoácido depende da alocação de todos os outros.

O uso de SIAs justifica-se pelos bons resultados que os mesmos vêm obtendo na resolução de problemas de diversas áreas (exemplos destas aplicações podem ser encontrados na Seção 3.4). Em particular, o PPEP foi resolvido com sucesso através do uso de SIAs em Cutello, Nicosia e Pavone (2004) e Cutello et al. (2005).

Outro fator motivador para o uso de SIAs é que ainda existem muitos desafios a serem superados nesta área (TIMMIS, 2005), alguns dos quais foram abordados neste trabalho. Aproximar o modelo computacional do modelo biológico e disseminar a técnica foram alguns dos desafios tratados neste trabalho.

## 1.2 **Objetivos**

Esta dissertação tem como principal objetivo desenvolver modelos computacionais baseados em SIAs para resolver o problema da Predição da Estrutura de Proteínas no modelo 3D HP.

Os modelos desenvolvidos são baseados em dois princípios dos SIAs: princípio da seleção clonal e princípio da rede imune. Um dos objetivos deste trabalho é comparar o desempenho (qualidade das soluções e tempo computacional) destas duas técnicas no problema da predição da estrutura de proteínas no modelo 3D HP.

Trabalhos recentes (CUTELLO; NICOSIA; PAVONE, 2004; CUTELLO et al., 2005) introduziram um novo tipo de operador nos Sistemas Imunológicos Artificiais, o operador de *aging* e relataram melhoras decorrentes do uso deste operador. Neste trabalho é proposta uma versão com maior inspiração biológica para o operador de *aging*, denominada operador de *aging fuzzy*. Pretende-se avaliar a influência que esta nova versão do operador de *aging* tem tanto nos modelos baseados no princípio

da seleção clonal quanto naqueles baseados no princípio da rede imune.

Os modelos implementados trabalham com dois tipos de espaço de busca: o primeiro permite apenas soluções factíveis para o problema enquanto o segundo permite soluções infactíveis. Esta dissertação busca aferir o impacto da utilização destes dois tipos de espaços de busca, tanto com relação à qualidade das soluções quanto em tempo computacional necessário para encontrá-las.

Na tentativa de melhorar a qualidade dos resultados encontrados pelos modelos desenvolvidos, introduz-se uma etapa de maturação intensiva da afinidade. Esta maturação é implementada através de uma busca tabu. A contribuição dada por esta nova etapa também deve ser avaliada.

Além da comparação entre os modelos desenvolvidos durante este trabalho, pretende-se comparar os modelos que apresentam os melhores resultados com alguns resultados reportados na literatura.

## 1.3 Organização da Dissertação

Esta dissertação está dividida em sete capítulos. O Capítulo 2 descreve o problema da predição da estrutura de proteínas e o modelo de representação do problema utilizado. O Capítulo 3 apresenta uma breve introdução aos Sistemas Imunológicos Artificiais. Os Sistemas de Inferência *Fuzzy* e a Busca Tabu, que neste trabalho funcionam como técnicas auxiliares aos SIAs, são tratados de forma sucinta no Capítulo 4. O Capítulo 5 descreve a metodologia utilizada no desenvolvimento dos algoritmos propostos. As simulações computacionais e a discussão dos resultados são apresentadas no Capítulo 6. Por fim, o Capítulo 7 traz as conclusões e pesquisas futuras.

---

# *Predição da Estrutura de Proteínas*

---

---

As proteínas são polímeros ou cadeias de aminoácidos, onde os monômeros <sup>1</sup> de aminoácidos são ligados uns aos outros covalentemente. Existem vinte aminoácidos formadores de proteínas, cada qual com propriedades químicas distintas, conforme apresentado no Apêndice A.

Os aminoácidos são formados por um carbono central ( $C\alpha$ ), ao qual está ligado um grupo amina ( $NH_2$ ) e um grupo carboxila ( $COOH$ ). Esta parte é comum a todos os aminoácidos e é chamada cadeia polipeptídica, pois os aminoácidos se ligam entre si através de ligações covalentes peptídicas. O que diferencia um aminoácido do outro é sua cadeia lateral, também chamada radical (R), que se encontra ligada ao  $C\alpha$  (MELO et al., visitado em 04/2006). A Figura 2.1 ilustra a estrutura de um aminoácido.

As proteínas constituem a maior parte da massa celular (excetuando a água),

---

<sup>1</sup>Molécula, de massa molecular geralmente pequena, capaz de ligar-se a outras moléculas da mesma espécie, constituindo longas cadeias que formam um polímero (FERREIRA, 1999).

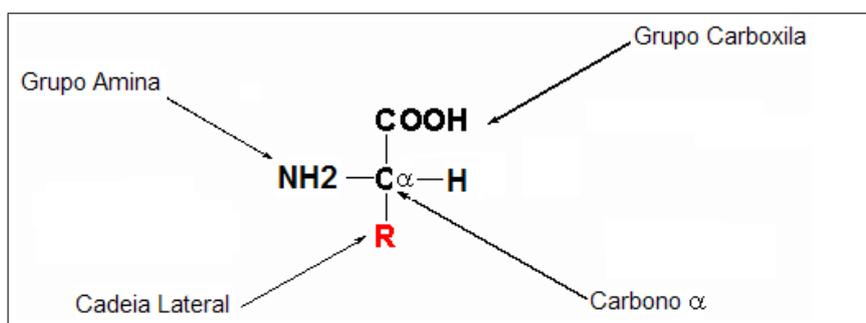


Figura 2.1: Estrutura básica de um aminoácido (MELO et al., visitado em 04/2006).

além de executar quase todas as funções celulares. Sob o ponto de vista químico, as proteínas possuem a estrutura mais complexa e sofisticada entre as moléculas conhecidas.

Uma proteína pode se dobrar de inúmeras formas, formando uma estrutura tridimensional, também chamada de conformação. O dobramento de uma proteína se dá por ligações não-covalentes que podem ser de três tipos: hidrogênio, iônica e atração de Van der Waals. Individualmente, as ligações não-covalentes são de 30 a 300 vezes mais fracas do que as ligações covalentes. Entretanto, muitas ligações não-covalentes podem formar duas regiões fortemente acopladas entre si. Assim, a estabilidade do dobramento da proteína é determinada pela força combinada de várias ligações não-covalentes (ALBERTS et al., 1997). A Figura 2.2 ilustra os três tipos de ligações não-covalentes.

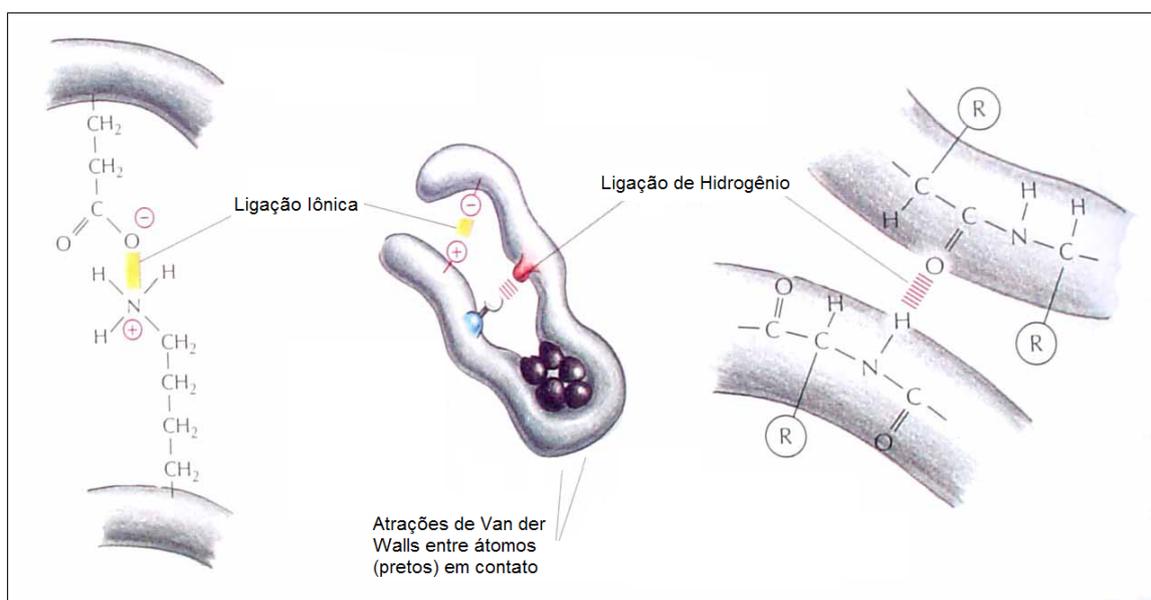


Figura 2.2: Tipos de ligações não-covalentes (adaptado de (ALBERTS et al., 2002)).

Uma quarta força, a hidrofobicidade, também tem uma função importante para determinar a forma de uma proteína. As moléculas hidrofóbicas, incluindo a cadeia lateral dos aminoácidos, tendem a se agrupar no interior da molécula de proteína em ambiente aquoso para minimizar o contato com a água. Por outro lado, as cadeias laterais polares tendem a se agrupar no exterior da molécula podendo formar ligações de hidrogênio com outras moléculas polares. Portanto, um fator importante no dobramento de proteína é a distribuição das cadeias laterais polares e não-polares ao longo da cadeia polipeptídica. O modelo computacional adotado neste trabalho leva em consideração apenas esta última força para avaliar o dobramento de uma proteína.

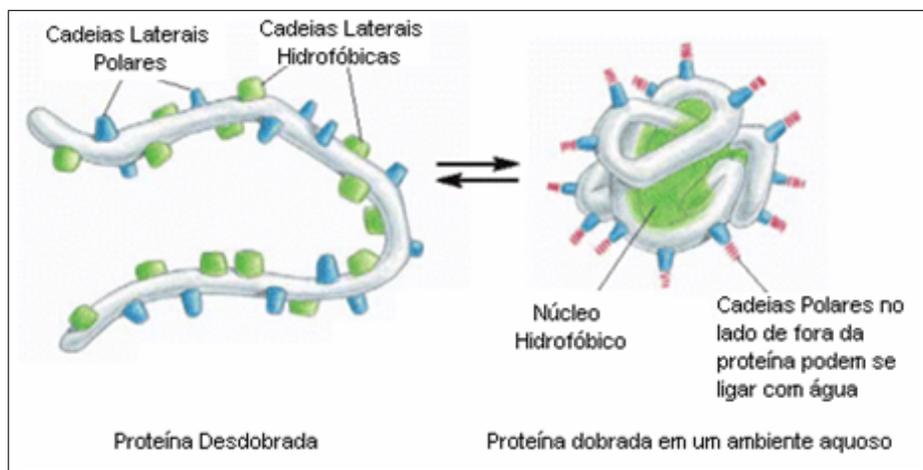


Figura 2.3: Forças hidrofóbicas atuando sobre a proteína (adaptado de (ALBERTS et al., 2002)).

Como resultado destas interações, cada proteína tem uma estrutura tridimensional particular, que é determinada pela ordem dos aminoácidos na cadeia polipeptídica. A conformação final da proteína é aquela que minimiza a energia livre. Uma proteína pode ser desdobrada, ou desnaturada, através de tratamento com certos solventes. Estes solventes quebram as ligações não-covalentes que ajudam a determinar a conformação da proteína. Este tratamento converte a proteína em uma cadeia polipeptídica flexível que perde sua forma natural. Removido o solvente, a proteína, em geral, dobra-se novamente em sua conformação original. Este fenômeno indica que toda informação necessária para especificar a conformação de uma proteína está contida na sua seqüência de aminoácidos (ALBERTS et al., 2002).

A estrutura de uma proteína é difícil de ser visualizada, pois mesmo pequenas proteínas são formadas por milhares de átomos ligados covalentemente e não-covalentemente. Os biólogos distinguem quatro níveis de organização de uma prote-

ína: estrutura primária, estrutura secundária, estrutura terciária e estrutura quaternária.

A seqüência de aminoácidos de uma proteína é conhecida como **estrutura primária** e é o nível estrutural mais simples de onde todos os outros níveis são derivados. Um exemplo de estrutura primária de uma proteína pode ser visto na Figura 2.4. A Tabela A.1 contendo os aminoácidos formadores de proteínas pode ser encontrada no Anexo A.

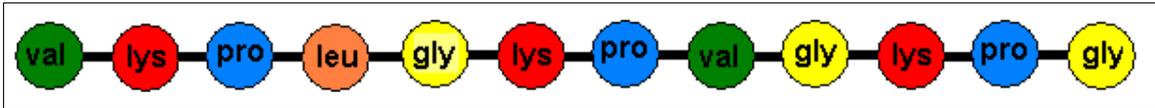


Figura 2.4: Estrutura primária de uma proteína.

Partes da cadeia polipeptídica que formam padrões estruturais, tais como  $\alpha$ -hlices e  $\beta$ -sheets, constituem a **estrutura secundária** das proteínas. Estes padrões podem ser observados na Figura 2.5.

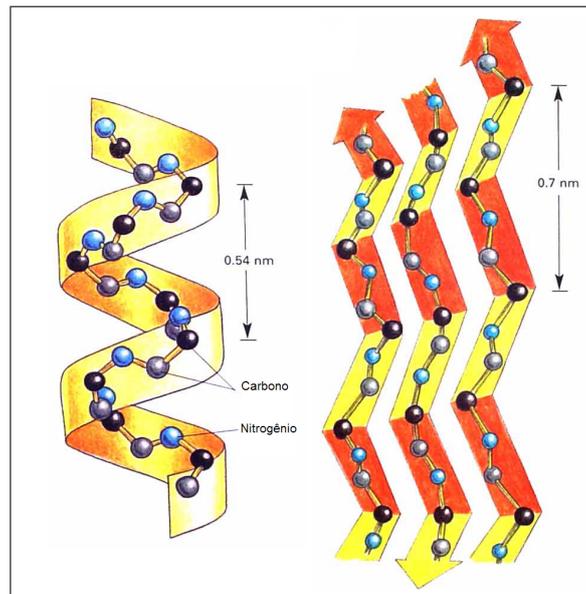


Figura 2.5: Padrões estruturais que compõem a estrutura secundária das proteínas.  $\alpha$ -hlices e  $\beta$ -sheets, respectivamente (adaptado de (ALBERTS et al., 2002)).

A estrutura tridimensional completa de uma proteína é chamada de **estrutura terciária**. Esta estrutura pode ser vista na Figura 2.6.

Finalmente, se uma molécula de proteína é formada como um complexo de uma ou mais cadeias polipeptídicas, esta organização é chamada de **estrutura quaternária**. A estrutura quaternária de uma proteína pode ser vista na Figura 2.7.

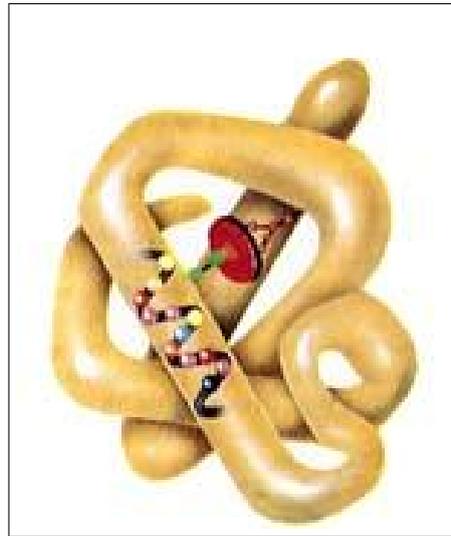


Figura 2.6: Estrutura terciária de uma proteína (MELO et al., visitado em 04/2006).



Figura 2.7: Estrutura quaternária de uma proteína (MELO et al., visitado em 04/2006).

As proteínas têm papel fundamental nos organismos vivos em virtude das mais variadas funções biológicas, como por exemplo: atividades catalíticas, defesa do organismo, transporte e armazenamento de nutrientes, regulação, entre outros.

A função que uma proteína exerce dentro de um organismo é dependente da sua estrutura tridimensional. Por isso, a determinação da estrutura de uma proteína é fundamental para a determinação da sua função biológica.

Existem formas experimentais de determinação da estrutura de proteínas como Cristalografia por Raio X, Microscopia *Cryoelectron* e Espectroscopia por (*Nuclear Magnetic Resonance*) NMR. A Cristalografia por Raio X provê as estruturas tridimensionais mais detalhadas porém requer a cristalização da proteína, o que é um processo custoso. A Microscopia *Cryoelectron* é mais utilizada para proteínas complexas, as quais são difíceis de ser cristalizadas. Apenas recentemente pequenas proteínas têm tido sua estrutura determinada por Espectroscopia NMR, a qual é uma técnica mais simples porém limitada a pequenas proteínas (ALBERTS et al., 2005).

A determinação da estrutura de proteínas através desses métodos exige sofisticados métodos físicos e complexa análise dos dados experimentais (ALBERTS et al., 2005). Além disso, estas técnicas são dispendiosas em relação a tempo e recursos financeiros. Segundo Alberts et al. (2005), cerca de 10 mil proteínas tiveram suas estruturas tri-dimensionais determinadas pela Cristalografia por Raio X.

Devido às dificuldades encontradas nos métodos experimentais, tem sido dada atenção especial aos métodos computacionais. Os métodos computacionais que abordam o dobramento de proteínas podem ser classificados em dois grupos: métodos *ab initio* e os métodos baseados em padrões (KOLINSKI; SKOLNICKB, 2004).

A primeira abordagem visa dobrar uma proteína sem nenhum conhecimento prévio de sua estrutura. Este método explora uma superfície de energia para identificar uma conformação de energia mínima, a qual acredita-se que seja o estado nativo da proteína. Infelizmente, o grande número de possíveis superfícies complica o processo de busca, o que leva pesquisadores a trabalhar com formas minimalistas de representar o dobramento de uma proteína na tentativa de restringir o espaço de busca.

Já os métodos baseados em padrões tentam encontrar a estrutura de uma pro-

teína comparando-a com proteínas que já têm suas estruturas determinadas. Os métodos baseados em padrões utilizam as estruturas de proteínas homólogas (semelhantes, de acordo com alguma medida de semelhança) à proteína alvo para determinar a sua estrutura, supondo que proteínas semelhantes possuem estruturas semelhantes.

Os métodos *ab initio* são computacionalmente mais complexos, gerando resultados inferiores àqueles obtidos com os métodos baseados em padrões, contudo, esses métodos podem ser aplicados para a determinação da estrutura de qualquer proteína (KOLINSKI; SKOLNICKB, 2004). Os métodos baseados em padrões são mais simples do que os *ab initio*, pois possuem mais informações (conhecimento) nas quais se basearem. A principal desvantagem é que eles são dependentes da existência de proteínas homólogas cujas estruturas já tenham sido determinadas com precisão através de métodos experimentais.

Os métodos computacionais trabalham sobre modelos que representam o dobramento de uma proteína de forma simplificada (KOLINSKI; SKOLNICKB, 2004). Os resultados obtidos no dobramento de proteínas é bastante dependente do modelo de representação utilizado. O modelo pode ser simples (fazendo com que o dobramento encontrado quase não lembre a proteína real) ou complexo (representando quase que fielmente a proteína dobrada). Os modelos de representação podem ser divididos em modelos baseados em *lattices* ou modelos livres. Os modelos baseados em *lattices* discretizam os valores que os parâmetros podem assumir enquanto os modelos livres permitem que os valores dos parâmetros assumam qualquer valor contínuo. As Figuras 2.8 e 2.9 apresentam, respectivamente, modelos de representação em *lattice* e livre.

Neste trabalho é realizado o dobramento de proteínas utilizando um método *ab initio* sobre um modelo baseado em *lattice*. O modelo utilizado é bastante difundido e conhecido como modelo hidrofóbico polar tri-dimensional (3D HP). Devido à sua importância neste trabalho, este modelo será descrito na seção seguinte.

## 2.1 Modelo Hidrofóbico-Polar Tridimensional

Devido à complexidade do Problema de Predição da Estrutura de Proteínas (PPEP), os modelos simplificados têm se tornado muito populares. Dentre esses

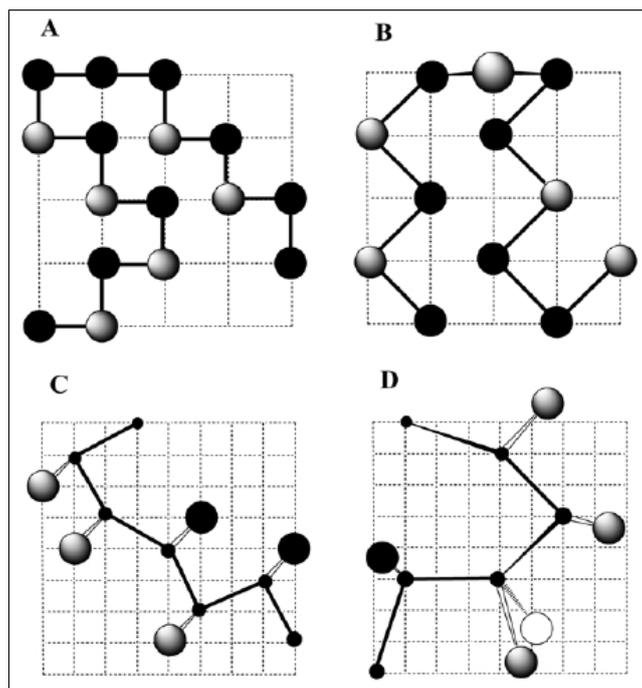


Figura 2.8: Modelos de Representação de Dobramentos baseados em lattices. (A) *Lattice* cúbico simples - quatro possíveis direções no *lattice*. (B) *Lattice* cúbico - oito possíveis direções no *lattice*. (C) 210 *lattice* - vinte e quatro possíveis direções no *lattice*. (D) 310 *lattice* - noventa possíveis direções no *lattice* (KOLINSKI; SKOLNICKB, 2004).

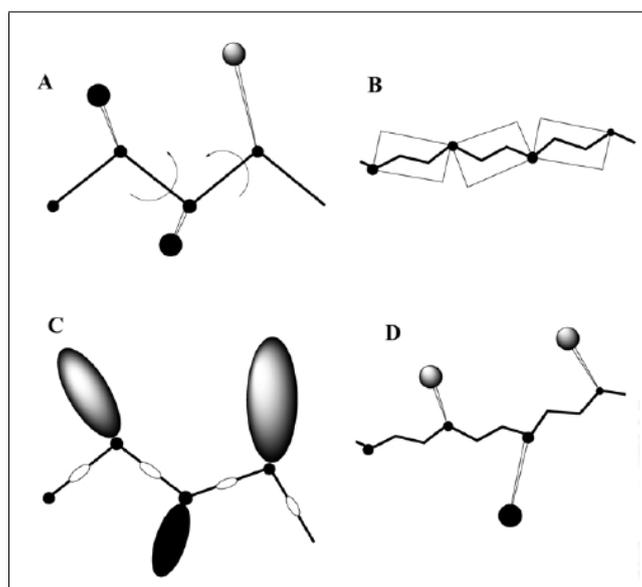


Figura 2.9: Modelos de Representação de dobramentos livres. (A) Modelo que considera dois átomos por resíduo, considerando apenas ângulos de torção. (B) Modelo onde três átomos devem ser movimentados em conjunto. (C) Modelo que considera dois centros de interação por resíduo: o centro da ligação peptídica e o centro da cadeia lateral. (D) Modelo que representa todos os átomos da cadeia principal e possui uma representação reduzida para a cadeia lateral. (KOLINSKI; SKOLNICKB, 2004).

modelos, o modelo Hidrofóbico-Polar (HP) (LAU; DILL, 1989) é o mais estudado e aplicado.

O HP é um modelo de energia livre baseado em dois importantes fatos (LAU; DILL, 1989) (RICHARDS, 1977):

- A interação hidrofóbica é a força guia do dobramento de proteínas e a hidrofobicidade dos aminoácidos é a principal força para o desenvolvimento da conformação nativa de pequenas proteínas globulares;
- As estruturas nativas de muitas proteínas são compactas e têm núcleos compactos com grande concentração de resíduos hidrofóbicos, assim como uma área mínima de superfície não-polar exposta no solvente.

No modelo HP os vinte (20) aminoácidos que compõem as proteínas são divididos em duas categorias: resíduos Hidrofóbicos/Não-polares (H) e resíduos Hidrofílicos/Polares (P), esta classificação pode ser encontrada no Anexo A. Então, a estrutura primária de uma proteína pode ser representada como uma *string* sobre o alfabeto {H,P}.

Como dois resíduos<sup>2</sup> não podem ocupar o mesmo lugar no espaço, conformações factíveis de uma seqüência no modelo HP são restritas a um caminho sem colisões em um *lattice*. Para o modelo 2D HP, um *lattice* quadrado bi-dimensional é tipicamente utilizado, enquanto o modelo 3D HP é baseado em um *lattice* cúbico tri-dimensional. Um exemplo de representação no modelo 3D HP de uma proteína com 50 resíduos é mostrado na Figura 2.10.

A representação de um caminho no *lattice* geralmente é feita através de coordenadas internas. Enquanto o sistema de coordenadas cartesianas mapeia os aminoácidos relativo a uma geometria espacial absoluta, o esquema de coordenadas internas define todas as posições relativas a uma estrutura arbitrária (como a primeira ligação em uma seqüência de aminoácidos). As coordenadas internas dependem da topologia do *lattice* considerado. Neste trabalho a representação dos dobramentos foi feita em um *lattice* tri-dimensional cúbico, onde cada posição (aminoácido) tem no máximo seis vizinhos (COTTA, 2003).

---

<sup>2</sup>Um resíduo de aminoácido é um aminoácido que se uniu a outro por uma ligação peptídica.

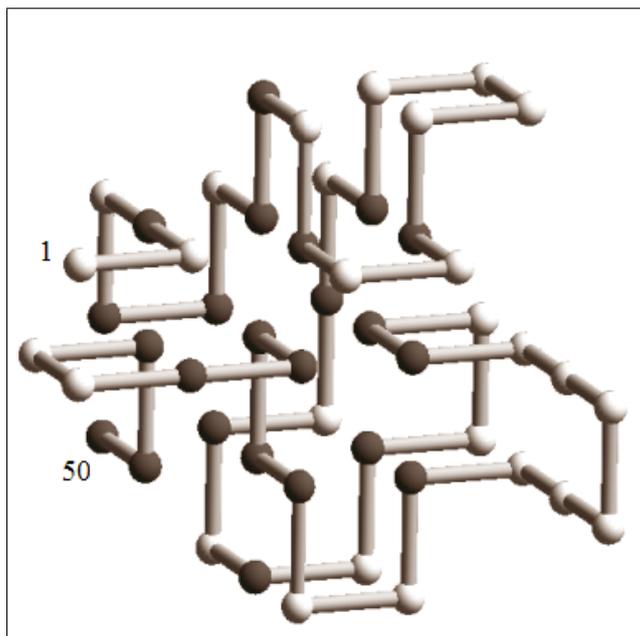


Figura 2.10: Representação do dobramento de uma proteína no modelo 3D HP: as esferas escuras representam resíduos hidrofóbicos e as esferas claras representam resíduos polares (SHMYGEL-SKA; HOOS, 2005).

Os dois principais esquemas para a representação dos movimentos internos são o absoluto e o relativo. O esquema absoluto adota um sistema de referência absoluto e todos os movimentos são especificados de acordo com esse referencial. Para o *lattice* cúbico um dobramento  $d$  é representado como uma *string* de comprimento  $L - 1$  no alfabeto {Norte, Sul, Leste, Oeste, Acima, Abaixo}, onde  $L$  é o total de aminoácidos. Já o esquema relativo não possui nenhum ponto de referência fixo, o que faz com que cada coordenada dependa da sua coordenada anterior. No *lattice* cúbico um dobramento utilizando esse esquema de coordenadas corresponde a uma *string* de comprimento  $L - 1$  no alfabeto {A frente, Girar Acima, Girar Abaixo, Girar a Esquerda, Girar a Direita}. Os dois esquemas possuem espaço de busca proporcional ao tamanho da proteína menos um porque o primeiro aminoácido é fixo (UNGER; MOULT, 1993). A Figura 2.11 ilustra os dois esquemas de representação de coordenadas internas. Neste trabalho foi adotado o esquema absoluto de representação, pois trabalhos anteriores compararam os dois tipos de representação e melhores resultados foram obtidos com o uso dessa representação (COTTA, 2003) (CUTELLO et al., 2005).

Uma conformação factível no modelo HP é associada a um nível de energia livre. Nesse modelo, a energia livre de uma conformação é proporcional ao número de contatos topológicos entre resíduos hidrofóbicos que não são vizinhos (contatos

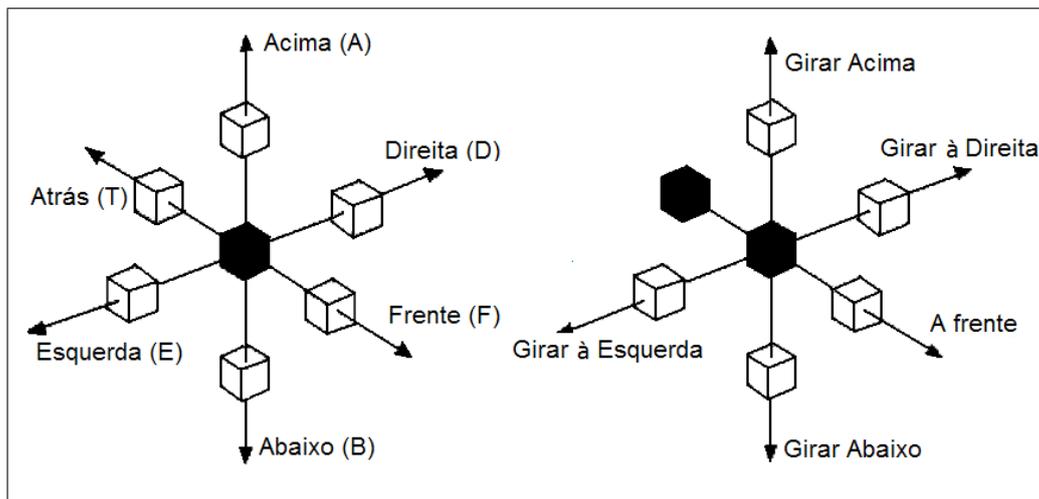


Figura 2.11: Representação do caminho no *lattice* por coordenadas: (a) Movimentos Absolutos: o cubo preto representa a localização corrente; (b) Movimentos Relativos: os cubos pretos representam a localização anterior e a localização corrente (adaptado de (COTTA, 2003)).

não-locais) na seqüência dada. Mais especificamente, a energia livre de uma certa conformação com  $\eta$  contatos hidrofóbicos não-locais é  $-\eta$ . O PPEP no modelo HP pode ser formalmente definido como: dada uma seqüência  $s = s_1s_2 \cdots s_n$  no alfabeto  $\{H,P\}$  encontre uma conformação de mínima energia de  $s$ , isto é, encontre  $c^* \in C(s)$  tal que  $E(c^*) = \min\{E(c) | c \in C\}$ , onde  $C$  é o conjunto de todas as conformações factíveis para  $s$  (SHMYGELSKA; HOOS, 2005).

Apesar de serem modelos reduzidos, foi provado que tanto no modelo 2D HP quanto no modelo 3D HP, o PPEP é NP-árduo (CRESCENZI et al., 1998; BERGER; LEIGHTON, 1998). O uso de métodos heurísticos para resolver o Problema da Predição da Estrutura de Proteínas no modelo HP é bastante justificável, pois não existem métodos determinísticos capazes de resolver esse problema.

## 2.2 Abordagens Relacionadas para a Predição da Estrutura de Proteínas no Modelo HP

A maioria das pesquisas no PPEP têm sido feitas no modelo 2D HP. Em (LOPES; SCAPIN, 2005) foi desenvolvido um Algoritmo Genético com operadores especiais para resolver o PPEP. O algoritmo obteve bons resultados em seqüências extraídas de um *benchmark* e de proteínas reais. Os Métodos de Monte-Carlo estão entre os melhores algoritmos para a Predição da Estrutura de Proteínas no

modelo 2D HP. Um exemplo dessa metodologia é o *Pruned Enriched Rosenbluth Method* (PERM) (HSU et al., 2003). O PERM é um algoritmo que foi responsável pela descoberta de alguns dos melhores resultados em *benchmarks*. Lau e Dill (1989) introduziram o algoritmo *Evolutionary Monte Carlo* (EMC). Esse é um Algoritmo Evolucionário que trabalha com uma população de indivíduos que realizam a otimização de Monte-Carlo. Esse algoritmo encontrou o melhor resultado conhecido para uma das seqüências do *benchmark* HP Tortilla (HART; ISTRAIL, visitado em 08/2005). A Otimização por Colônias de Formigas (OCF) também foi utilizada para resolver o problema de Predição da Estrutura de Proteínas no modelo 2D HP. Em Shmygelska e Hoos (2003), uma OCF com formigas melhoradas foi aplicada ao problema e alcançou resultados comparáveis a métodos do estado-da-arte. Em Cutello, Nicosia e Pavone (2004) foi aplicado um Sistema Imunológico Artificial com operador de *aging* determinístico, o qual obteve bons resultados.

Trabalhos mais recentes têm focado no modelo 3D HP. Cotta (2003) propôs um Algoritmo Evolucionário hibridizado com *Backtracking*. Ele comparou os resultados obtidos utilizando a representação relativa com aqueles obtidos com a representação absoluta. Também comparou o método utilizando espaço apenas com soluções factíveis, utilizando penalidades e fazendo reparos nas soluções infactíveis. Os melhores resultados foram obtidos utilizando a representação absoluta e reparando as soluções infactíveis. Em Shmygelska e Hoos (2005), uma OCF foi utilizada para solucionar o PPEP no 3D HP. Os resultados encontrados foram favoravelmente comparáveis com métodos do estado da arte. Uma Otimização por Colônias de Formigas paralelas foi utilizada em (CHU; TILL; ZOMAYA, 2005). De acordo com os autores, o método paralelo supera a implementação de uma única colônia tanto em termos de tempo de CPU quanto em qualidade das soluções. Em Cutello et al. (2005) foi utilizado um SIA baseado no princípio da seleção clonal com operador de *aging* determinístico e células B de memória. Os resultados foram comparados favoravelmente com aqueles obtidos em Cotta (2003). Em Blazewicz, Lukasiak e Milostan (2005) a técnica Busca Tabu foi aplicada como único método para resolver o Problema da Predição da Estrutura de Proteínas no modelo 3D HP. Os resultados obtidos motivam o uso desta metaheurística como técnica complementar a outras técnicas.

---

## *Sistema Imunológico Artificial*

---

---

Os sistemas imunológicos artificiais, como outras técnicas inspiradas na natureza, tentam extrair dos sistemas biológicos idéias para desenvolver ferramentas que resolvam problemas computacionais (DE CASTRO; TIMMIS, 2002b; DASGUPTA, 1998).

O Sistema Imunológico (SI) natural é um mecanismo complexo responsável pela defesa do organismo (em especial dos animais vertebrados) contra o ataque de microorganismos externos que podem causar doenças (bactérias, fungos, vírus, entre outros) chamados de antígenos (DE CASTRO; TIMMIS, 2002a). O propósito principal do SI é reconhecer todas as moléculas presentes no organismo e classificá-las como próprias (antígenos próprios) e não-próprias (patogênicas). Para as moléculas rotuladas como não-próprias, o sistema imunológico é ativado e introduz um tipo apropriado de defesa a partir das células brancas ou leucócitos (DE CASTRO, 2001). Nesta defesa, anticorpos<sup>1</sup> são produzidos com a função principal de neutralizar e eliminar o antígeno que estimulou a sua produção.

O sistema imunológico pode produzir uma resposta imediata (característica do

---

<sup>1</sup>Anticorpos são proteínas produzidas pelos plasmócitos (linfócitos B).

sistema imune inato) à invasão dos antígenos, assim como pode produzir uma resposta mais lenta e duradoura (função do sistema imune adaptativo). Os macrófagos e granulócitos responsáveis pela resposta imunológica inata estão disponíveis nos organismos de forma constante, não exigindo a pré-infecção para o combate a um determinado invasor (DE CASTRO, 2001).

Já o sistema imune adaptativo, representado pelos linfócitos B e linfócitos T, tem a capacidade de produzir uma resposta específica para um dado antígeno. Os linfócitos possuem uma memória imunológica, pois reconhecem o reaparecimento do antígeno e evitam uma nova contaminação pela doença, aperfeiçoando-se a cada encontro com um invasor ou similares (DE CASTRO, 2001).

Os Sistemas Imunológicos Artificiais (SIAs) são compostos por metodologias inteligentes inspiradas no sistema imunológico natural para a solução de problemas do mundo real (DASGUPTA, 1998). Existem vários algoritmos e aplicações para os SIAs (DE CASTRO; VON ZUBEN, 2000a). Os principais mecanismos herdados dos sistemas imunológicos reais são:

- **Memória Imunológica e Maturação de Afinidade:** quando o sistema imunológico reconhece um antígeno, ele faz uma clonagem das células B que têm maior afinidade com o invasor (antígeno). Parte das células que são clonadas se tornam parte da memória do sistema de forma a facilitar combates futuros contra esse mesmo antígeno ou semelhantes. A outra parte sofre uma hipermutação clonal que pode melhorar a sua eficiência no combate ao invasor. Dessa forma, uma segunda resposta a um antígeno é provavelmente mais eficiente do que a primeira;
- **Reconhecimento de Padrões:** cada anticorpo é capaz de reconhecer uma determinada quantidade de antígenos com os quais ele se associa;
- **Diversidade Imunológica:** através da hipermutação somática e da recombinação genética é garantida a diversidade imunológica;
- **Princípio da Seleção Clonal:** esse princípio estabelece que apenas a célula (anticorpo) capaz de reconhecer um determinado estímulo antigênico irá se proliferar, sendo portanto selecionada em detrimento das outras;
- **Distinção Próprio/Não-Próprio:** o sistema imunológico é capaz de distinguir entre células invasoras e células do próprio organismo, sendo que apenas

as primeiras devem ser combatidas;

- **Teoria da Rede Imunológica:** sugere que o sistema imunológico é composto por uma rede regulada de células e moléculas que se reconhecem mesmo na ausência de antígenos.

Esses princípios serviram como base para a implementação de diversos algoritmos. Neste trabalho são desenvolvidos modelos computacionais baseados no princípio da seleção clonal e na teoria da rede imunológica. Por isso, esses dois princípios serão melhor explicados nas Seções 3.1 e 3.2.

## 3.1 Princípio da Seleção Clonal

O princípio da seleção clonal explica as características básicas da resposta imune adaptativa ao ataque de um antígeno. Este princípio estabelece que somente a célula que é capaz de reconhecer um determinado estímulo antigênico deve proliferar-se, sendo selecionada em detrimento das outras (DE CASTRO; VON ZUBEN, 2000b).

Quando ocorre o ataque de um antígeno a um organismo, as células do SI respondem a esse ataque através de um conjunto (população) de anticorpos. Cada célula é capaz de reconhecer um número restrito de invasores, além de ter o poder de reconhecer antígenos próprios daquele organismo. As células reconhecedoras de antígenos próprios são eliminadas do repertório de anticorpos antes que sejam capazes de exercer uma resposta (seleção negativa). Já aquela célula que melhor reconhece o antígeno estranho é selecionada para proliferar-se. Após a proliferação, os clones sofrem um processo de mutação somática a altas taxas na tentativa de melhorar a afinidade dos clones com o antígeno invasor, chamada de hipermutação somática. A prole (clones) gerada a partir deste anticorpo diferencia-se em células de memória (células com um longo ciclo de vida) ou em plasmócitos, células que excretam anticorpos que irão atacar o invasor. A Figura 3.1 ilustra o princípio da seleção clonal (DE CASTRO; TIMMIS, 2002a).

Para que o sistema imunológico seja eficiente é necessário que ele seja capaz de adaptar a resposta imune a cada ataque sofrido pelo organismo e também que o número de anticorpos que irá atacar uma população de antígenos seja suficiente para eliminá-la.

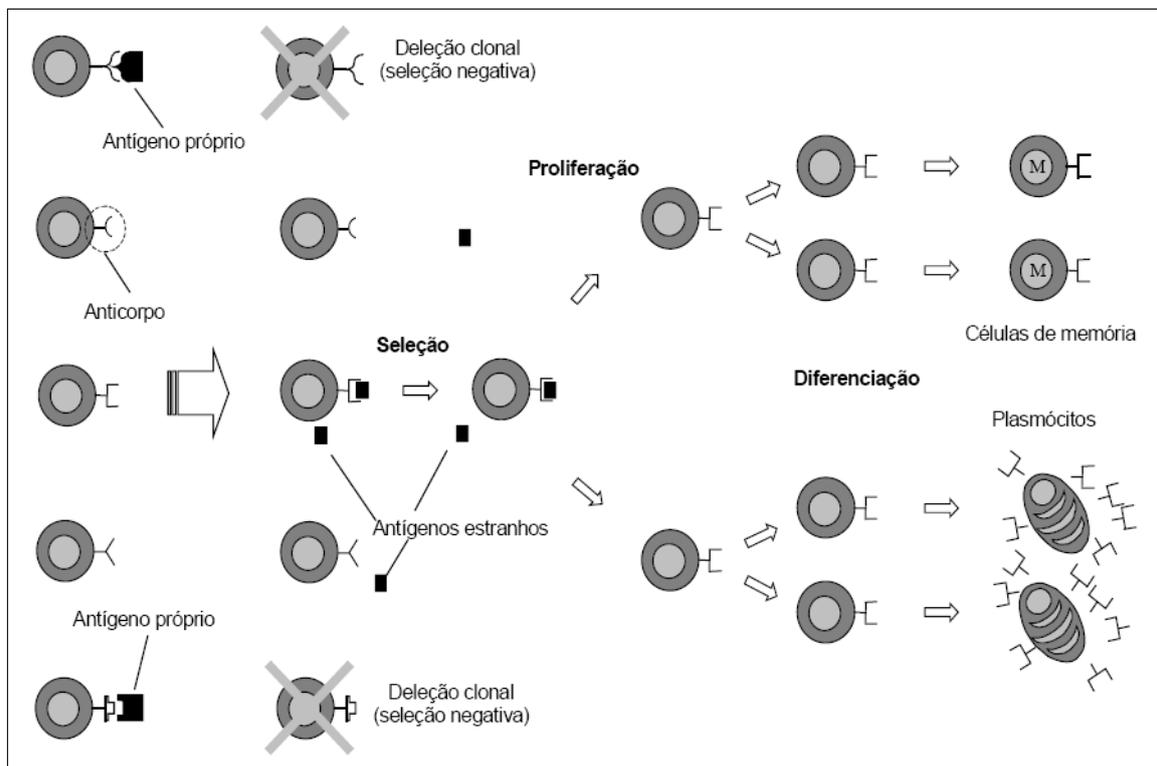


Figura 3.1: Princípio da seleção clonal (DE CASTRO, 2001).

No processo evolutivo do SI um antígeno é apresentado a um organismo diversas vezes. Em uma primeira exposição a quantidade de anticorpos, com alta afinidade com tal antígeno, disponíveis para a resposta imune é pequena. Em um segundo ataque o sistema imunológico já é capaz de produzir uma resposta mais rápida e precisa, pois ele guarda no organismo células capazes de gerar anticorpos (células B) com alta afinidade àquele antígeno, essas células são chamadas de células de memória. Então quanto maior é a exposição de um organismo a determinado invasor mais eficiente é a resposta imunológica gerada, o que demonstra a capacidade de aprendizagem dos SIs. A Figura 3.2 ilustra a memória imunológica e a aprendizagem por reforço apresentada pelo sistema imunológico biológico.

Os anticorpos que fazem parte de uma resposta imunológica secundária possuem, no geral, maior afinidade com o antígeno do que os anticorpos presentes na resposta primária. Isso é possível graças ao fenômeno chamado maturação da afinidade. Esta maturação dos anticorpos acontece devido a alterações nas células B durante a expansão clonal (hipermutação somática). Essas alterações podem gerar células que irão resultar em anticorpos de maior afinidade com o antígeno ou então gerar células de baixa afinidade ou que reconheçam antígenos próprios. Essas células que

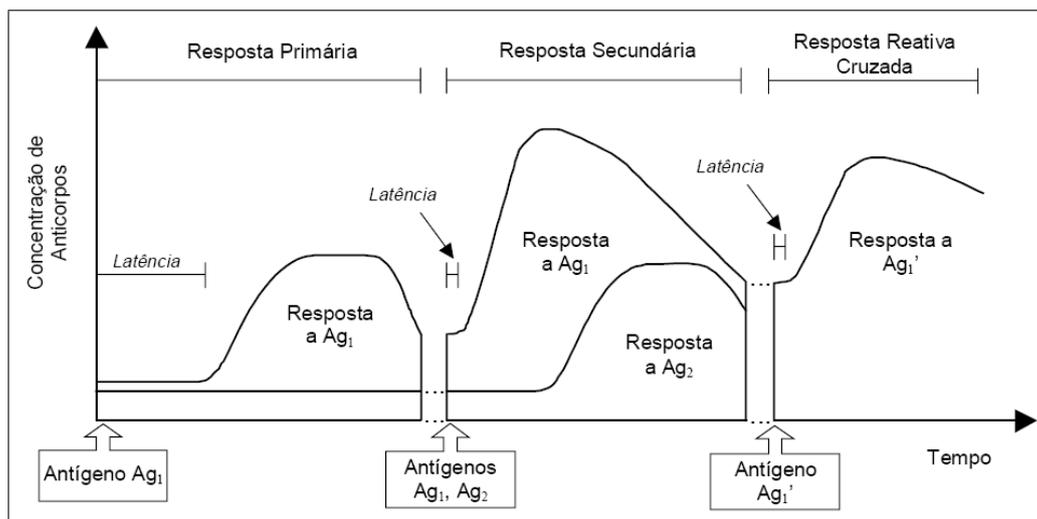


Figura 3.2: Resposta Primária, Secundária e Reativa Cruzada. Após a exposição primária de um antígeno  $Ag_1$ , novas exposições do mesmo antígeno (resposta secundária) ou de um antígeno similar  $Ag_1'$  (reação cruzada) farão com que o sistema imunológico gere uma resposta mais rápida e eficiente. A resposta primária de  $Ag_2$ , um antígeno diferente de  $Ag_1$ , é bem semelhante à resposta primária de  $Ag_1$  (DE CASTRO, 2001).

resultarão em anticorpos de baixa afinidade são eliminadas do repertório de células imunológicas (DE CASTRO, 2001). A maturação acontece a cada re-exposição do antígeno ao organismo e faz com que a afinidade dos anticorpos cresça constantemente (Figura 3.3).

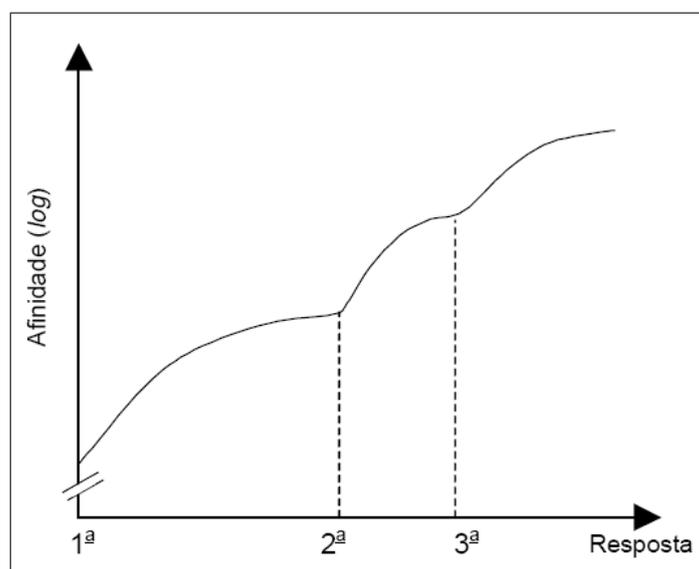


Figura 3.3: Maturação da afinidade ao longo das respostas imunes (DE CASTRO, 2001).

A quantidade de mutação que uma célula B recebe é regulada pelo processo de seleção. A seleção das células B de melhor afinidade diz qual célula ainda deve con-

tinuar sendo mutada ou não. As células que possuem alta afinidade com o antígeno devem ser pouco (ou nada) mutadas enquanto aquelas que possuem baixa afinidade devem continuar sofrendo mutações a altas taxas (DE CASTRO; TIMMIS, 2002a).

## 3.2 Teoria da Rede Imune

A teoria da Rede Imunológica, ou Rede Idiotípica, foi proposta por Jerne (1974) para explicar as capacidades de memória e aprendizado dos sistemas imunes. A teoria da rede imune define o SI como uma rede regulada de células e moléculas que se reconhecem mesmo na ausência de antígenos. A principal hipótese desta teoria diz que a memória imune é mantida mesmo sem a presença de antígenos devido às interações entre as células do sistema imunológico.

Um antígeno possui apenas regiões que podem ser reconhecidas por um anticorpo, essas regiões são chamadas de epitopo. Já um anticorpo possui uma região de reconhecimento de antígenos (paratopo) e também uma região que pode ser reconhecida por outros anticorpos (idiotopo). Assim, a produção de um anticorpo  $Ab_1$  estimula (ou suprime) a produção de outro anticorpo  $Ab_2$  que estimula (ou suprime) a produção de outro anticorpo  $Ab_3$  e assim por diante (PARISI, 1990). A Figura 3.4 ilustra a dinâmica do sistema imunológico de acordo com a teoria de rede imune. Se um anticorpo reconhece uma célula sua produção é estimulada pela rede. Caso um anticorpo seja reconhecido por outro anticorpo sua produção é inibida (supressão). Desta forma, a rede possui um mecanismo auto-regulatório de produção de anticorpos quer um antígeno esteja presente no sistema ou não.

As redes imunológicas possuem três características principais (BERSINI; VARELA, 1990):

- **Estrutura:** descrição dos componentes celulares e moleculares da rede e suas interconexões, desconsiderando as interações entre estes componentes;
- **Dinâmica:** interação constante entre os diversos componentes do sistema;
- **Meta-dinâmica:** o sistema imunológico pode interagir com qualquer elemento, para isso a rede precisa estar sempre produzindo novos anticorpos. Como o repertório imunológico tem tamanho constante, para que novos anticorpos sejam gerados outros precisam morrer o que faz com que o sistema

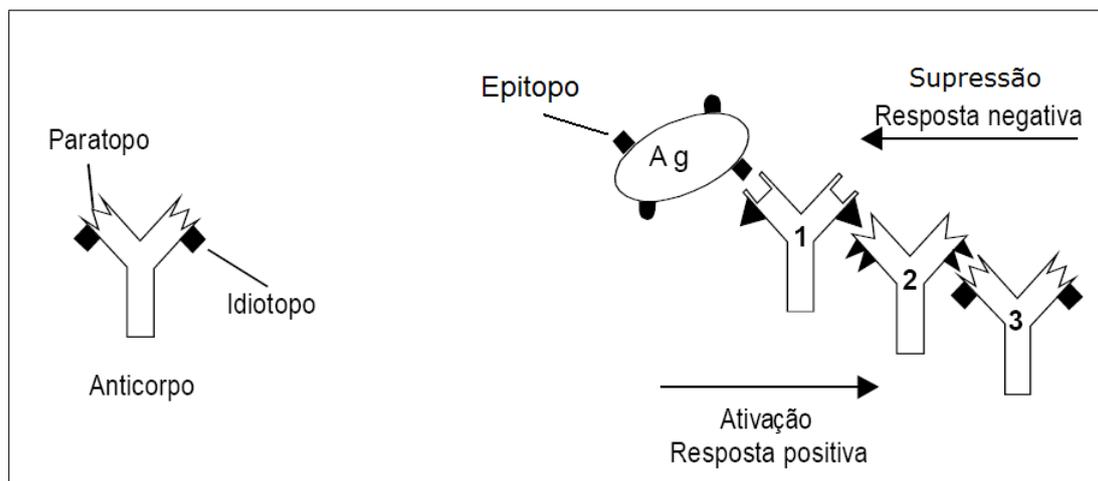


Figura 3.4: Teoria da Rede Imunológica. (a) Anticorpo com destaque para o paratopo e idiotopo. (b) Respostas positivas e negativas resultantes da interação de anticorpos com antígenos (epitopo) e com anticorpos (idiotopo) (DE CASTRO, 2001).

imunológico esteja em constante renovação. Essa renovação garante diversidade para que o SI possa reagir a novos ataques.

A teoria da rede imune provê uma descrição mais complexa e precisa do funcionamento do sistema imunológico como um sistema em constante evolução para se adequar ao seu ambiente. A visão do sistema imune como uma rede agrega menos importância a cada célula, delegando maior importância ao sistema como um todo. No princípio da seleção clonal um único clone é responsável por combater o antígeno, na teoria da rede imune o sistema como um todo é responsável por isso (dada a forte correlação entre as várias células do sistema). Deste modo, a teoria da rede imune provê uma abstração mais robusta do funcionamento do sistema imunológico.

Segundo DE CASTRO (2001), a teoria da rede é particularmente interessante para o desenvolvimento de ferramentas computacionais, pois ela fornece uma medida aproximada de propriedades emergentes como aprendizagem e memória, tolerância ao próprio, tamanho e diversidade de populações celulares. Através do mecanismo de estímulo e supressão, a rede é capaz de aprender e memorizar quais células ela deve ou não produzir (o que ela deve ou não fazer). A tolerância ao próprio diz respeito à capacidade do sistema de não atacar células e moléculas próprias ao organismo, respondendo apenas àquelas que são externas ao organismo. Apesar do tamanho do repertório imune ser constante as populações celulares podem variar em tamanho e diversidade. Uma população celular, neste contexto, é um conjunto de células ou moléculas que reconhecem idiotopos e epitopos semelhantes. O tamanho

e a diversidade das populações celulares são reguladas pelos estímulos e inibições inerentes a rede imune.

### 3.3 Sistema Imunológico Artificial

A modelagem de um sistema imunológico artificial envolve pelo menos três etapas: a representação formal dos elementos do sistema, a medida de afinidade entre estes elementos e o algoritmo que representa o princípio imune implementado. As Subseções 3.3.1 e 3.3.2 descrevem conceitos relacionados à representação, medida de afinidade e biblioteca gênica, enquanto a Subseção 3.3.3 descreve os algoritmos de engenharia imunológica nos quais os algoritmos propostos neste trabalho se basearam.

#### 3.3.1 Espaço de Formas

O espaço de formas é um formalismo proposto em (PERELSON; OSTER, 1979) para descrever matematicamente os anticorpos e antígenos e quantificar seu grau de interação. Conforme descrito anteriormente, um paratopo<sup>2</sup> reconhece um idiotopo ou epitopo. Contudo, não foi explicado que este reconhecimento se dá através de complementaridade química e física. Não é necessária a total complementaridade entre um paratopo e um idiotopo ou epitopo para que ocorra o reconhecimento (um exemplo é demonstrado na Figura 3.5). Quanto maior o nível de complementaridade, maior maior o nível de reconhecimento, também conhecido como afinidade (GOMES, 2006).

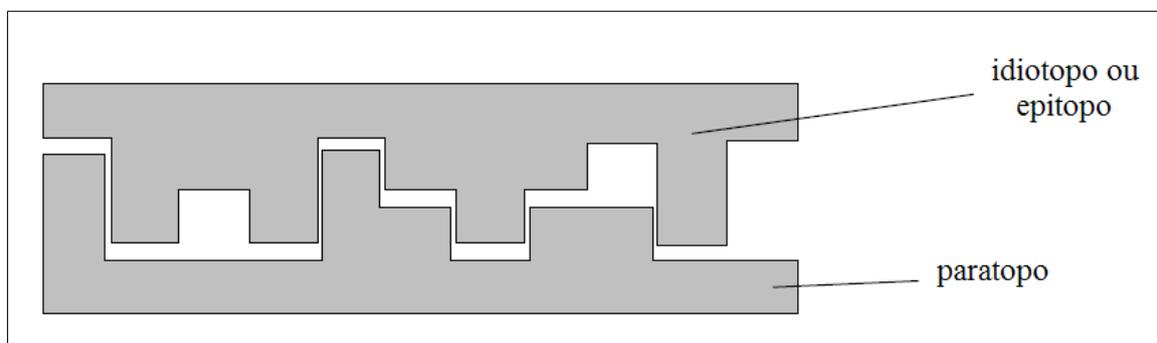


Figura 3.5: Reconhecimento por complementaridade (DE CASTRO, 2001).

---

<sup>2</sup>Paratopo é a parte do anticorpo que reconhece corpos estranhos.

No formalismo do espaço de formas, um paratopo possui um número pré-definido de pontos de ligação (pontos de reconhecimento) que pode ou não ser o mesmo número de pontos a serem reconhecidos num epitopo ou idiotopo. Como os valores que estes pontos podem assumir (pontos do paratopo, idiotopo e epitopo) estão contidos no mesmo espaço (ou seja, podem assumir o mesmo conjunto de valores) e como neste espaço sempre existe um valor complementar, um número finito de anticorpos é capaz de reconhecer um conjunto potencialmente infinito de idiotopos/epitopos.

Dentro do espaço de formas existem regiões que um paratopo é capaz de reconhecer. Estas regiões são vizinhas no espaço de formas e determinadas pelo limiar de afinidade, o qual determina o grau mínimo de complementaridade entre um paratopo e um idiotopo ou epitopo para que o primeiro seja capaz de reconhecer o segundo.

Sob o ponto de vista computacional, geralmente não existe a distinção entre um paratopo, idiotopo ou epitopo. Ou seja, na maior parte das implementações computacionais a parte de um anticorpo que reconhece e é reconhecida é a mesma. Outra simplificação que geralmente é feita é que o grau de afinidade é medido por similaridade e não por complementaridade.

A medida de afinidade utilizada depende do problema e geralmente corresponde a uma medida de distância. Mas no caso particular de problemas de otimização, como é o caso da Predição da Estrutura de Proteínas no modelo 3D HP, a medida de afinidade é, em geral, uma função matemática que deve ser minimizada ou maximizada.

### 3.3.2 Bibliotecas Gênicas

O material genético de um anticorpo está armazenado em cinco bibliotecas separadas e distintas, conhecidas por bibliotecas gênicas (DE CASTRO, 2001). A produção de um anticorpo se dá através da concatenação de componentes selecionados aleatoriamente a partir de cada uma das bibliotecas.

Um modelo computacional biologicamente plausível de construção do repertório de anticorpos envolve a utilização de bibliotecas gênicas. Um número variável de bibliotecas gênicas pode ser utilizado e não necessariamente todas as bibliotecas devem ter o mesmo tamanho. A Figura 3.6 dá uma visão pictórica do processo.

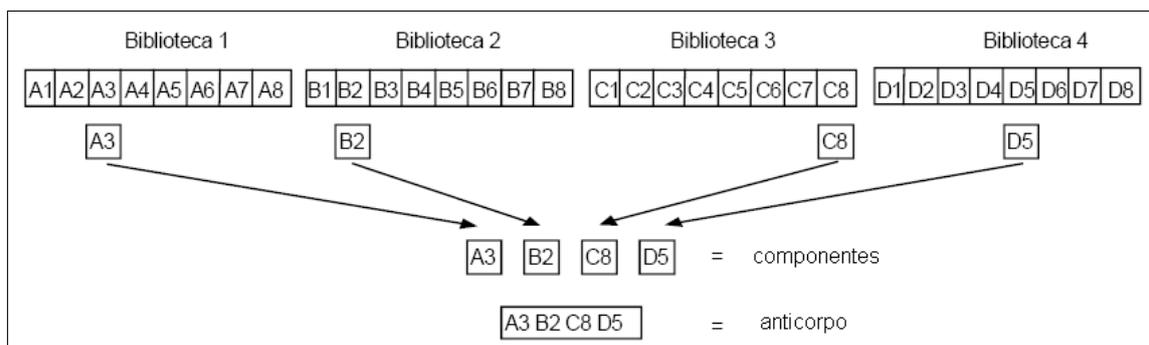


Figura 3.6: Processo de construção de um anticorpo utilizando biblioteca gênica (DE CASTRO, 2001).

### 3.3.3 Algoritmos

Existem diversos algoritmos baseados em princípios dos sistemas imunológicos, dentre os quais destacam-se o Clonalg (DE CASTRO; VON ZUBEN, 2000b) (DE CASTRO, 2001) (DE CASTRO; VON ZUBEN, 2002), o opt-aiNet (TIMMIS; EDMONDS, 2004) (DE CASTRO; TIMMIS, 2002a) e o copt-aiNet (GOMES, 2006) (FRANÇA; GOMES, 2006). Como os modelos computacionais propostos neste trabalho são variações destes algoritmos, os mesmos serão brevemente descritos nas subseções a seguir.

#### Clonalg

O algoritmo Clonalg (*Clonal Selection Algorithm*) é o algoritmo baseado em sistemas imunológicos mais amplamente difundido. Como o nome sugere, ele foi inspirado pelo princípio da seleção clonal, em particular na geração e proliferação de clones, na presença de antígenos e na modificação dos clones através da aplicação da hipermutação somática.

O Clonalg foi inicialmente proposto para resolver problemas de aprendizado de máquina e reconhecimento de padrões e foi posteriormente adaptado para resolver problemas de otimização (DE CASTRO, 2001).

Existem algumas diferenças entre a versão do algoritmo utilizada para problemas de reconhecimento de padrões e a versão utilizada para otimização. Neste trabalho será focada apenas a versão para resolver problemas de otimização. A versão canônica do algoritmo Clonalg para otimização é dada pelo Algoritmo 1.

---

**Algoritmo 1** Forma Básica do Algoritmo Clonalg (DE CASTRO; VON ZUBEN, 2000b)

---

**Clonalg**(Tam<sub>POP</sub>, Fator<sub>DUP</sub>, Taxa<sub>HIP</sub>)  
 gerações  $\leftarrow$  0;  
 POP = Inicialização(Tam<sub>POP</sub>);  
 Avalia(POP);  
**Enquanto**(!Condição\_de\_Parada) **faça**  
   POP<sub>c</sub>  $\leftarrow$  Clonagem(POP, dup);  
   POP<sub>h</sub>  $\leftarrow$  Hipermutação(POP<sub>c</sub>, Taxa<sub>HIP</sub>);  
   Avalia(POP<sub>h</sub>);  
   POP  $\leftarrow$  Seleção(POP, POP<sub>h</sub>);  
   Substitui  $d$  piores anticorpos por anticorpos novos;  
**Fim Enquanto**

---

O algoritmo Clonalg recebe como parâmetros o tamanho da população (Tam<sub>POP</sub>), a taxa de hipermutação (Taxa<sub>HIP</sub>) e o número de clones que será gerado a partir cada anticorpo da população (Fator<sub>DUP</sub>). Além disso, é necessário definir a representação dos anticorpos (que pode ser binária, real ou inteira) e a medida de afinidade entre os anticorpos e os antígenos.

A geração da população inicial costuma ser aleatória, mas pode fazer uso da teoria das bibliotecas gênicas (ou qualquer outra metaheurística), a qual é seguida pela avaliação da população. A clonagem replica cada um dos anticorpos da população FATOR<sub>DUP</sub> vezes. Esta é a principal característica que diferencia a versão do Clonalg para otimização da versão para reconhecimento de padrões (DE CASTRO, 2001), já que no segundo o número de clones gerados por cada anticorpo é proporcional à afinidade deste anticorpo com o antígeno atual. Os clones são então submetidos a um processo de hipermutação somática, o qual geralmente corresponde a uma mutação com taxa inversamente proporcional à afinidade do clone. A população de clones é avaliada e os melhores anticorpos entre a população original e os clones hipermutados são selecionados para compor a próxima população. Após isto, os  $d$  piores anticorpos são retirados da população e substituídos por novos anticorpos gerados aleatoriamente. Este processo se repete até que a condição de parada seja alcançada.

O algoritmo possui algumas características importantes, dentre as quais destacam-se a capacidade de auto-adaptação da mutação, introdução de diversidade na população e preservação dos melhores anticorpos.

O algoritmo básico tem sido modificado recentemente. Em (CUTELLO; NICOSIA;

PAVONE, 2004) são analisados diferentes tipos de operadores de hipermutação. Uma versão adaptativa do algoritmo Clonalg é proposta em (GARRETT, 2004). O uso de um operador de *aging* é proposto em (CUTELLO et al., 2005). Em (WATKINS, 2005) são propostas diferentes versões paralelas de um algoritmo Clonalg bastante modificado (o AIRS, que é utilizado para o reconhecimento de padrões). O uso da metáfora da seleção clonal para a solução de problemas de otimização multi-objetivo é investigado em (COELLO; CORTES, 2005).

Neste trabalho, será avaliado o uso no algoritmo Clonalg no PPEP. São propostas várias alterações do modelo supracitado e o impacto disso é avaliado como pode ser visto no Capítulo 6.

## Opt-Ainet

Existem diversos modelos computacionais baseados na teoria da Rede Imune (GALEANO; VELOZA-SUAN; GONZÁLEZ, 2005). Os primeiros modelos de redes imunes foram desenvolvidos para tratar problemas de compressão e agrupamento de dados (DE CASTRO; TIMMIS, 2002a; DE CASTRO, 2001). O algoritmo opt-aiNet (*Artificial Immune Network for Optimization*) proposto em DE CASTRO e Timmis (2002a) foi desenvolvido para tratar problemas de otimização multimodal. O algoritmo pode ser considerado uma extensão do algoritmo Clonalg com passos envolvendo as interações anticorpos-anticorpos e não apenas as interações antígenos-anticorpos. A Opt-aiNet é apresentada no Algoritmo 2.

O algoritmo inicia-se com a geração aleatória e avaliação da população inicial. Em seguida o algoritmo entra no seu laço principal, que é executado até que o critério de parada seja satisfeito. O critério de parada proposto em (DE CASTRO; TIMMIS, 2002a) é guiado pela variação do tamanho da população: o laço principal é interrompido quando o tamanho da população não variar entre duas iterações consecutivas.

Por ter sido inspirado no algoritmo Clonalg, ele possui um ciclo de maturação da afinidade que é repetido até que a rede atinja sua estabilidade. No caso da Opt-aiNet, a estabilidade da rede é medida pela diferença entre a afinidade média da população atual e a afinidade média da população anterior. Assim como ocorre com a versão do Clonalg para otimização, a Opt-aiNet também gera um número

---

**Algoritmo 2** Forma básica do algoritmo opt-Ainet (DE CASTRO; TIMMIS, 2002a)

---

**Rede Imune**(Tam<sub>POP\_INI</sub>, Fator<sub>DUP</sub>, Taxa<sub>HIP</sub>, Limiar<sub>SUP</sub>)

gerações ← 0;

POP = Inicialização(Tam<sub>POP\_INI</sub>);

Avalia(POP);

**Enquanto**(!Condição\_de\_Parada) **faça**

**”Enquanto**(Instabilidade\_da\_Rede)

    POP<sub>c</sub> ← Clonagem(POP, dup);

    POP<sub>h</sub> ← Hipermutação(POP<sub>c</sub>, Taxa<sub>HIP</sub>);

    Avalia(POP<sub>h</sub>);

    POP ← Seleção(POP, POP<sub>h</sub>);

    gerações ← gerações+1;

**Fim Enquanto**

  POP ← Supressão(POP, Limiar<sub>SUP</sub>);

  Introduza  $d\%$  \* TAM<sub>POP\_ATUAL</sub> novos anticorpos;

**Fim Enquanto**

---

fixo (Fator<sub>DUP</sub>) de clones de cada célula (anticorpo) da rede, ou seja, o número de clones não é proporcional à afinidade da célula. Após a clonagem, ocorre a hipermutação dos clones. Esta hipermutação, geralmente, é aplicada a taxas inversamente proporcionais à afinidade de cada clone. Os clones de uma célula formam uma subpopulação, da qual a melhor célula é selecionada e substitui sua célula pai caso sua afinidade seja superior.

Após a estabilidade da rede ser alcançada, ocorre uma etapa de supressão. Esta etapa corresponde à interação entre as células da rede, onde tenta-se eliminar a redundância de células na população. O primeiro critério proposto para a eliminação de células redundantes correspondia a eliminar qualquer uma de duas células que possuíam grande similaridade, porém Timmis e Edmonds (2004) sugerem que a célula eliminada seja aquela de menor afinidade (solução de qualidade inferior). Depois disso, é feita uma inserção de novas células geradas aleatoriamente para permitir a exploração de novas regiões do espaço de busca.

As características principais da opt-aiNet são (DE CASTRO; TIMMIS, 2002a):

- Mecanismo determinístico e elitista de seleção: a seleção de qual célula irá fazer parte da próxima população corresponde a um torneio entre a célula pai e seus clones hipermutados. Esta seleção favorece células de melhor afinidade (elitismo) e ocorre de maneira determinística (a melhor célula sempre é escolhida);

- Determinação automática do tamanho da população: através da etapa de supressão e de introdução de novas células, a rede é capaz de controlar automaticamente seu tamanho, ou seja, se o problema possuir muitos ótimos a rede vai tender a se expandir para criar células que representem os mesmos;
- Combinação de busca local e busca global: a rede realiza uma busca local na vizinhança de cada célula pai através do operador de hipermutação. Como, por causa da supressão, as células tendem a estar espalhadas pelo espaço de busca, a rede exhibe a propriedade de busca global;
- Critério de convergência automático: o critério de parada não precisa ser definido pois ele já é característico da rede e tende a se adaptar às necessidades de cada problema;
- Capacidade de localizar e manter soluções ótimas estáveis: como o tamanho da população é auto ajustável, a rede pode crescer de maneira a acomodar novas soluções (sub-)ótimas encontradas e, a menos que existam soluções ótimas muito próximas, a rede irá manter automaticamente todas as soluções ótimas encontradas.

## Copt-Ainet

A *copt-aiNet* (*Artificial Immune Network for Combinatorial Optimization*) é uma abordagem de rede para a solução de problemas combinatórios (GOMES, 2006). A *copt-aiNet* pode ser considerada uma extensão da *opt-aiNet* com a adição dos passos de maturação fraca da afinidade, maturação intensiva da afinidade e tamanho mínimo da rede. A *copt-aiNet* inspira-se em três conceitos oriundos do sistema imunológico biológico (GOMES, 2006):

- Teoria da Rede Imunológica (detalhes na Seção 3.2);
- Seleção Clonal e Maturação da Afinidade (detalhes na Seção 3.1);
- Bibliotecas Gênicas (detalhes na Subseção 3.3.2).

A forma básica da *copt-aiNet* é apresentada pelo Algoritmo 3.

---

**Algoritmo 3** Forma básica do algoritmo copt-Ainet (FRANÇA; GOMES, 2006)

---

**Rede Imune**(Tam<sub>POP\_INI</sub>, Fator<sub>DUP</sub>, Taxa<sub>HIP</sub>, Limiar<sub>SUP</sub>)  
 gerações ← 0;  
 POP = Inicialização(Tam<sub>POP\_INI</sub>);  
 Avalia(POP);  
**Enquanto**(!Condição\_de\_Parada)**faça**  
   **Enquanto**(Instabilidade\_da\_Rede)  
     POP<sub>c</sub> ← Clonagem(POP, dup);  
     POP<sub>h</sub> ← Hipermutação(POP<sub>c</sub>, Taxa<sub>HIP</sub>);  
     Avalia(POP<sub>h</sub>);  
     POP ← Seleção(POP, POP<sub>h</sub>);  
     gerações ← gerações+1;  
**Fim Enquanto**  
 POP ← Supressão(POP, Limiar<sub>SUP</sub>);  
**Se** (Tam<sub>POP\_ATUAL</sub> < Tam<sub>POP\_MIN</sub>) **então**  
   Gerar novos anticorpos para completar a população;  
**Fim Se**  
**Se** !(K\_melhores sofreram melhorias durante K\_Iterações) **então**  
   POP ← MaturaçãoFrac(POP);  
**Fim Se**  
**Fim Enquanto**  
 POP ← MaturaçãoIntesiva(K\_melhores);

---

Semelhante ao que ocorre na rede opt-aiNet, e em concordância com o princípio da seleção clonal, as células imunes passam por uma etapa de clonagem seguida do processo de maturação da afinidade através da hipermacromutação. A hipermutação é aplicada a taxas inversamente proporcionais a afinidade das células da rede. Na etapa de supressão os anticorpos similares são identificados e aqueles que possuem menores afinidades com o antígeno são eliminados da população.

Após a etapa de supressão o tamanho mínimo que a população pode assumir é verificado. Se o número de células for menor que o tamanho mínimo da rede (parâmetro fixo do algoritmo), novos anticorpos são gerados e inseridos na população pelo processo de recombinação gênica (utilizando o conceito de bibliotecas gênicas).

Se durante um determinado número de iterações do algoritmo nenhum dos *k\_melhores* anticorpos tiver sua afinidade melhorada toda a população passa pelo processo de maturação fraca da afinidade. Diferente da opt-aiNet, a copt-aiNet possui um critério de parada baseado na convergência das melhores células, ou seja, se nenhuma das *k\_melhores* células tiver a sua afinidade melhorada no decorrer de um número de gerações pré-determinado o algoritmo pára.

Quando o critério de parada é alcançado os  $k$ -*melhores* anticorpos são submetidos ao processo de maturação intensiva da afinidade, como uma última tentativa do algoritmo de melhorar a afinidade destas células.

Na versão original do algoritmo (FRANÇA; GOMES, 2006) tanto a maturação fraca da afinidade como a maturação intensiva da afinidade são implementadas através de uma Busca Tabu. Ambas as etapas de maturação melhoram a capacidade de intensificação da busca da rede.

Neste trabalho serão implementados diversos modelos computacionais inspirados tanto na rede opt-aiNet como na rede copt-aiNet.

### 3.4 Aplicações

Apesar de ser uma área relativamente nova (cerca de dez anos) já existem aplicações dos Sistemas Imunológicos Artificiais em diversas áreas (DE CASTRO, 2001). Abaixo serão comentadas algumas áreas e aplicações de SIAs para elas:

- **Mineração de Dados:** é o processo de análise de conjuntos de dados que tem por objetivo a descoberta de padrões interessantes e que possam representar informações úteis. Existem diversos SIAs desenvolvidos para tratar este problema. Em Alves et al. (2004a) é proposto um novo algoritmo de indução de regras de classificação em mineração de dados, chamado IFRAIS. O IFRAIS busca descobrir um conjunto de regras *fuzzy*, através de um processo evolucionário baseado em SIA e procedimentos usuais em mineração de dados. O trabalho reporta resultados para várias bases de dados, alcançando resultados comparáveis a populares algoritmos para mineração de dados. Um algoritmo baseado nas metáforas de seleção negativa e seleção clonal é proposto em Lee et al. (2003) para consultas adaptativas em *web* semântica;
- **Busca e Otimização:** são problemas que consistem na determinação de uma solução ótima dentro de um espaço de busca. Exemplos de uso de SIAs são encontrados em: (FRANÇA; VON ZUBEN; DE CASTRO, 2005), (COELLO; CORTES, 2005), (WALKER; GARRETT, 2003), (CUTELLO; NICOSIA; PAVONE, 2003) e (COSTA et al., 2002). Em (FRANÇA; VON ZUBEN; DE CASTRO, 2005) é proposta uma rede imune que estende a rede

opt-aiNet para ambientes dinâmicos. Esta rede foi denominada dopt-aiNet. A dopt-aiNet possui algumas características relevantes como ajuste automático do limiar de supressão e da taxa de mutação e novos tipos de mutação e foi aplicada a problemas de otimização estáticos e dinâmicos obtendo bons resultados. Uma alternativa para a solução de problemas multi-objetivos com ou sem restrições é proposta em Coello e Cortes (2005). Neste trabalho um algoritmo baseado no princípio da seleção clonal, fazendo uso de memória secundária e função de aptidão baseada na dominância dos anticorpos é aplicado a funções clássicas de *benchmarks*. O problema de coloração de grafos é resolvido por um algoritmo imune baseado no princípio da seleção clonal que incorpora um processo de busca local em (CUTELLO; NICOSIA; PAVONE, 2003);

- **Aprendizagem de Máquina:** é o estudo de algoritmos que permitem a uma máquina aprender. Nos seguintes trabalhos dessa área foram utilizados SIAs: Watkins, Timmis e Boggess (2004) e Bersini (1999). Em Watkins, Timmis e Boggess (2004) é proposto um algoritmo para classificação que faz uso do princípio da seleção clonal e ARB (*Antigenic Recognition Ball*). Os ARBs são utilizados para restringir a proliferação de anticorpos redundantes;
- **Robótica:** problemas encontrados ou decorrentes da construção de robôs. Algumas aplicações de SIAs para esse problema podem ser encontradas em Ishiguro et al. (1998), Canham, Jackson e A. (2003) e Vargas et al. (2003). Ishiguro et al. (1998) propuseram um SIA baseado no princípio de moderação para controlar a movimentação de um robô com seis pernas. A implementação de uma rede imuno-genética para a movimentação autônoma de um robô do tipo Khepera II é proposta em Vargas et al. (2003). Nesta rede o sistema imune implementa o processo de decisão enquanto o algoritmo evolutivo define a estrutura da rede;
- **Controle de Processos:** criação de sistemas que automatizam o controle de máquinas e processos quer sejam industriais ou não. Aplicações de SIA nessa área são feitas em: (DING; REN, 2000), (KIM, 2001) e (LAW; WONG, 2004). Um controle *fuzzy* auto ajustável inspirado no mecanismo de *feedback* do sistema imune é proposto em Ding e Ren (2000) para controlar a temperatura da pele de pessoas hipertermia. Em Kim (2001) é proposta uma rede imune para ajustar os parâmetros de um controlador PID (Proporcional-Integral-Derivativo);

- **Segurança Computacional e de Rede:** visa identificar e tratar possíveis invasores nos sistemas computacionais. Nos seguintes trabalhos dessa área foram utilizados SIAs: (ANCHOR et al., 2002), (WANG; HIRSBRUNNER, 2002) e (DE PAULA; DE CASTRO; GEUS, 2004). Em (ANCHOR et al., 2002) é proposto um algoritmo baseado no princípio da seleção positiva que identifica possíveis ataques a uma rede de computadores. Um sistema que faz uso de princípios do sistema imune inato, da teoria do perigo, da seleção e seleção negativa é proposto em DE PAULA, DE CASTRO e Geus (2004) para identificar ataques de *buffer overflow*;
- **Detecção de Falhas e Anomalias:** visa identificar possíveis invasores em sistemas computacionais e defeitos em qualquer tipo de sistema. O uso de sistemas imunológicos artificiais para a solução deste tipo de problema é abordado em Bradley e Tyrrell (2000), Lee et al. (2005), González e Dasgupta (2003) e Greensmith, Aickelin e Twycross (2004). Em Lee et al. (2005) a detecção de falha num contador é resolvido pelo uso de um sistema imune com ênfase na diversidade populacional, a qual é evoluída por evolução simbiótica. A detecção de intrusos é tratada através dos princípios da teoria do perigo em Greensmith, Aickelin e Twycross (2004);
- **Predição da Estrutura de Proteínas:** O algoritmo Clonalg Cutello, Nicosia e Pavone (2004) foi aplicado ao PPEP no modelo 2D HP. Em (CUTELLO et al., 2005; CUTELLO et al., 2006) o algoritmo Clonalg, usando operador de *aging* puro, foi aplicado ao problema da predição de proteínas nos modelos 2D e 3D HP. Modelos baseados no algoritmo Clonalg trabalhando em representações livres de *lattice* foram utilizados nos trabalhos de Cutello, Narzisi e Nicosia (2006) e Anile A. M. Cutello et al. (2006). O primeiro trabalho trata o PPEP como sendo um problema de otimização multi-modal, enquanto o segundo trata o problema como sendo multi-objetivo. Os dois utilizam o CHARMM<sup>3</sup> (MACKERELL et al., 1998) como função de energia a ser minimizada.
- **Outros:** outras aplicações de sistemas imunes podem ser encontradas em: Zhou, Dasgupta D.; Zhiling e Hongmei (2006), Rahman, Yasin e Abdullah (2004), Ishak, Abidin e Rahman (2004) e Campelo et al. (2006).

---

<sup>3</sup>CHARMM (*Chemistry at HARvard Macromolecular Mechanics*) é um programa para simulações de macromoléculas, incluindo minimização de energia, dinâmica molecular e simulações de Monte Carlo.

---

## *Técnicas Auxiliares*

---

---

Para aprimorar o desempenho dos Sistemas Imunológicos Artificiais, foram adicionadas aos modelos propostos técnicas auxiliares. Foram utilizadas duas técnicas: Sistemas de Inferência *Fuzzy* (SIF) (ZADEH, 1965) e Busca Tabu (BT) (GLOVER; LAGUNA, 1993).

Neste trabalho, um SIF é utilizado para determinar a possibilidade de morte de um anticorpo decidindo quando ele deve ser eliminado da população. Já a BT é implementada no estágio de maturação intensiva da afinidade aplicada periodicamente à população, na tentativa de melhorar sua afinidade.

Estes dois tópicos são amplos e envolveriam uma longa explanação sobre os vários conceitos associados. No entanto, o objetivo aqui é apenas apresentar as informações mais relevantes e necessárias para o entendimento da metodologia. Portanto, estas duas técnicas são brevemente descritas nas seções seguintes.

## 4.1 Sistema de Inferência Fuzzy

Os Sistemas de Inferência *Fuzzy* (SIF) são modelos computacionais baseados em conceitos da teoria de conjuntos *fuzzy*, regras *fuzzy* e raciocínio *fuzzy* (PEDRYCZ; GOMIDE, 1998).

A teoria de conjuntos *fuzzy* foi proposta por Zadeh (1965) com o objetivo de fornecer uma ferramenta matemática para o tratamento de informações imprecisas ou vagas. Os conjuntos *fuzzy* formam a base teórica para o entendimento dos SIFs e serão brevemente descritos na Subseção 4.1.1.

Os sistemas *fuzzy* têm sido aplicados com sucesso em diversas áreas, tais como: controle automático, classificação e reconhecimento de padrões, tomada de decisão, sistemas inteligentes, previsão de séries temporais e robótica (JANG; SUN; MIZUTANI, 1997).

Segundo Jang, Sun e Mizutani (1997) a estrutura básica de um sistema de inferência *fuzzy* possui três componentes conceituais:

- Base de dados: definição das funções de pertinência usadas nas regras *fuzzy*;
- Base de Regras: especifica o total de regras e quais são elas; e
- Mecanismo de raciocínio: procedimento de inferência sobre as regras e fatos para gerar saídas (ou conclusões) adequadas.

### 4.1.1 Introdução aos Conjuntos Fuzzy

Segundo Delgado (2002), a noção de conjunto ocorre freqüentemente quando se tenta organizar, resumir e generalizar o conhecimento a respeito de objetos. Seja  $\mathbf{X}$  uma coleção de objetos denominados genericamente por  $x$ . Então, um conjunto  $A$  é definido por uma coleção de pares ordenados

$$A = \{(x, \mu_A(x)) \mid x \in \mathbf{X}\} .$$

A função  $\mu_A(x)$  é denominada função de pertinência e determina com que grau um objeto  $x$  pertence a um conjunto  $A$  e  $\mathbf{X}$  é chamado de universo. Em conjuntos

clássicos, apenas dois valores para  $\mu_A(x)$  são permitidos: o elemento pertence ou não pertence a um determinado conjunto. Na teoria dos conjuntos *fuzzy* a transição entre pertencer e não pertencer é gradual.

Como exemplo, considere  $\mathbf{X} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$  uma coleção de números inteiros. Seja  $A$  um conjunto *fuzzy* que define “números inteiros próximos a 5”, dado por

$$A = \{(1, 0), (2, 0.4), (3, 0.6), (4, 0.8), (5, 1), (6, 0.8), (7, 0.6), (8, 0.4), (9, 0)\}.$$

Então, o conjunto *fuzzy*  $A$  pode ser definido como uma coleção de objetos com valores de pertinência variando entre 0 (exclusão completa) e 1 (pertinência completa). Os conjuntos *fuzzy* representam, portanto, uma generalização dos conjuntos clássicos. O conceito de função de pertinência, fundamental para a teoria dos conjuntos *fuzzy*, será discutido a seguir.

## Funções de Pertinência

Toda a teoria dos conjuntos está baseada no conceito de pertinência. Um conjunto *fuzzy* é definido pela função de pertinência  $\mu_A(x)$  que estabelece para cada  $x$  um grau de pertinência ao conjunto  $A$ , com  $\mu_A(x) \in [0, 1]$  (DELGADO, 2002). Os conjuntos clássicos podem ser vistos como um caso particular dos conjuntos *fuzzy*, no qual apenas os limites do intervalo são utilizados na definição da função de pertinência:  $\mu_A(x) \in \{0, 1\}$ ,  $x \in \mathbf{X}$ .

Por exemplo, considerando-se o universo contínuo  $\mathbf{X} \in \mathfrak{R}$ , o conceito de “números próximos a 5” pode ser expresso de forma diferente dependendo da definição da função de pertinência associada:

- “números próximos a 5” na concepção clássica

$$\mu_A(x) = \begin{cases} 0 & \text{se } x \leq 4.5 \\ 1 & \text{se } 4.5 < x \leq 5.5 \\ 0 & \text{se } x > 5.5 \end{cases}$$

- “números próximos a 5” na concepção de conjuntos *fuzzy*

$$\mu_A(x) = \begin{cases} 0 & \text{se } x \leq 4.5 \\ \frac{x-4.5}{0.5} & \text{se } 4.5 < x \leq 5 \\ \frac{5.5-x}{0.5} & \text{se } 5 < x \leq 5.5 \\ 0 & \text{se } x > 5.5 \end{cases}$$

A Figura 4.1 ilustra as diferenças entre as funções de pertinência no caso de conjuntos clássicos e *fuzzy*.

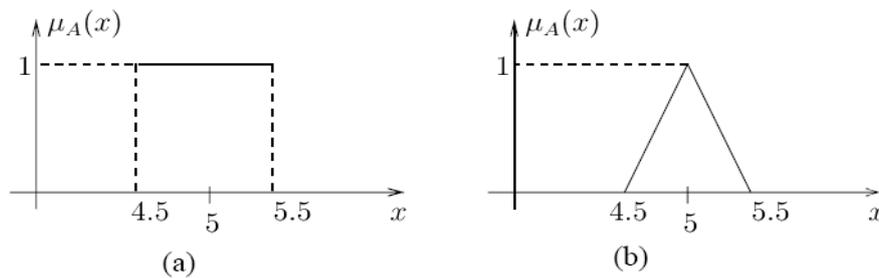


Figura 4.1: Funções de pertinência: (a) Conjuntos clássicos, (b) Conjuntos *fuzzy* (DELGADO, 2002).

Em geral, o formato das funções de pertinência é restrito a uma certa classe de funções, representadas por alguns parâmetros específicos. Os formatos mais comuns são: triangular, trapezoidal e Gaussiana. Além dos formatos tradicionais existe uma forma bastante utilizada em aplicações práticas: o conjunto unitário (*singleton*).

- Função Triangular: parâmetros  $(a, m, b)$ , com  $a \leq m \leq b$

$$\mu_A(x) = \begin{cases} 0 & \text{se } x \leq a \\ \frac{x-a}{m-a} & \text{se } a < x \leq m \\ \frac{b-x}{b-m} & \text{se } m < x \leq b \\ 0 & \text{se } x > b \end{cases}$$

- Função Trapezoidal: parâmetros  $(a, m, n, b)$ , com  $a \leq m, n \leq b$  e  $m < n$

$$\mu_A(x) = \begin{cases} 0 & \text{se } x \leq a \\ \frac{x-a}{m-a} & \text{se } a < x \leq m \\ 1 & \text{se } m < x \leq n \\ \frac{b-x}{b-n} & \text{se } n < x \leq b \\ 0 & \text{se } x > b \end{cases}$$

- Função Gaussiana: parâmetros  $(m, \sigma_k)$ , com  $\sigma_k > 0$

$$\mu_A(x) = \exp^{-\sigma_k(x-m)^2};$$

- Conjunto Unitário (*singleton*): parâmetros  $(m, h)$

$$\mu_A(x) = \begin{cases} h & \text{se } x = m \\ 0 & \text{caso contrário} \end{cases}$$

A Figura 4.2 ilustra estes diferentes formatos, especificando os parâmetros associados a cada caso.

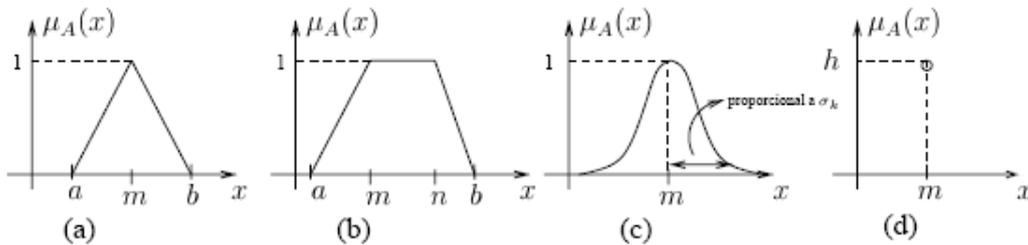


Figura 4.2: Diferentes formatos das funções de pertinência: (a) Triangular, (b) Trapezoidal, (c) Gaussiana, (d) *Singleton*.

A escolha do formato mais adequado nem sempre é óbvia, podendo inclusive não estar ao alcance do conhecimento de um especialista para a aplicação em questão. No entanto, existem sistemas *fuzzy* cujos parâmetros das funções de pertinência são completamente definidos pelo especialista. Nestes casos, a escolha de funções triangulares e trapezoidais é mais comum porque a idéia de se definirem regiões de pertinência total, média e nula é mais intuitiva do que a especificação do valor modal e dispersão, conceitos ligados às funções Gaussianas (JANG; SUN; MIZUTANI, 1997). Neste trabalho, as funções de pertinência são definidas pelo especialista e têm formatos triangulares e trapezoidais.

### 4.1.2 Base de Dados

A base de dados contém informações relacionadas às variáveis lingüísticas do sistema e suas partições. Uma variável lingüística pode ser definida como uma variável onde os valores são palavras ou sentenças ao invés de números (PEDRYCZ; GOMIDE, 1998).

A definição formal de uma variável lingüística envolve a cobertura do domínio da variável com uma série de conjuntos *fuzzy* e a definição do significado de cada conjunto. Segundo Zadeh (1965) uma variável lingüística é dada por uma quintupla:  $\langle X, T(x), \mathbf{X}, G, M \rangle$ , onde  $X$  corresponde ao nome da variável lingüística,  $T(x)$  representa o conjunto de termos lingüísticos (cada elemento de  $T(x)$  corresponde a um rótulo  $l$  dos termos que a variável pode assumir),  $\mathbf{X}$  é o universo da variável  $X$ ,  $G$  corresponde à gramática para a geração de termos ou rótulos e  $M$  representa uma regra que associa a cada rótulo  $l$  um conjunto fuzzy no universo  $\mathbf{X}$  (representando o seu significado).

Um exemplo de uma variável lingüística pode ser visto na Figura 4.3. Neste caso,  $X = \text{Afinidade de um Anticorpo}$  e o conjunto de termos lingüístico é dado por  $T(\text{Afinidade de um Anticorpo}) = \{\text{Baixa}, \text{Média}, \text{Alta}\}$ . Os três termos lingüísticos são associados a funções de pertinência trapezoidais. O universo  $\mathbf{X}$  da variável está limitado ao intervalo  $[0,1]$ .

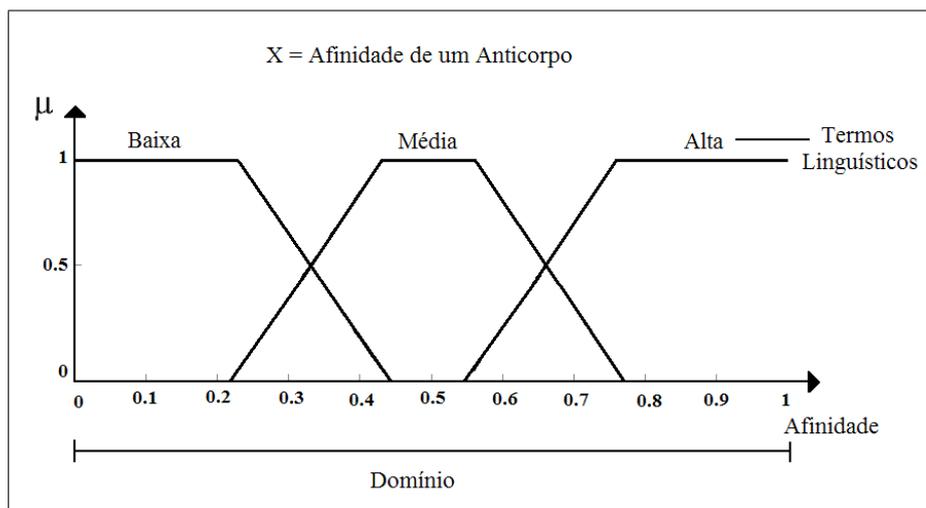


Figura 4.3: Exemplo de variável lingüística.

As variáveis lingüísticas são amplamente utilizadas na construção de regras *fuzzy* do tipo SE-ENTÃO, as quais utilizam diferentes combinações de rótulos nos an-

tecedentes e conseqüentes das regras, conforme explicado a seguir.

### 4.1.3 Base de Regras

As regras *fuzzy* são apropriadas para representar o conhecimento lingüístico/empírico, assim como o modo de raciocínio humano que nos fazemos capazes de concluir sobre ambientes imprecisos e complexos (JANG; SUN; MIZUTANI, 1997), mesmo sem regras e dados exatos. As regras *fuzzy* (ZADEH, 1965) são expressões na forma: SE <antecedente> ENTÃO <conseqüente>. Antecedentes e conseqüentes são proposições do tipo  $Z \text{ é } l$  onde  $Z$  é uma variável lingüística e  $l$  um rótulo da variável lingüística associado a um conjunto *fuzzy*. Como exemplo temos a regra, SE pressão é alta ENTÃO temperatura é baixa, onde  $X = \text{pressão}$ ,  $T(X) = \{\text{baixa, média, alta}\}$  e  $Y = \text{temperatura}$ ,  $T(Y) = \{\text{alta, baixa}\}$ ; e cada rótulo está associado a um conjunto *fuzzy* diferente representado por uma função de pertinência específica.

As regras podem ter múltiplas proposições como antecedente e conseqüente formando regras do tipo: SE  $X_1 \text{ é } A_1 \text{ E } X_2 \text{ é } A_2 \text{ E } \dots \text{ E } X_n \text{ é } A_n$  ENTÃO  $Y_1 \text{ é } B_1 \dots \text{ E } Y_s \text{ é } B_s$ . Um exemplo de regras com múltiplas proposições é “Se pressão é Alta E volume é Pequeno ENTÃO temperatura é Alta”. Quando o antecedente e o conseqüente das regras são formados por múltiplas proposições existe a necessidade dessas proposições serem agregadas. As normas-t são consideradas operadores de agregação válidos para regras unidas por E, cujos representantes mais conhecidos são o mínimo e o produto algébrico. Além da agregação dos antecedentes e conseqüentes é necessária a definição da semântica da regra. A semântica da regra é a função que define uma regra *fuzzy* como uma relação *fuzzy*  $R$  no produto cartesiano  $(\mathbf{X}_1 \times \mathbf{X}_2 \dots \mathbf{X}_n) \times (\mathbf{Y}_1 \dots \mathbf{Y}_n)$  (JANG; SUN; MIZUTANI, 1997). Existem duas classes principais de funções que derivam relações *fuzzy* induzidas por regras: conjunções (normas-t) e disjunções (normas-s) (PEDRYCZ; GOMIDE, 1998). Neste trabalho serão utilizadas conjunções *fuzzy* (semântica dada por uma norma-t).

Uma base de regras fuzzy é formada por um conjunto de regras onde a saída final do sistema deve considerar a influência de cada regra individual. Portanto, após a definição da semântica da regra que irá influenciar na derivação da saída individual da regra, é necessário definir como estas respostas individuais serão agregadas para se obter a saída *fuzzy* final. O operador mais usual para realizar esta tarefa é o *max*,

mas operadores de médias também são utilizados (PEDRYCZ; GOMIDE, 1998).

Existem diversos modelos de sistemas de inferência *fuzzy* descritos na literatura. Os modelos de inferência *fuzzy* diferem basicamente no conseqüente das regras. Os principais modelos são (PEDRYCZ; GOMIDE, 1998):

- Modelo Mamdani: apresenta conjuntos *fuzzy* nos antecedentes e conseqüentes das regras. A saída final, representada por conjunto *fuzzy*, resulta da agregação das saídas das diversas regras. Este modelo demanda processo de defuzzificação para obtenção de saída não *fuzzy*;
- Modelo Takagi-Sugeno: conseqüentes das regras são uma combinação (normalmente linear) das variáveis de entrada mais uma constante, e a saída final é a média ponderada (em função do nível de ativação) da saída de cada regra;

Neste trabalho será empregado o modelo Mamdani, cujo mecanismo de inferência é detalhado a seguir. Esta escolha se deve ao fato deste ser o modelo mais apropriado quando não se dispõe de dados para treinamento e ser um modelo de fácil interpretação.

#### 4.1.4 Mecanismos de Raciocínio

Os sistemas *fuzzy* utilizam uma computação baseada na regra composicional de inferência e no raciocínio *fuzzy* (PEDRYCZ; GOMIDE, 1998). De uma forma geral, tem-se:

$$\begin{array}{ll}
 \text{(fato } P_i\text{):} & X_1 \text{ é } A_1 \text{ E } X_2 \text{ é } A_2 \text{ E } \dots \text{ E } X_n \text{ é } A_n \\
 \text{(regra } R_1\text{):} & \text{Se } X_1 \text{ é } A_1^1 \text{ E } \dots \text{ E } X_n \text{ é } A_n^1 \text{ então } Y_1 \text{ é } B_1^1 \text{ E } \dots \text{ E } Y_s \text{ é } B_s^1 \\
 & \vdots \\
 \text{(regra } R_m\text{):} & \text{Se } X_1 \text{ é } A_1^m \text{ E } \dots \text{ E } X_n \text{ é } A_n^m \text{ então } Y_1 \text{ é } B_1^m \text{ E } \dots \text{ E } Y_s \text{ é } B_s^m \\
 \hline
 \text{(conclusão } P_o\text{):} & Y_1 \text{ é } B_1 \text{ E } Y_2 \text{ é } B_2 \text{ E } \dots \text{ E } Y_s \text{ é } B_s .
 \end{array}$$

A forma como a conclusão é extraída a partir do fato  $P_i$  e das regras  $R_1$  a  $R_m$  define diferentes mecanismos de inferência utilizados por sistemas *fuzzy*. Existem dois mecanismos de inferência principais: inferência composicional e inferência esca-

lonada. O esquema de inferência escalonada simplifica muito a computação em SIFs. Por isso, a maioria das aplicações práticas utiliza SIFs baseados neste mecanismo.

O mecanismo de inferência adotado neste trabalho é o de inferência escalonada, por isso ele será descrito a seguir.

## Inferência Escalonada

Segundo Pedrycz e Gomide (1998), o esquema geral para a inferência escalonada para um sistema *fuzzy* composto por  $m$  regras *fuzzy* ( $R_j$ ,  $j = 1, \dots, m$ ) envolve os seguintes passos:

**Passo 1 Matching:** para cada regra, computar o nível de similaridade (*matching*)  $\mu_{m_k}$ ,  $k = 1, \dots, n$ , entre cada proposição atômica<sup>1</sup> (associada à variável  $k$ ) do antecedente da regra e a proposição atômica correspondente no fato (ou proposição de entrada  $P_i$ ). A operação de *matching* pode ser relacionada com a composição **sup-min**, o que, no caso de uma entrada não *fuzzy*  $u_k$ , resulta no *matching* dado por  $\mu_{m_k} = \mu_{A_k}(u_k)$ .

**Passo 2 Agregação dos Antecedentes:** para cada regra, computar o grau de ativação  $\mu_j$  da regra  $R_j$  da seguinte forma:

$$\mathbf{A}_a = \mu_j = \mathbf{t}_{k=1}^n \mu_{m_k}, j = 1 \dots, m \quad (4.1)$$

**Passo 3 Derivação da conclusão individual  $P_{oj}$ :** para cada regra, computar o valor inferido, baseado no resultado da agregação dos antecedentes  $\mu_j$  e na semântica  $f$  da regra escolhida.

$$P_{oj}(y) = B^{j*}(y) = f(\mu_j, \mu_{Bj}(y)) \quad (4.2)$$

Neste trabalho, a semântica da regra é sempre uma norma-t.

**Passo 4 Derivação do resultado final  $P_o$ :** Computar o valor inferido do conjunto

---

<sup>1</sup>proposições atômicas representam unidades de informação básicas das regras e são dadas por  $X$  é  $A$  ou ainda  $Y$  é  $B$

completo de regras através da agregação dos valores inferidos no passo 3:

$$P_o(y) = B(y) = \mathbf{A}_R (B^{j*}) \quad (4.3)$$

Como exemplo, suponha o seguinte esquema de raciocínio *fuzzy* envolvendo regras *fuzzy* com duas variáveis lingüísticas  $X_1$  e  $X_2$  no antecedente e uma variável lingüística  $Y$  no conseqüente,

$$\begin{array}{ll} \text{(fato } P_i): & X_1 \text{ é } A_1 \text{ E } X_2 \text{ é } A_2 \\ \text{(regra } R_1): & \text{Se } X_1 \text{ é } A_1^1 \text{ E } X_2 \text{ é } A_2^1 \text{ então } Y \text{ é } B^1 \\ & \vdots \\ \text{(regras } R_m): & \text{Se } X_1 \text{ é } A_1^m \text{ E } X_2 \text{ é } A_2^m \text{ então } Y \text{ é } B^m \\ \hline \text{(conclusão } P_o): & Y \text{ é } B. \end{array}$$

A utilização da inferência escalonada na implementação do mecanismo de raciocínio *fuzzy* descrito anteriormente depende da fixação de alguns parâmetros. Suponha, por exemplo:

- agregação de antecedentes  $\mathbf{A}_a$ : produto algébrico;
- semântica das regras  $f$ :  $f = \min$ ;
- a agregação das regras  $\mathbf{A}_R$ : max.

Então:

**Passo 1** *Matching* na regra  $R_j$ :

$$\mu_{m_1}^j = \sup_{x_1} [\mu_{A_1}(x_1) \wedge \mu_{A_1^j}(x_1)];$$

$$\mu_{m_2}^j = \sup_{x_2} [\mu_{A_2}(x_2) \wedge \mu_{A_2^j}(x_2)].$$

**Passo 2** Agregação dos Antecedentes:

$$\mu_j = \mu_{m_1}^j \cdot \mu_{m_2}^j$$

**Passo 3** Derivação individual  $P_{oj}(y)$ :

$$P_{oj}(y) = B^{j*}(y) = f_p(\mu_j, \mu_{B^j}(y)) = \min(\mu_j, \mu_{B^j}(y)).$$

**Passo 4** Conclusão final

$$P_o(y) = B(y) = \mathbf{A}_R (B^{j*}) = \max [B^{j1*}, \dots, B^{m1*}] .$$

**Inferência Escalonada no Modelo Mamdani**

A característica básica do modelo de Mamdani é o fato de utilizar conjuntos *fuzzy* nos conseqüentes das regras *fuzzy*. As etapas de cálculo de *matching* e agregação dos antecedentes seguem os passos normais da inferência escalonada. A obtenção da conclusão de cada regra depende da semântica escolhida e dos resultados do *matching* e agregação dos antecedentes (DELGADO, 2002). Da agregação das conclusões inferidas de cada regra, resulta um conjunto *fuzzy*. Nos casos em que uma saída não-*fuzzy* é necessária, utiliza-se um dos métodos de defuzzificação (LEE, 1990a; LEE, 1990b). O método utilizado neste trabalho foi o centro de gravidade que é dado por:

$$\text{Saída} = \frac{\int_y \mu_B(y) y dy}{\int_y \mu_B(y) dy} \quad (4.4)$$

onde  $B$  é um conjunto *fuzzy* definido no universo  $\mathbf{Y}$ .

Suponha um sistema *fuzzy* do tipo Mamdani, composto por duas regras *fuzzy* na forma

$$R_j : \text{ Se } X_1 \text{ é } A_1^j \text{ E } X_2 \text{ é } A_2^j \text{ então } Y \text{ é } B^j, \quad j = 1, 2.$$

onde  $X_1$  e  $X_2$  são variáveis lingüísticas e  $A_1^j$  e  $A_2^j$  são conjuntos *fuzzy* nos universos  $\mathbf{X}_1$  e  $\mathbf{X}_2$ , respectivamente. A Figura 4.4 ilustra o processo de inferência (escalonada) para duas entradas não-*fuzzy* (representadas por valores em posições arbitrárias  $u_1$  e  $u_2$ ). Neste exemplo, a agregação dos antecedentes é dada pelo *min*, a semântica das regras é fixada como  $f = f_m = \min$  e a união das saídas de cada regra é computada pelo operador *max*.

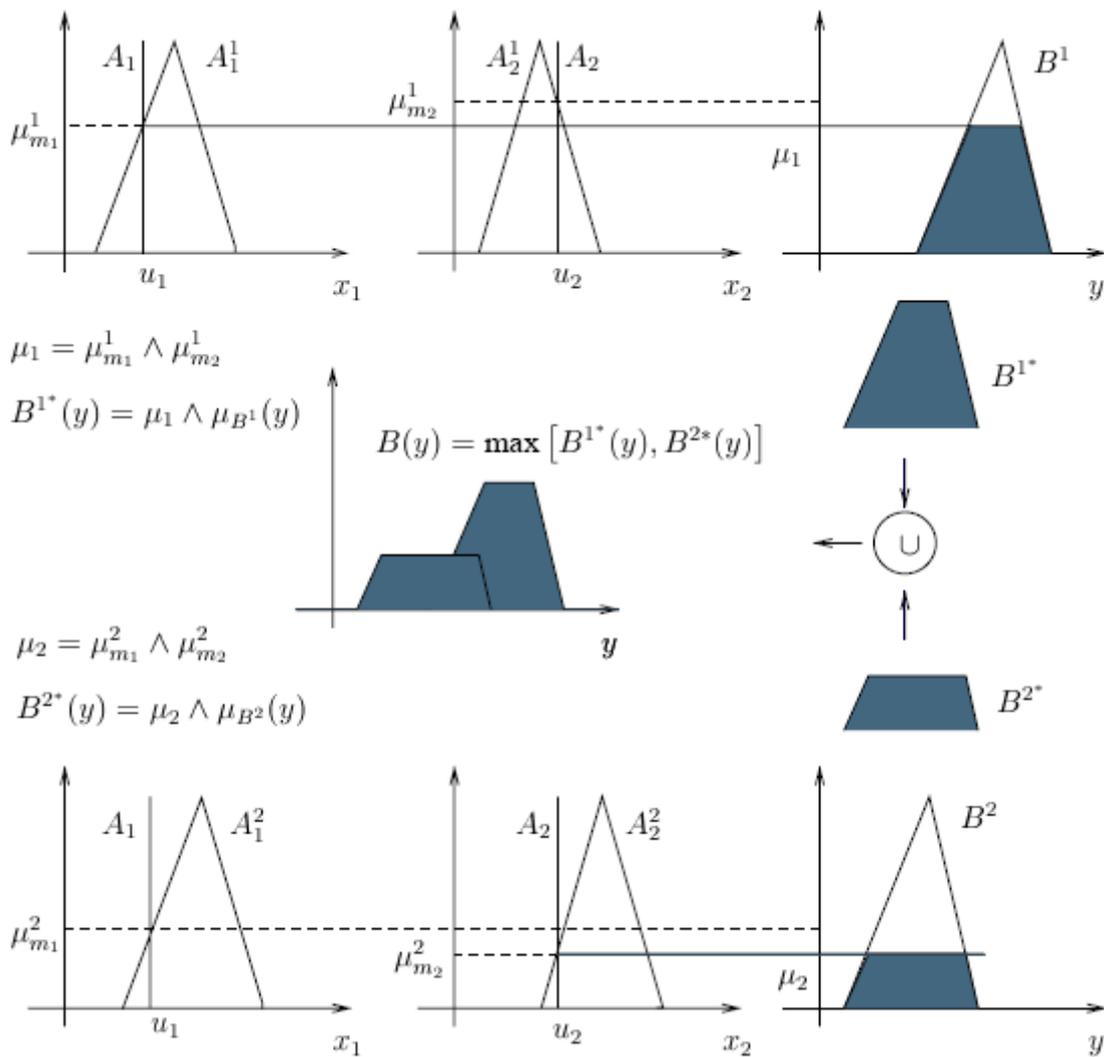


Figura 4.4: Inferência no modelo Mamdani.((DELGADO, 2002))

## 4.2 Busca Tabu

A Busca Tabu é uma metaheurística que visa guiar um procedimento de busca local para a geração de soluções ótimas ou muito próximas do ótimo. Ela, geralmente, é aplicada no contexto de otimização combinatorial e foi usada com sucesso em diversos problemas como telecomunicações, roteamento, alocação, *scheduling*, entre outros (GLOVER, 1995).

Um aspecto importante da BT é o uso sistemático de memória adaptativa ao longo do processo de busca (HERTZ; TAILLARD; WERRA, 1995). A memória adaptativa permite que a Busca Tabu explore de maneira eficiente e econômica o espaço de busca. O uso de memória adaptativa distingue a BT de técnicas estocásticas como o *Simulated Annealing* e os Algoritmos Genéticos e de técnicas que fazem uso de estruturas rígidas de memória como os métodos de *branch-and-bound* (GLOVER; LAGUNA, 1993).

Os conceitos de vizinhança, memória adaptativa e critério de aspiração são fundamentais para o entendimento do funcionamento da Busca Tabu e são explanados nas Subseções 4.2.1, 4.2.2 e 4.2.3, respectivamente. Já a Subseção 4.2.4 apresenta o algoritmo básico que descreve o funcionamento de uma BT.

### 4.2.1 Vizinhança

A vizinhança de uma solução é o conjunto de todas as soluções que podem ser obtidas por movimentos (transformações locais) aplicados a esta solução. Formalmente, tem-se:

$$Vz(x) = \{y \mid y \text{ é uma nova solução resultante de um movimento aplicado a } x\} \quad (4.5)$$

Os movimentos que podem ser aplicados durante a transformação de uma solução para gerar a vizinhança é o fator que define a estrutura da vizinhança. A estrutura de uma vizinhança pode ser simples (movimentos simples) ou complexa (movimentos complexos). Geralmente, estruturas complexas permitem uma melhor exploração do espaço de busca. O espaço de busca utilizado pode restringir os tipos de movimentos que podem ser utilizados. Por exemplo, pode-se definir que movimentos que resultem

em soluções infactíveis não são movimentos válidos. Outro fator importante diz respeito à representação das soluções, certos tipos de movimentos só fazem sentido em certas representações.

Segundo Gendreau (2002) a definição da vizinhança utilizada é um aspecto importante do projeto de uma Busca Tabu eficiente.

### 4.2.2 Memória Adaptativa

O uso de memória adaptativa é uma característica marcante da Busca Tabu. Existem diversas classificações para os tipos de memórias utilizadas em uma Busca Tabu. Elas podem ser de curto, médio ou longo prazo, de fatos recentes, de frequência, de eventos críticos, entre outros (GLOVER; LAGUNA, 1993)(GLOVER, 1995).

A memória de curto prazo é mais comumente empregada. A lista tabu é um tipo de memória de curto prazo. Seu objetivo é evitar a repetição de soluções previamente encontradas (soluções tabu). O tamanho da lista tabu, também chamado de *tenure*, determina quanto tempo um movimento ou solução deve permanecer como tabu. É mais comum armazenarem-se movimentos ao invés de soluções, visto que os primeiros são menores e costumam permitir mecanismos de busca mais eficientes na lista.

A memória de médio prazo comumente armazena soluções de elite (ou boas soluções) e é utilizada durante estágios de reinício da busca. Um reinício da busca é como se fosse uma nova busca, mas as memórias não costumam ser reiniciadas. O reinício permite explorar boas regiões do espaço de busca que haviam sido preteridas em favor de outras que se acreditava serem melhores. Porém, essas regiões favorecidas podem não levar ao ótimo, o que instiga a exploração de outras regiões do espaço de busca.

Já a memória de longo prazo armazena estatística sobre a frequência de certos componentes nas soluções. Ela pode ser utilizada para gerar diversificação (evitando gerar componentes que aparecem frequentemente) ou para intensificar a busca (fixando os componentes mais frequentes na solução e permitindo apenas movimentos que não afetem estes componentes).

O uso eficiente das memórias é importante para a qualidade das soluções obtidas pela Busca Tabu.

### 4.2.3 Critérios de Aspiração

O uso de uma lista tabu pode fazer com que uma boa solução (até mesmo a melhor solução) seja desconsiderada da vizinhança de uma solução por se tratar uma solução tabu. Contudo, a escolha destas boas soluções poderia melhorar os resultados encontrados pelo método. Para evitar que tais boas soluções sejam ignoradas, foram introduzidos os critérios de aspiração. Um critério de aspiração é uma condição que deve ser atingida para que uma solução tabu faça parte das soluções consideradas pela busca.

Um critério de aspiração bastante utilizado é o de permitir que uma solução que seja melhor do que a melhor solução encontrada até o momento seja considerada. Este critério, em particular, só se aplica quando a lista tabu armazena movimentos e não soluções completas. Outros critérios de aspiração podem envolver características como diversidade, qualidade das soluções e informações mantidas nas memórias de médio e longo prazos.

### 4.2.4 Algoritmo da Busca Tabu

O pseudo-código da Busca Tabu é apresentado no Algoritmo 4.

---

#### Algoritmo 4 Busca Tabu

---

```

Escolha uma solução inicial  $x \in S$ ;
 $fk = 0$ ;
 $x^* = x$ ;
Enquanto (!Critério_de_Parada) faça
     $Vz'(x) = Vz(x) - T + A$ ;
     $x = \text{Melhor}(Vz'(x))$ ;
    Se ( $x$  é melhor que  $x^*$ ) então
         $x^* = x$ ;
    Fim Se
    Atualizar_Lista(T);
     $k = k + 1$ ;
Fim Enquanto

```

---

O algoritmo começa com a escolha de uma solução inicial ( $x$ ), a qual é considerada como sendo a melhor solução encontrada até o momento ( $x^*$ ) e a iteração atual é inicializada em zero. Uma lista tabu ( $T$ ) também é criada, mas ainda não contém nenhum movimento ou indivíduo tabu. Após isto, são analisadas todas as soluções na vizinhança de  $x$  ( $Vz(x)$ ). Deste conjunto de possíveis soluções são eliminadas aquelas que são consideradas tabu (ou por estarem presentes em  $T$  ou por serem geradas através de um movimento armazenado em  $T$ ) e adicionadas aquelas que são tabu mas atendem a um critério de aspiração (conjunto  $A$ ), formando o conjunto  $Vz'(x)$  de possíveis soluções a serem consideradas. Deste conjunto é escolhida a melhor solução, a qual passa a ser considerada a solução atual ( $x$ ). Se esta solução for melhor do que a melhor solução encontrada até o momento ( $x^*$ ) ela é considerada como sendo o novo valor de  $x^*$ . De qualquer maneira a lista tabu é atualizada para refletir a escolha da nova solução atual. Esta atualização é feita através da inserção desta solução ( $x$ ) em  $T$  ou através da inserção do movimento que deu origem a  $x$ . A iteração atual é incrementada e o processo se repete até que o critério de parada seja atingido.

---

## *Metodologia Proposta*

---

---

Neste trabalho foram desenvolvidos diversos modelos para a Predição da Estrutura de Proteínas baseados no algoritmo Clonalg e no algoritmo da Rede Imune. O desenvolvimento destes modelos tem o propósito de responder às seguintes questões:

- Qual a influência do uso de um sistema de inferência *fuzzy* no operador de *aging*?
- Modelos que trabalham com espaço de soluções inactíveis são comparáveis àqueles que trabalham apenas com anticorpos factíveis?
- Qual o impacto do uso da metaheurística Busca Tabu na melhoria da afinidade dos anticorpos?
- O modelo final é capaz de apresentar melhores resultados do que aqueles encontrados na literatura?

Todos os modelos propostos implementam dois tipos de entidades: antígenos e anticorpos. Neste trabalho, os antígenos são representados pelas seqüências das proteínas a serem dobradas e os anticorpos representam possíveis dobramentos para

estas seqüências. A representação dos antígenos é feita por vetores booleanos onde o valor verdadeiro (V) representa um resíduo hidrofóbico (H) e o valor falso (F) representa um resíduo polar (P). Os anticorpos são representados por um caminho no *lattice* cúbico. A Figura 5.1 ilustra um exemplo de anticorpo produzido pela metodologia proposta. Neste exemplo, a seqüência de direções {B, E, B, E, A, E, A, D, T, D, F, F, B, E, B, E, A, A, D}<sup>1</sup> foi obtida na presença de um antígeno contendo 20 resíduos, cabe salientar que existem apenas 19 posições pois a primeira posição é fixa. Este antígeno, na representação booleana, representa o seguinte vetor {F, F, F, F, F, F, F, F, V, F, F, V, V, F, V, F, F, F, F, V}. Observa-se na Figura 5.1 que a solução é factível (a estrutura não contém colisões) e que o nível de energia obtido foi  $\eta = -4$ , ou seja, há quatro ligações hidrofóbicas não locais identificadas pelas linhas verdes.

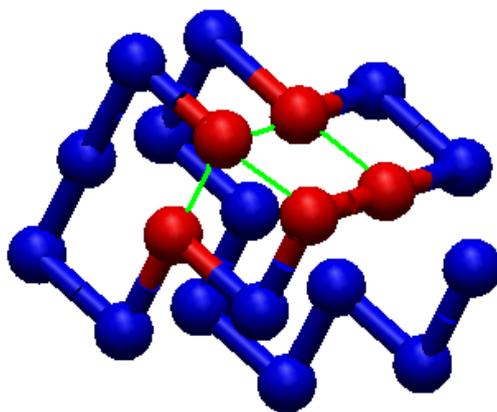


Figura 5.1: Exemplo de estrutura predita pelos modelos propostos.

A Seção 5.1 apresenta a metodologia adotada no desenvolvimento dos modelos baseados no algoritmo Clonalg e a Seção 5.2 apresenta a metodologia adotada no desenvolvimento dos modelos baseados na Rede Imunológica. Detalhes sobre cada modelo implementado podem ser encontrados no Capítulo 6.

Os algoritmos foram codificados na linguagem Java 2 Standard Edition (J2SE) utilizando a versão 5.0 no ambiente de desenvolvimento NetBeans na versão 5.5.

<sup>1</sup>Lembrando que B = Abaixo, A = Acima, E = Esquerda, D = Direita, F = Frente e T = Atrás

## 5.1 Clonalgs

Os modelos baseados no Clonalg implementados neste trabalho adotam a forma geral apresentada no Algoritmo 5. É importante salientar que, por ser geral, o algoritmo apresenta alguns estágios que não são utilizados em determinados modelos (para detalhes de cada modelo ver a Subseção 6.1.1).

---

**Algoritmo 5** Algoritmo Clonalg adotado

---

```

Clonalg(TamPOP, FatorDUP, IdadeMAX, TaxaHIP, MI)
  gerações ← 0;
  POP = Inicialização(TamPOP);
  Avalia(POP);
  Enquanto(!Condição_de_Parada) faça
    POPc ← Clonagem(POP, FatorDUP);
    POPh ← Hipermutação(POPc, TaxaHIP);
    POPm ← Hipermacromutação(POPc);
    Avalia(POPh);
    Avalia(POPm);
    OperadorAging(POP, POPh, POPm, IdadeMAX, POPMorte, POPa);
    POPa ← POPa ∪ MaturaçãoFracca(POPMorte);
    POP ← Seleção(POPa);
    gerações ← gerações+1;
    Se Num_Selecionados < TamPOP então
      Gerar novos anticorpos para completar a população;
    Fim Se
    Se ((Num_Avaliações / MI)==0) então
      POP ← MaturaçãoIntensiva(POP);
    Fim Se
  Fim Enquanto

```

---

A seguir será dada uma descrição geral do comportamento do algoritmo Clonalg. Os principais estágios do algoritmo serão melhor detalhados em seções posteriores.

### 5.1.1 Descrição Geral do Clonalg

A geração da população inicial é feita aleatoriamente (com distribuição uniforme). Existem modelos baseados neste algoritmo que trabalham apenas com o espaço de soluções factíveis e outros que permitem soluções infactíveis aplicando penalidades a elas no processo de avaliação. Após a inicialização da população, cada anticorpo é avaliado pela função `Avalia()` que recebe como parâmetro uma população de anticorpos. A Afinidade de um anticorpo é medida pelo o número de con-

tatos hidrofóbicos não-locais (detalhado na Seção 2.1) e pelo seu número de colisões (caso ele seja inactivo), o que transforma o problema de minimização de energia em um problema de maximização do número de contatos hidrofóbicos não-locais e minimização das colisões. A função de avaliação, portanto, calcula o número de contatos/Ligações Hidrofóbicas Não-Locais (LHNL) e o número de colisões de cada anticorpo da população (colisões). Um anticorpo activo tem por valor de afinidade o número de ligações hidrofóbicas não-locais, quanto mais ligações melhor é o anticorpo. Já um anticorpo inactivo é penalizado. A penalidade corresponde a ignorar todas as suas ligações hidrofóbicas não-locais ( $LHNL = 0$ ) e subtrair deste valor o número de colisões que este anticorpo possui. A afinidade é dada pela Equação 5.1:

$$\begin{aligned} \textit{Afinidade} &= LHNL, & \text{se colisões} &= 0 \\ \textit{Afinidade} &= -col, & \text{caso contrário.} & \end{aligned} \quad (5.1)$$

Onde *Afinidade* é a afinidade associada ao anticorpo, *LHNL* é o número de ligações hidrofóbicas não-locais do anticorpo e *col* é o número de colisões que o anticorpo tem quando representado no *lattice*.

Este método de penalidade favorece os anticorpos activos com várias ligações hidrofóbicas e anticorpos inactivos com poucas colisões, ou seja, que com poucas mutações podem ser transformados em activos. No caso de anticorpos com colisões, o número de ligações hidrofóbicas não-locais é desconsiderado porque a maioria dos *LHNLs* serão eliminados para que as colisões desapareçam.

A *Condição\_de\_Parada* é uma variável verdadeira sempre que o processo evolucionário termina. Neste trabalho, a condição de parada utilizada em todos os modelos do Clonalgs foi um número máximo de avaliações.

O operador de *Clonagem()* produz algumas cópias (clones) de cada anticorpo. Este operador gera uma população intermediária ( $POP_c$ ) de tamanho igual a  $Tam_{POP} * Fator_{DUP}$ , onde  $Tam_{POP}$  é o tamanho da população inicial e  $Fator_{DUP}$  é o parâmetro que define o número de cópias de cada anticorpo. Durante a expansão clonal, os clones herdaram a idade de seus pais.

A função *Hipermutação()* recebe dois parâmetros - a população de clones e a taxa de hipermutação ( $Taxa_{HIP}$ ) - e retorna uma população intermediária ( $POP_h$ ). A função *Hipermacromutação()* recebe apenas a população de clones e também

retorna uma população intermediária ( $POP_m$ ). Os operadores de Hipermutação e de Hipermacromutação são explanados na Seção 5.1.2.

É importante salientar que a Hipermutação e a Hipermacromutação não geram anticorpos inactíveis (nem mesmo quando a população inicial permite anticorpos com infactibilidade) e nem permitem a existência de anticorpos iguais na população.

Uma importante contribuição deste trabalho consiste na análise do operador de *aging*, o qual é usado para decidir quais anticorpos serão eliminados da população antes da etapa de seleção e antende, entre outros, ao critério de idade. Dois tipos de funções de `Aging()` serão testadas e os detalhes de cada função são fornecidos na Subseção 5.1.4.

A população  $POP_{morte}$  é formada pelos anticorpos que são sentenciados à morte pelo operador de `Aging()` enquanto que  $POP_a$  é formada pelos anticorpos que sobreviveram ao operador de `Aging()`.

Antes da seleção da população que irá fazer parte da próxima geração, um processo de maturação da afinidade fraco é aplicado à  $POP_{morte}$  como a última tentativa de melhorar a afinidade dessa população. Os anticorpos resultantes desta maturação são adicionados a  $POP_a$ . Esse estágio de maturação é explicado na subseção 5.1.3.

A função de `Seleção()` recebe como parâmetro a população de clones ( $POP_a$ ) e retorna os  $Tam_{POP}$  melhores anticorpos para compor a população da próxima geração.

Se forem selecionados menos anticorpos que  $Tam_{POP}$ , novos anticorpos são gerados aleatoriamente para completar a nova população (da mesma forma como foi gerada a população inicial, ou seja, podendo gerar anticorpos inactíveis no caso do método de avaliação baseado em penalidades).

Finalmente, a cada  $MI$  avaliações o processo de maturação intensiva é aplicado. Este estágio de maturação é baseado na Busca Tabu e é aplicado a toda a população, sendo melhor descrito na Subseção 5.1.3.

Neste ponto cabe destacar as principais diferenças entre o modelo descrito e o algoritmo Clonalg tradicional:

- Operador de *aging*;
- Operador de *hipermacromutação*;
- Maturação fraca da afinidade;
- Maturação forte da afinidade;
- Critério de inserção de novos anticorpos.

## 5.1.2 Operadores de Mutação

Durante o processo de evolução dois tipos de mutação são aplicados: a hipermutação (inversamente proporcional à afinidade) e a Hipermacromutação (independente da afinidade do anticorpo).

### Hipermutação

O operador de hipermutação somática muta um anticorpo de forma inversamente proporcional à sua afinidade, ou seja, quanto maior a afinidade de um anticorpo, menor é o número de posições desse anticorpo que sofrerão alteração e vice-versa. O número máximo de mutações que um anticorpo irá sofrer é calculado de acordo com a Equação 5.2.

$$M_{max}(x) = \begin{cases} (1 + \frac{E^*}{LHNL(x)}) \cdot \alpha, & \text{se número colisões} = 0 \text{ e } LHNL(x) > 0 \\ (1 + E^*) \cdot \alpha + \alpha, & \text{caso contrário} \end{cases} \quad (5.2)$$

onde  $\alpha = \text{Taxa}_{HIP} * \text{Tam}_{PROT}$  identifica o produto da taxa de hipermutação pelo tamanho da proteína,  $LHNL(x)$  é o número de ligações hidrofóbicas não-locais de  $x$  e  $E^*$  corresponde ao ótimo reportado na literatura.

O Algoritmo 6 apresenta os passos do processo de Hipermutação somática.

Verifica-se que o número efetivo de mutações ( $M$ ) realizadas é limitado a  $M_{max}$ ,

---

**Algoritmo 6** Hipermutação

---

**Para** (Cada anticorpo da População) **faça**  
 Calcule  $M_{max}$ ; //de acordo com a equação 5.2  
 Num\_Mutações = 0;  
**Enquanto**(Num\_Mutações  $\leq$   $M_{max}$  & PMC = falso) **faça**  
 Selecionar aleatoriamente uma posição;  
 Substituir o valor dessa posição por outro gerado aleatoriamente;  
 Num\_Mutações = Num\_Mutações + 1;  
**Fim Enquanto**  
**Fim Para**

---

mas depende efetivamente da geração de um anticorpo factível ou não, conforme detalhado na definição da Primeira Mutação Construtiva (PMC).

## Hipermacromutação

O operador de Hipermacromutação aplicado em todos os modelos desenvolvidos é o mesmo utilizado em (CUTELLO et al., 2005). Este operador tenta mutar cada anticorpo  $M$  vezes sempre gerando anticorpos factíveis. O número máximo de mutações é  $M_{max} = j - i + 1$ ,  $1 \leq i, j \leq Tam_{PROT}$ , onde  $i$  e  $j$  são dois inteiros aleatórios tal que  $(i + 1) \leq j \leq Tam_{PROT}$ . O número de mutações independe da afinidade do anticorpo sendo hipermacromutado. Este operador seleciona aleatoriamente a direção da perturbação, tanto da posição  $i$  para a posição  $j$  (da esquerda para a direita) quanto da posição  $j$  para a posição  $i$  (da direita para esquerda). O Algoritmo 7 contém os passos deste operador.

---

**Algoritmo 7** Hipermacromutação

---

**Para** (Cada anticorpo da População) **faça**  
 Determine o valor de  $i$ ;  
 Determine o valor de  $j$ ;  
**Se**(Direção é de  $i$  para  $j$ )**então**  
**For** ( $k = i$ ;  $k \leq j$  & PMC = falso;  $k++$ )  
 Substitua o valor da posição  $k$  por um gerado aleatoriamente;  
**Fim For**  
**Senão**  
**For** ( $k = j$ ;  $k \geq i$  & PMC = falso;  $k--$ )  
 Substitua o valor da posição  $k$  por um gerado aleatoriamente;  
**Fim For**  
**Fim Se**  
**Fim Para**

---

## Primeira Mutação Construtiva

Para evitar a convergência prematura e melhor explorar o espaço de busca, foi aplicada a chamada Primeira Mutação Construtiva (PMC), descrita por Cutello et al. (2005). Este mecanismo é aplicado durante as fases de Hipermutação e Hiper-macromutação.

Seu funcionamento é o seguinte: se uma mutação resulta em um anticorpo factível (PMC = verdadeiro) os processos de Hipermutação e Hiper-macromutação movem-se para o próximo anticorpo. Portanto, o número efetivo de mutações que um anticorpo sofre é igual a  $M$ , onde  $1 \leq M \leq M_{max}$  e  $M$  corresponde ao total de mutações realizadas até que se encontre um anticorpo factível. Se após  $M_{max}$  mutações nenhum anticorpo factível é encontrado o anticorpo sendo mutado é descartado. PMC é igual a falso enquanto um anticorpo factível não for encontrado.

### 5.1.3 Estágios de Maturação da Afinidade

Os dois estágios de maturação da afinidade aplicados são detalhados a seguir. O estágio de maturação fraca possui passos semelhantes à Hipermutação e à Hiper-macromutação e é executado após o operador de *Aging*. O estágio de maturação intensiva é baseado na técnica de Busca Tabu e é utilizado periodicamente sempre que a função de avaliação é executada  $MI$  vezes.

Estes dois estágios realizam um processo de intensificação da busca em contraste com a natureza exploratória dos operadores de Hipermutação e Hiper-macromutação.

#### Maturação Fraca da Afinidade

Este processo de maturação é aplicado para tratar os anticorpos cuja morte é iminente ( $POP_{morte}$ ). Os passos da maturação fraca da afinidade são apresentados no Algoritmo 8. Primeiramente, a maturação fraca da afinidade tenta melhorar a afinidade de um anticorpo por um processo análogo à Hipermutação mas com algumas pequenas diferenças: depois que a posição a ser mutada é definida o processo de maturação fraca testa todas as possíveis direções (outros movimentos no *lattice*) para esta posição antes de selecionar uma outra posição de mutação.

---

**Algoritmo 8** Maturação Fraca da Afinidade
 

---

```

Para (Cada anticorpo de POPmorte) faça
  //Aplicar variação de Hipermutação
  Calcule Mmax;
  Posições_Mutadas = 0;
  Enquanto(Posições_Mutadas <= Mmax & (Infactível || !Melhorou)) faça
    Selecionar aleatoriamente uma posição;
    Direções_Testadas = 0;
    Enquanto (Direções_Testadas < Num_Direções-1 & Infactível &
    !Melhorou) faça
      Substituir o valor dessa posição por outro gerado aleatoriamente;
      Direções_Testadas = Direções_Testadas + 1;
    Fim Enquanto
    Posições_Mutadas = Posições_Mutadas + 1;
  Fim Enquanto
  Se (!Melhorou) então
    //Aplicar variação da Hipermacromutação
    Determine o valor de i e j;
    Se(Direção é de i para j)então
      For (k = i; k <= j & (Infactível || !Melhorou); k++)
        Direções_Testadas = 0;
        Enquanto (Direções_Testadas < Num_Direções-1 & Infactível
        & !Melhorou) faça
          Substitua o valor da posição k por um gerado
          aleatoriamente;
          Direções_Testadas = Direções_Testadas + 1;
        Fim Enquanto
      Fim For
    Senão
      For (k = j; k => i & (Infactível || !Melhorou); k--)
        Direções_Testadas = 0;
        Enquanto (Direções_Testadas < Num_Direções-1 & Infactível
        & !Melhorou) faça
          Substitua o valor da posição k por um gerado
          aleatoriamente;
          Direções_Testadas = Direções_Testadas + 1;
        Fim Enquanto
      Fim For
    Fim Se
  Fim Se
  Se (!Melhorou) então
    Descartar o anticorpo;
  Fim Se
Fim Para

```

---

Além disso, o processo só termina quando a afinidade do anticorpo é melhor que a original (valor medido antes de mutar a primeira posição) ou o número máximo de tentativas é alcançado.

Caso esta primeira etapa da maturação fraca da afinidade não seja capaz de melhorar a afinidade do anticorpo, ocorre a aplicação de uma etapa semelhante à Hipermacromutação. Nesta fase todas as possíveis direções para todas as posições do intervalo  $[i, j]$  (gerado aleatoriamente) são testadas e o processo pára quando a afinidade do anticorpo é melhorada ou após todas as posições serem modificadas.

Se após a aplicação dessas duas etapas da maturação fraca da afinidade o anticorpo não tiver sua afinidade melhorada ele é descartado da população (ou seja, não é inserido na população  $POP_a$ ).

Cabe destacar que este algoritmo não faz uso da primeira mutação construtiva, sendo o número efetivo de mutações ( $M$ ) igual ao total de mutações realizadas até que um anticorpo com afinidade melhor que a afinidade do anticorpo original seja encontrado. Então, o valor de  $M$  é limitado ao intervalo  $1 \leq M \leq ((M_{max} * (\text{Num\_Direções} - 1)) + ((j - i + 1) * (\text{Num\_Direções} - 1)))$ , onde a primeira parcela da soma corresponde ao número máximo de mutações realizadas na etapa semelhante à hipermutação e a segunda parcela corresponde ao número máximo de mutações realizadas na etapa semelhante à hipermacromutação.

## Maturação Intensiva da Afinidade

O estágio de maturação intensiva foi proposto para melhorar a qualidade dos anticorpos gerados pelo Sistema Imunológico Artificial.

Neste trabalho, o processo de maturação da afinidade intensivo é implementado através da metaheurística Busca Tabu. Essa escolha foi motivada pelos resultados obtidos em (BLAZEWICZ; LUKASIAK; MILOSTAN, 2005). Para a implementação desta maturação foi utilizado o *framework* OpenTS disponível em <http://www.coin-or.org/OpenTS/index.html>.

Para a utilização do OpenTS é necessário definir os seguintes parâmetros:

- **Vizinhança de uma Solução:** uma solução é vizinha de outra se é possível

transformar a solução original na outra solução através de uma única mutação. Essa vizinhança foi escolhida por ser bastante suave. A Figura 5.2 ilustra a vizinhança utilizada;

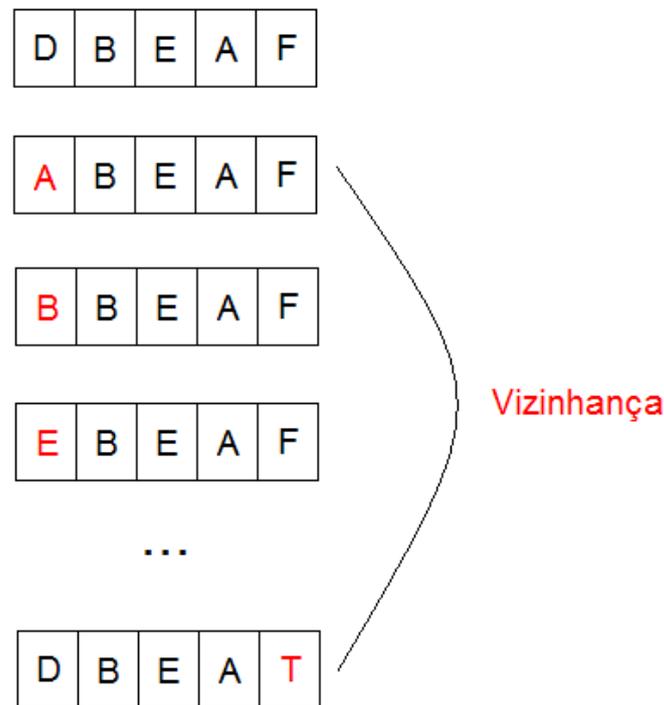


Figura 5.2: Exemplo de vizinhança para uma proteína com 6 resíduos.

- **Lista Tabu:** armazena os movimentos que são considerados tabu. Os movimentos são representados pelo par ordenado posição-valor, onde posição é a posição que foi modificada e valor é o valor que foi atribuído após a modificação. A lista tabu é ilustrada na Figura 5.3;

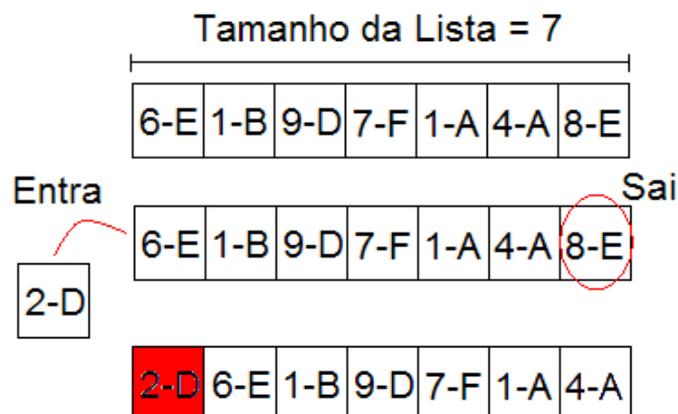


Figura 5.3: Exemplo de uma lista tabu de tamanho 7.

- **Tamanho da Lista Tabu:** foi definido como sendo 20. Isso faz com que a busca tabu seja a maior responsável por encontrar a energia ótima para proteínas pequenas e evita a perda de diversidade nas proteínas grandes;
- **Critério de Aspiração:** uma solução decorrente de um movimento tabu somente é aceita se ela for a melhor solução encontrada até o momento;
- **Função Objetivo:** é a mesma utilizada na etapa de seleção do Clonalg;
- **Solução:** é um anticorpo.

O processo de maturação intensiva é descrito no Algoritmo 9, de acordo com o *framework*.

---

**Algoritmo 9** Maturação Intensiva da Afinidade

---

Definir a Função Objetivo;  
Criar a Lista Tabu;  
Definir os Movimentos Válidos;  
Definir o Número de Iterações;  
Definir o Critério de Aspiração;  
**Para** (Cada anticorpo de POP) **faça**  
    Executar a Busca Tabu; //conforme definido no Algoritmo 4  
**Fim Para**

---

### 5.1.4 Operadores de *Aging*

Os operadores de *Aging* foram desenvolvidos para evitar a convergência prematura do algoritmo e melhorar a diversidade da população. Em sua versão original (CUTELLO et al., 2005), o operador de *Aging* tinha a inspiração biológica de que um anticorpo possui um ciclo de vida limitado. Nesta idéia original todos os anticorpos possuíam o mesmo ciclo de vida, ou seja, todos morriam (eram eliminados da população) após atingirem uma idade máxima (que era igual para todos os anticorpos).

Neste trabalho essa idéia é aprimorada levando-se em consideração também a importância de um anticorpo para a manutenção da saúde de um organismo (capacidade de reconhecimento de antígenos - afinidade) e o quanto ele contribui para a diversidade do repertório de anticorpos de um organismo. Esta idéia é mais próxima do que ocorre na natureza, onde cada anticorpo possui um ciclo de vida próprio.

Além disso, aplicação do operador de *Aging* gera uma sub-população que deve ser submetida a uma etapa de maturação fraca da afinidade na tentativa de melhorar a afinidade dos anticorpos cuja morte é iminente.

Neste trabalho foram testados dois tipos de operadores de *Aging*: um operador de *Aging* Puro (*Aging*) e um operador de *Aging Fuzzy* (*Aging Fuzzy*), ambos contendo a etapa de maturação fraca da afinidade.

Os operadores de *Aging* influenciam os demais da seguinte forma:

- Operador de Expansão Clonal: todo Clone herda a idade de seu pai;
- Operadores de Mutação: depois das fases de Hipermutação e Hipermacromutação, um anticorpo tem sua idade inicializada se o anticorpo mutado tem afinidade melhor que a afinidade de seu pai;
- Operador de Seleção: se menos anticorpos que o tamanho da população sobreviveram ao operador de *Aging*, novos anticorpos são gerados aleatoriamente para completar a população.

## Operador de *Aging* Puro

No operador de *Aging* Puro o número máximo de gerações que um anticorpo pode permanecer na população é determinado pelo parâmetro  $Idade_{MAX}$  do Sistema Imunológico Artificial. Assim que um anticorpo atinge o total de  $Idade_{MAX}$  gerações sua morte é considerada iminente. Quando isto acontece um processo de maturação fraca da afinidade é aplicado a este anticorpo. Se este processo é capaz de melhorar sua afinidade ele permanece na população senão ele morre.

O operador de *Aging* Puro adotado preserva o princípio da “igualdade de oportunidades” do operador de *Aging* original descrito em (CUTELLO et al., 2005), o qual impõe uma igual oportunidade de busca para todos os anticorpos da população sem fazer uso de qualquer mecanismo de busca local. Esse operador foi aqui nomeado Puro porque depende puramente da idade de um anticorpo.

## Operador de *Aging Fuzzy*

Apesar da igualdade de oportunidades ser relevante, no operador de *Aging Fuzzy* optou-se por privilegiar anticorpos jovens com boa afinidade e grande diversidade (assim como ocorre na natureza e em concordância com o princípio da sobrevivência do mais apto). Com isso, a determinação da iminência de morte de um anticorpo leva em consideração além da idade, a afinidade e a diversidade. A diversidade mede a distância entre um anticorpo e os demais da população. A medida de distância adota neste trabalho foi a de Hamming.

Um Sistema de Inferência Fuzzy (SIF) do tipo Mamdani (PEDRYCZ; GOMIDE, 1998) foi utilizado para determinar a possibilidade de morte de um anticorpo. Para a implementação do SIF em java foi utilizado um pacote que pode ser encontrado em: [http://people.clarkson.edu/~esazonov/neural\\_fuzzy/loadsway/LoadSway.htm](http://people.clarkson.edu/~esazonov/neural_fuzzy/loadsway/LoadSway.htm).

Este sistema considera três variáveis de entrada: Idade, Afinidade e Diversidade; e uma variável de saída: Possibilidade de Morte. Na base de dados adotada, as variáveis lingüísticas têm os seguintes termos lingüísticos ou rótulos:

- $T(\text{Idade}) = \{\text{Jovem}, \text{Adulto}, \text{Idoso}\};$
- $T(\text{Afinidade}) = \{\text{Baixa}, \text{Média}, \text{Alta}\};$
- $T(\text{Diversidade}) = \{\text{Baixa}, \text{Média}, \text{Alta}\};$
- $T(\text{Poss. de Morte}) = \{\text{Muito Baixa}, \text{Baixa}, \text{Média}, \text{Alta}, \text{Muito Alta}\}$

Os conjuntos fuzzy são ilustrados na Figura 5.4. As variáveis diversidade e afinidade são normalizadas de forma dinâmica no intervalo  $[0,1]$ , ou seja, esses valores são atualizados a cada geração do algoritmo.

A base de regras do SIF pode ser vista na Tabela 5.1. A construção desta base de regras deriva de três princípios:

- **Elitismo:** é importante manter as melhores soluções na população com o objetivo de melhorá-las ainda mais. Contudo essas soluções ainda devem poder ser eliminadas caso elas estejam estagnadas (correspondem a máximos locais) ou sejam redundantes no repertório de anticorpos;

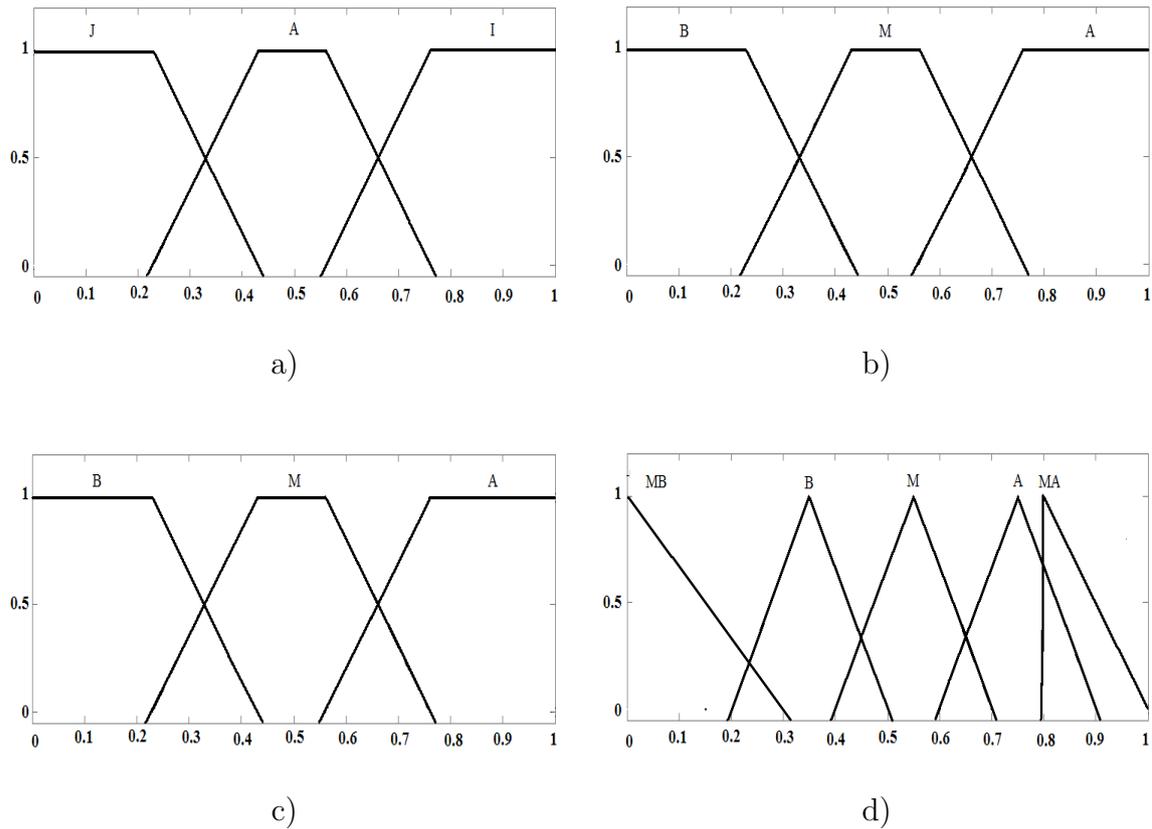


Figura 5.4: Funções de pertinência: a) Variável Idade. b) Variável Afinidade. c) Variável Diversidade. d) Variável Possibilidade de Morte.

- **Diversidade:** a diversidade dos anticorpos da população deve ser alta para que haja uma melhor exploração do espaço de busca;
- **Estagnação:** se um anticorpo está há muito tempo na população e não foi possível melhorá-lo, ele possui uma grande possibilidade de ser um máximo local e, portanto, deve ser eliminado para que outras regiões do espaço de busca possam ser exploradas.

O primeiro princípio visa manter uma população com soluções de boa qualidade enquanto os dois últimos princípios visam evitar a convergência prematura do algoritmo.

Os parâmetros do SIF utilizado neste trabalho foram definidos como:

- Agregação (E) dos antecedentes: mínimo
- Semântica da regra: conjunção mandami (mínimo)

Tabela 5.1: Regras fuzzy utilizadas no sistema de inferência fuzzy

Se Idade é <b>J</b> e Afinidade é <b>B</b> e Diversidade é <b>B</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>J</b> e Afinidade é <b>B</b> e Diversidade é <b>B</b> então Poss. de Morte é <b>A</b>
Se Idade é <b>J</b> e Afinidade é <b>B</b> e Diversidade é <b>B</b> então Poss. de Morte é <b>MA</b>
Se Idade é <b>A</b> e Afinidade é <b>M</b> e Diversidade é <b>B</b> então Poss. de Morte é <b>B</b>
Se Idade é <b>A</b> e Afinidade é <b>M</b> e Diversidade é <b>B</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>A</b> e Afinidade é <b>M</b> e Diversidade é <b>B</b> então Poss. de Morte é <b>A</b>
Se Idade é <b>I</b> e Afinidade é <b>A</b> e Diversidade é <b>B</b> então Poss. de Morte é <b>MB</b>
Se Idade é <b>I</b> e Afinidade é <b>A</b> e Diversidade é <b>B</b> então Poss. de Morte é <b>B</b>
Se Idade é <b>I</b> e Afinidade é <b>A</b> e Diversidade é <b>B</b> então Poss. de Morte é <b>A</b>
Se Idade é <b>J</b> e Afinidade é <b>B</b> e Diversidade é <b>M</b> então Poss. de Morte é <b>B</b>
Se Idade é <b>J</b> e Afinidade é <b>B</b> e Diversidade é <b>M</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>J</b> e Afinidade é <b>B</b> e Diversidade é <b>M</b> então Poss. de Morte é <b>MA</b>
Se Idade é <b>A</b> e Afinidade é <b>M</b> e Diversidade é <b>M</b> então Poss. de Morte é <b>MB</b>
Se Idade é <b>A</b> e Afinidade é <b>M</b> e Diversidade é <b>M</b> então Poss. de Morte é <b>B</b>
Se Idade é <b>A</b> e Afinidade é <b>M</b> e Diversidade é <b>M</b> então Poss. de Morte é <b>A</b>
Se Idade é <b>I</b> e Afinidade é <b>A</b> e Diversidade é <b>M</b> então Poss. de Morte é <b>MB</b>
Se Idade é <b>I</b> e Afinidade é <b>A</b> e Diversidade é <b>M</b> então Poss. de Morte é <b>B</b>
Se Idade é <b>I</b> e Afinidade é <b>A</b> e Diversidade é <b>M</b> então Poss. de Morte é <b>MB</b>
Se Idade é <b>J</b> e Afinidade é <b>B</b> e Diversidade é <b>A</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>J</b> e Afinidade é <b>B</b> e Diversidade é <b>A</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>J</b> e Afinidade é <b>B</b> e Diversidade é <b>A</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>A</b> e Afinidade é <b>M</b> e Diversidade é <b>A</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>A</b> e Afinidade é <b>M</b> e Diversidade é <b>A</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>A</b> e Afinidade é <b>M</b> e Diversidade é <b>A</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>I</b> e Afinidade é <b>A</b> e Diversidade é <b>A</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>I</b> e Afinidade é <b>A</b> e Diversidade é <b>A</b> então Poss. de Morte é <b>M</b>
Se Idade é <b>I</b> e Afinidade é <b>A</b> e Diversidade é <b>A</b> então Poss. de Morte é <b>M</b>

- Agregação das regras: máximo.
- Método de defuzzificação: o centro de massa ou centróide.

A definição do número de termos lingüísticos e a escolha dos parâmetros acima baseou-se nos parâmetros mais comuns encontrados na literatura (LEE, 1990a; LEE, 1990b).

A saída do SIF é usada para classificar um anticorpo. Esta classificação determina se um anticorpo irá sobreviver, está na iminência de morte devendo ser tratado pelo estágio de maturação fraca da afinidade ou deve ser eliminado da população.

Se um anticorpo tiver Possibilidade de Morte inferior a 0.5 ele é mantido automaticamente na população, se ele tem Possibilidade de Morte entre 0.5 e 0.85 ele passa pelo estágio de maturação fraca da afinidade (ou seja, ele está na iminência de morrer) e se o anticorpo tem Possibilidade de Morte superior a 0.85 ele morre.

### 5.1.5 Diferenças com Relação à Literatura

Os modelos de Clonalg propostos nesta dissertação possuem características semelhantes de diferentes com relação a trabalhos encontrados na literatura. A Tabela 5.2 apresenta as diferenças entre os modelos de Clonalg propostos (CUTELLO et al., 2005), (CUTELLO; NICOSIA; PAVONE, 2004) e os modelos propostos neste trabalho.

Tabela 5.2: Diferenças entre os modelos de Clonalg

Característica	Modelos Propostos	Abordagem Cutello et al
Aging	Puro e <i>Fuzzy</i>	Puro
Maturação Fraca	Faz uso após o <i>aging</i>	Não faz uso
Maturação Intensiva	Faz uso periódico	Não faz uso
Uso de Células de memória	Não faz uso	Armazena melhores células B
Espaço de Busca	Factível e Infactível	Factível

## 5.2 Redes Imunes

Os modelos baseados na teoria da Rede Imune implementados neste trabalho adotam a forma geral apresentada no Algoritmo 10. É importante salientar que, por ser geral, o algoritmo apresenta alguns estágios que não são utilizados em determinados modelos (para detalhes de cada modelo ver a Subseção 6.1.2).

---

**Algoritmo 10** Algoritmo de Rede Imunológica adotado

---

```

Rede Imune(TamPOP, FatorDUP, IdadeMAX, TaxaHIP, LimiarEST,
LimiarSUP, TamanhoMIN, PercSEL, K_melhores, M_)
  gerações ← 0;
  POP = Inicialização(TamPOP);
  Avalia(POP);
  Enquanto (!Condição_de_Parada) faça
    Enquanto (!Estabilidade_da_Rede)
      POPc ← Clonagem(POP, dup);
      POPh ← Hipermacromutação(POPc, TaxaHIP);
      POPm ← Hipermacromutação(POPc);
      Avalia(POPh);
      Avalia(POPm);
      OperadorAging(POP, POPh, POPm, IdadeMAX, POPMorte, POPa);
      POPa ← POPa ∪ MaturaçãoFrac(POPMorte);
      POP ← Seleção(POPa);
      gerações ← gerações + 1;
    Fim Enquanto
    POP ← Supressão(POP, LimiarSUP);
    Se (TamanhoPOPatual < TamanhoMIN) então
      Gerar novos anticorpos para completar a população;
    Fim Se
    Se !(K_melhores sofreram melhorias durante K_Iterações) então
      POP ← MaturaçãoFrac(POP);
    Fim Se
  Fim Enquanto
  POP ← MaturaçãoIntensiva(POP);

```

---

No algoritmo **Rede Imune** existem também duas formas de geração da população inicial: geração da população somente com anticorpos factíveis e a geração de uma população que permite a existência de anticorpos infactíveis. Após a geração da população inicial (POP) a função **Avalia()** recebe POP como parâmetro e retorna esta população com seus anticorpos avaliados.

A condição de parada utilizada foi um número pré-definido de avaliações da função avaliação. A variável **Condição\_de\_Parada** é verdadeira quando o processo de evolução do algoritmo deve parar.

A variável `Estabilidade_da_Rede` é verdadeira sempre que o valor da diferença entre a afinidade média da população atual e a afinidade média da população da geração anterior for menor que a condição de estabilidade, a qual é representada pela constante `LimiarEST`.

Enquanto a rede encontra-se instável as funções de clonagem (`Clonagem()`), hipermutação (`Hipermutação()`), hipermacromutação (`Hipermacromutação()`), avaliação (`Avalia()`), *aging* (`OperadorAging()`) e seleção (`Seleção()`) são aplicadas. Estas funções são as mesmas utilizadas nos modelos baseados no algoritmo Clonalg e suas descrições podem ser encontradas na Seção 5.1.

Depois que a rede se estabiliza, a população passa pela etapa de supressão. Esta etapa tem a intenção de eliminar possíveis redundâncias de anticorpos na população, ajudando a regular o seu tamanho. A função `Supressão()` verifica se dois anticorpos são muito semelhantes, se isto acontece ela elimina da população o anticorpo que tiver menor afinidade. A medida de semelhança adotada neste trabalho é o número de posições idênticas entre dois anticorpos (ou seja, a similaridade é o oposto da diversidade utilizada pelo Clonalg padrão). Se após a supressão restarem menos anticorpos do que o tamanho mínimo da população (`TamanhoMIN`) a população é completada com anticorpos gerados da mesma maneira que a geração da população inicial. Uma segunda forma de completar a população é através da geração de anticorpos por recombinação, a qual é explicada, na Subseção 5.2.1 e é baseada na teoria das bibliotecas gênicas.

Após a etapa de supressão é verificado se houve melhoria na afinidade de algum dos *k<sub>melhores</sub>* anticorpos da população durante `kIterações` do algoritmo. Caso não ocorra melhoria em nenhum dos *k<sub>melhores</sub>* anticorpos o processo de maturação fraca da afinidade é aplicado a toda a população.

Após a condição de parada ser alcançada toda a população passa por uma etapa de maturação intensiva da afinidade. As etapas de maturação fraca da afinidade e maturação intensiva da afinidade são descritas na Seção 5.1.

O algoritmo descrito acima possui características presentes no algoritmo Opt-ainet (DE CASTRO; TIMMIS, 2002b) e no algoritmo copt-ainet (FRANÇA; GOMES, 2006; GOMES, 2006). As semelhanças com a rede opt-ainet incluem a forma de verificação da estabilidade da rede e a forma de supressão. Já as características presentes também na rede copt-ainet são a forma de determinar o número de clones

que cada anticorpo irá gerar (fixo), a verificação da estabilidade da rede (que é igual à verificação feita pela opt-ainet), o mecanismo de supressão da rede (que é igual ao mecanismo da opt-ainet), a etapa de maturação fraca da afinidade e a etapa de maturação intensiva da afinidade. A forma de seleção aqui adotada não corresponde ao mecanismo de seleção utilizado nem pela opt-ainet e nem pela copt-ainet.

A opção por criar modelos de rede que não se enquadram na opt-ainet nem na copt-ainet se deve ao fato de que os modelos de rede utilizados neste trabalho foram uma extensão dos modelos baseados no Clonalg pela adição dos passos de supressão, ajuste automático do tamanho da população e recombinação.

### 5.2.1 Recombinação

O processo de recombinação adotado neste trabalho (inspirado na teoria das bibliotecas gênicas, comentada na Seção 3.3.2) tem o seguinte comportamento: durante todo o processo de evolução do SIA um repositório contendo os  $n$  melhores anticorpos da população é mantido. Para a geração de um novo anticorpo cada anticorpo do repertório é dividido em  $p$  partes, essas partes são selecionadas de forma aleatória para compor o novo anticorpo (chamado de anticorpo recombinado). O tamanho de cada parte é determinado de maneira aleatória e varia de 1 a  $2 * Tam_{PROT}/p$  (com exceção da última parte que é do tamanho necessário para completar o anticorpo). Cabe salientar que, para o problema tratado neste trabalho, a parte selecionada no anticorpo do repertório ocupa a mesma posição no anticorpo recombinado. Um exemplo de recombinação pode ser visto na Figura 5.5.

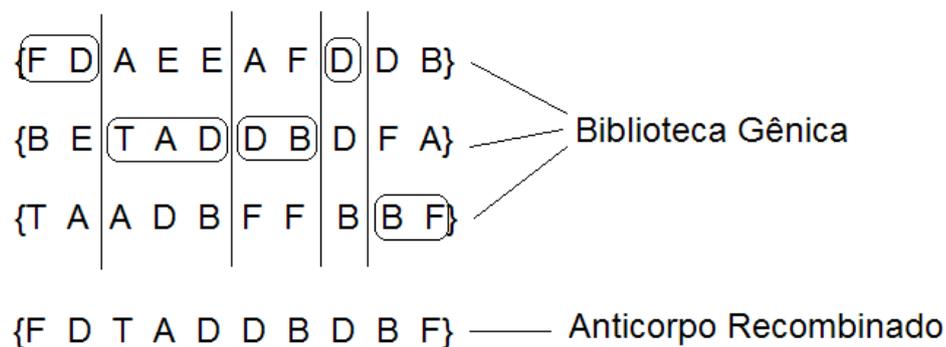


Figura 5.5: Geração de um anticorpo de tamanho 10 através de recombinação.

## 5.2.2 Diferenças com Relação à Literatura

Os modelos de Rede Imune propostos neste trabalho possuem características semelhante e diferentes com relação a dois tipos de rede encontradas na literatura. A Tabela 5.3 ilustra as diferenças entre os modelos de rede propostos nesta dissertação, a opt-Ainet e a copt-Ainet.

Tabela 5.3: Diferenças entre os modelos de Rede

<b>Característica</b>	<b>Modelos Propostos</b>	<b>Opt-Ainet</b>	<b>Copt-Ainet</b>
Clonagem	Fixo	Proporcional à Afinidade	Fixo
Operadores de Mutação	Hipermutação e Hipermacromutação	Hipermutação	Hipermutação
Seleção	Melhores anticorpos	Melhor clone substitui o pai	Melhor clone substitui o pai
Tamanho mínimo	Fixo	Não existe	Fixo
Maturação Fraca	Semelhante a Hipermutação e a Hipermacromutação	Não Existe	Busca Tabu Fraca
Maturação Intensiva	Busca Tabu	Não Existe	Busca Tabu Intensiva

---

## *Simulações e Resultados*

---

---

Este capítulo apresenta os modelos implementados (Seção 6.1) e os resultados obtidos com cada um destes modelos. A Seção 6.3 contém os resultados correspondentes aos modelos baseados no algoritmo Clonalg enquanto a Seção 6.4 mostra os resultados dos algoritmos baseados na Rede Imune. O confronto entre o melhor modelo baseado no algoritmo Clonalg e o melhor modelo baseado na teoria da Rede Imune é apresentado na Seção 6.5. A Seção 6.6 traz o resultado da comparação do melhor modelo obtido com alguns trabalhos da literatura.

Para cada modelo desenvolvido foram realizadas 50 execuções independentes. As simulações foram feitas em máquinas com processador Athlon 64 2.0 ou equivalentes e 512Mb de memória RAM<sup>1</sup>.

Para validação dos resultados, os modelos de SIAs desenvolvidos foram aplicados no *benchmark* Tortilla (HART; ISTRAIL, visitado em 08/2005). As seqüências e seus respectivos  $E^*$  são apresentados na Tabela 6.1.

---

<sup>1</sup>cada execução foi realizada com uma semente diferente para o gerador de números aleatórios, o que faz com que a população inicial seja diferente para cada execução

Tabela 6.1: Seqüências do *benchmark* para o modelo 3D HP utilizadas neste trabalho com os melhores valores de energia conhecidos -  $E^*$

N <sup>o</sup>	Tamanho	Seqüência	$E^*$
1	20	hphp <sub>2</sub> h <sub>2</sub> php <sub>2</sub> hph <sub>2</sub> p <sub>2</sub> hph	-11
2	24	h <sub>2</sub> p <sub>2</sub> (hp <sub>2</sub> ) <sub>6</sub> h <sub>2</sub>	-13
3	25	p <sub>2</sub> hp <sub>2</sub> (h <sub>2</sub> p <sub>4</sub> ) <sub>3</sub> h <sub>2</sub>	-9
4	36	p <sub>3</sub> h <sub>2</sub> p <sub>2</sub> h <sub>2</sub> p <sub>5</sub> h <sub>7</sub> p <sub>2</sub> h <sub>2</sub> p <sub>4</sub> h <sub>2</sub> p <sub>2</sub> hp <sub>2</sub>	-18
5	48	p <sub>2</sub> h(p <sub>2</sub> h <sub>2</sub> ) <sub>2</sub> p <sub>5</sub> h <sub>10</sub> p <sub>6</sub> (h <sub>2</sub> p <sub>2</sub> ) <sub>2</sub> hp <sub>2</sub> h <sub>5</sub>	-29
6	50	h <sub>2</sub> (ph) <sub>3</sub> ph <sub>4</sub> p(hp <sub>3</sub> ) <sub>2</sub> hp <sub>4</sub> (hp <sub>3</sub> ) <sub>2</sub> hph <sub>4</sub> (ph) <sub>3</sub> ph <sub>2</sub>	-26
7	60	p <sub>2</sub> h <sub>3</sub> ph <sub>8</sub> p <sub>3</sub> h <sub>10</sub> php <sub>3</sub> h <sub>12</sub> p <sub>4</sub> h <sub>6</sub> ph <sub>2</sub> php	-49

## 6.1 Modelos Implementados

Esta seção descreve os modelos baseados em SIA implementados neste trabalho. A subseção 6.1.1 apresenta os modelos baseados no princípio da seleção clonal, mais precisamente no algoritmo Clonalg. Já a subseção 6.1.2 apresenta os modelos baseados no princípio da rede imune, os quais possuem características dos algoritmo opt-aiNet e do algoritmo copt-aiNet.

### 6.1.1 Clonalgs

#### ClonalgI

O primeiro algoritmo baseado no Clonalg corresponde à sua forma padrão (descrita no Algoritmo 1), não possuindo as etapas de *Aging*, maturação fraca da afinidade e maturação intensiva da afinidade. Este algoritmo foi implementado com o intuito de servir de base para comparações com os modelos que possuem as melhorias propostas neste trabalho. Neste modelo o algoritmo Clonalg trabalha apenas com anticorpos factíveis.

#### ClonalgII

Percebeu-se que uma das deficiências do algoritmo ClonalgI era o fato de, muitas vezes, ele não conseguir melhorar anticorpos que estão há muito tempo na população. Para tratar essa deficiência no algoritmo ClonalgII foram introduzidos o operador

de *Aging* Puro e o estágio de maturação fraca da afinidade.

## ClonalgIII

Para tornar o algoritmo mais próximo do comportamento observado na natureza o operador de *Aging* Puro foi substituído pelo operador de *Aging Fuzzy*. Salvo esta mudança o ClonalgIII é idêntico ao ClonalgII.

## ClonalgI<sub>inf</sub>

Os três primeiros modelos do algoritmo Clonalg trabalham apenas com espaço de soluções factíveis. Isto aumenta consideravelmente o tempo computacional dos algoritmos, pois a geração de novos anticorpos é custosa. Para resolver este problema foram desenvolvidos modelos do algoritmo Clonalg que aceitam anticorpos infactíveis penalizando-os. O algoritmo ClonalgI<sub>inf</sub> é o mesmo algoritmo que o ClonalgI exceto por permitir soluções infactíveis na geração da população inicial. Cabe salientar que, durante o processo evolutivo, os operadores continuam gerando somente soluções factíveis. Isso ocorre para que a busca seja concentrada em regiões de soluções factíveis.

Cuidados especiais foram tomados para garantir que apenas anticorpos factíveis possam ser considerados como melhores soluções (um anticorpo factível nunca é substituído por um anticorpo infactível). Isto garante que todos resultados apresentados no Capítulo 6 correspondem a soluções válidas.

Esse algoritmo serve de base de comparação para os modelos que fazem uso de espaço infactível.

## ClonalgII<sub>inf</sub>

Para aferir a influência de uma população com Anticorpos infactíveis num algoritmo Clonalg que possua operador de *Aging* Puro e maturação fraca da afinidade foi desenvolvido o ClonalgII<sub>inf</sub>. Este algoritmo corresponde ao ClonalgII trabalhando em um espaço de busca considerando soluções infactíveis.

## ClonalgIII<sub>inf</sub>

O ClonalgIII<sub>inf</sub> procura tornar o ClonalgII<sub>inf</sub> mais próximo da natureza. Para isto, o operador de *Aging* Puro do ClonalgII<sub>inf</sub> foi novamente substituído pelo operador de *Aging Fuzzy*.

## ClonalgIII<sub>inf-BT</sub>

Na tentativa de melhorar o desempenho obtido pelo algoritmo ClonalgIII<sub>inf</sub> foi introduzida uma etapa de maturação intensiva da afinidade, a qual neste trabalho corresponde a uma Busca Tabu. O modelo contento esta alteração foi denominado ClonalgIII<sub>inf-BT</sub>.

A Tabela 6.2 sumariza as características opcionais presentes em cada versão do algoritmo Clonalg.

Tabela 6.2: Características dos algoritmos propostos baseados no Clonalg.

	CI	CII	CIII	CI <sub>inf</sub>	CII <sub>inf</sub>	CIII <sub>inf</sub>	CIII <sub>inf-BT</sub>
Infactibilidade	-	-	-	X	X	X	X
<i>Aging</i> Puro	-	X	-	-	X	-	-
<i>Aging Fuzzy</i>	-	-	X	-	-	X	X
Maturação Fraca	-	X	X	-	X	X	X
Maturação Intensiva	-	-	-	-	-	-	X

Alguns modelos de Clonalg aqui desenvolvidos (CI, CIII, CIII<sub>inf</sub> e CIII<sub>inf-BT</sub>) resultaram em um artigo aceito para o EvoCop 2007 (ALMEIDA; GONÇALVES; DELGADO, 2007).

### 6.1.2 Redes Imunes

A seqüência de modelos de Rede Imune proposta tem o intuito de testar se o comportamento deste algoritmo frente às modificações é semelhante ao comportamento do algoritmo Clonalg. Portanto, os modelos de Rede Imune são análogos aos modelos de Clonalg discutidos na seção anterior.

## RedeI

Assim como para o algoritmo Clonalg foi implementado um modelo de rede considerado básico. Porém, este modelo não pode ser considerado nem um modelo de Opt-anet e nem um modelo de Copt-ainet. A RedeI possui os passos contidos no Algoritmo 10 exceto os passos da função de *aging* (*OperadorAging()*) e do processo de maturação intensiva da afinidade. Este modelo de rede foi desenvolvido para servir como base de comparação para as versões seguintes, as quais propõem melhorias para a RedeI. A RedeI trabalha com o espaço de busca contendo apenas soluções factíveis.

## RedeII

A RedeII estende a RedeI com o uso do operador de *Aging* Puro para decidir quais anticorpos devem ser eliminados da população ou não. Este operador é idêntico àquele utilizado nos modelos baseados no Clonalg e sua descrição é feita na Subseção 5.1.4.

## RedeIII

A RedeIII corresponde a RedeII, porém ao invés do operador de *Aging* Puro ela faz uso do operador de *Aging* Fuzzy. Este operador foi descrito na Seção 5.1.4, pois é o mesmo operador utilizado nos modelos baseados no Clonalg. A RedeIII trabalha somente com soluções factíveis durante seu processo de evolução.

## RedeI<sub>inf</sub>

A RedeI<sub>inf</sub> busca averiguar o efeito da presença de anticorpos inactíveis na população. Este modelo permite a geração de soluções com colisões durante a geração da população inicial e na geração de anticorpos para completar a população após a supressão da rede. A RedeI<sub>inf</sub> corresponde a RedeI trabalhando em espaço de soluções inactíveis.

### **RedeI<sub>inf-Rec</sub>**

Para verificar o impacto que o mecanismo de recombinação tem no modelo RedeI<sub>inf</sub>, este mecanismo foi aplicado como forma de geração de anticorpos para completar a população após a supressão da rede. Então, a RedeI<sub>inf-Rec</sub> corresponde a RedeI<sub>inf</sub> utilizando a recombinação para a geração de soluções após a supressão ao invés da geração de anticorpos de forma aleatória. Além disso, este modelo permite que a população contenha soluções ineficazes tanto na geração da população inicial como pelo mecanismo de recombinação, pois a recombinação não garante que o anticorpo recombinado não contenha colisões.

### **RedeII<sub>inf</sub>**

O modelo RedeII<sub>inf</sub> é idêntico ao modelo RedeII, exceto por permitir que anticorpos ineficazes façam parte da população.

### **RedeIII<sub>inf</sub>**

A RedeIII<sub>inf</sub> corresponde à RedeIII, porém trabalha com espaço de soluções ineficazes da mesma maneira que os três modelos descritos anteriormente.

### **RedeI<sub>inf-BT</sub>**

Na tentativa de melhorar o resultados do modelo de rede que apresentou os melhores resultados (RedeI<sub>inf</sub>) este modelo foi incrementado com a etapa de maturação intensiva da afinidade.

A Tabela 6.3 sumariza as características opcionais presentes em cada modelo de Rede Imune.

Tabela 6.3: Características dos algoritmos propostos baseados na rede imune.

	RI	RII	RIII	RI <sub>inf</sub>	RI <sub>inf-Rec</sub>	RII <sub>inf</sub>	RIII <sub>inf</sub>	RI <sub>inf-BT</sub>
Infactibilidade	-	-	-	X	X	X	X	X
<i>Aging Puro</i>	-	X	-	-	-	X	-	-
<i>Aging Fuzzy</i>	-	-	X	-	-	-	X	-
Maturação Fraca	X	X	X	X	X	X	X	X
Maturação Intensiva	-	-	-	-	-	-	-	X
Recombinação	-	-	-	-	X	-	-	-

## 6.2 Ajuste dos Parâmetros

Foram realizadas algumas simulações para a escolha e verificação da sensibilidade dos parâmetros do algoritmo Clonalg. Como o objetivo do trabalho é aferir o impacto da inserção de novas características nos modelos (operador de *Aging Puro*, operador de *Aging Fuzzy* e estágios de Maturação da Afinidade), os parâmetros foram ajustados para os modelos ClonalgI e ClonalgII e aplicados em todos os outros modelos (inclusive nas Redes Imunes).

Para o ajuste de parâmetros foram utilizadas as proteínas de tamanho 24 e 48. Essa escolha foi feita levando-se em consideração o tamanho da proteína e a dificuldade para dobrá-la. Para cada conjunto de parâmetros foram realizadas vinte execuções do algoritmo, cada uma com uma semente diferente para o gerador de números aleatórios.

### 6.2.1 Ajuste dos Parâmetros dos Modelos Baseados no Clonalg

O modelo ClonalgI foi utilizado para a definição do tamanho da população de anticorpos ( $Tam_{POP}$ ), taxa de hipermutação ( $Taxa_{HIP}$ ) e número de clones de cada anticorpo ( $Fator_{DUP}$ ). O ClonalgII foi utilizado para ajustar a idade máxima que um anticorpo pode assumir antes de sua morte ficar iminente ( $Idade_{MAX}$ ), visto que o ClonalgI não possui operador de *aging*.

O conjunto inicial de parâmetros foi extraído de Cutello et al. (2005). Os autores do referido trabalho também empregaram um Sistema Imunológico Artificial

baseado no algoritmo Clonalg com operador de *aging* para a Predição da Estrutura de Proteínas no modelo 3D HP. Os parâmetros iniciais utilizados foram:  $Tam_{POP} = 10$ ,  $Taxa_{HIP} = 40\%$ ,  $Fator_{DUP} = 2$  e  $Idade_{MAX} = 5$ .

Neste trabalho foram considerados os seguintes valores para os parâmetros do algoritmo Clonalg:

- Tamanho da população = 5, 10, 15, 20 e 25;
- Taxa de hipermutação = 30%, 40%, 50%, 60% e 70%;
- Fator de duplicação = 2, 3, 4, 5 e 6;
- Idade máxima = 5, 10, 15, 20 e 25.

Existem 625 possíveis combinações para este conjunto de parâmetros. Testar todas estas combinações é uma tarefa bastante complicada e, mesmo que executada, não resolve o problema de otimização paramétrica o qual é mais complexo ainda do que a própria PEP no modelo 3D HP. Portanto, fez-se necessária a adoção de uma metodologia para simplificar os testes e definir quais combinações seriam usadas.

Os testes foram realizados na ordem apresentada acima pois acreditava-se que essa era uma decrescente de sensibilidade do algoritmo ao parâmetro em questão, ou seja, foram testados e ajustados primeiro os parâmetros que poderiam influenciar mais o desempenho do algoritmo.

Após todos os valores para um parâmetro serem testados, foram selecionados dois valores para este parâmetro. A escolha destes valores foi baseada nos valores médios de energia e também no número de ótimos (ou sub-ótimos) encontrados.

Enquanto um valor era testado os outros eram fixados nos dois valores selecionados em etapas anteriores do ajuste de parâmetros ou no valor definido no conjunto inicial de parâmetros, caso o parâmetro ainda não tenha sido ajustado.

Para o tamanho da população foram selecionados os valores 10 e 15. Após o ajuste da taxa de mutação foram escolhidas as combinações tamanho de população igual a 10 e taxa de hipermutação de 60% e tamanho da população igual a 15 e taxa de hipermutação de 50%. O ajuste do número de clones

- $Tam_{POP} = 10$ ;
- $Taxa_{HIP} = 60$ ;
- $Fator_{DUP} = 4$ ;
- $Idade_{MAX} = 5$ .

resultou nas combinações tamanho da população igual a 10, taxa de hipermutação de 60% e número de clones igual a 4 e tamanho de população igual a 15, taxa de hipermutação de 50% e número de clones igual a 4. Por último, depois do ajuste da idade máxima, foram selecionadas as combinações tamanho da população igual a 10, taxa de hipermutação de 60%, número de clones igual a 4 e idade máxima igual a 5 e tamanho de população igual a 15, taxa de hipermutação de 50%, número de clones igual a 4 e idade máxima igual a 10.

Dentre estas duas combinações a configuração final dos parâmetros escolhida foi a seguinte:

O critério de parada utilizado pelos algoritmos foi fixado em  $10^5$  avaliações para que fosse possível uma comparação dos resultados com (CUTELLO et al., 2005) e (COTTA, 2003).

Os parâmetros do SIF foram determinados manualmente com o auxílio de um especialista. Estes parâmetros foram detalhados na Subseção 5.1.4.

Os parâmetros da busca tabu foram baseados nos resultados obtidos no trabalho de Blazewicz, Lukasiak e Milostan (2005) e foram: tamanho da lista tabu igual a 20, número máximo de iterações igual a 200, vizinhança de uma mutação e o critério de aspiração melhoria dos anticorpos encontrados até o momento. A BT é realizada a cada  $MI$  avaliações, onde  $MI$  é igual a 10.000.

Durante os testes realizados para o ajuste dos parâmetros dos modelos baseados no algoritmo Clonalg foi possível observar que os resultados obtidos sofriam uma pequena variação ao redor dos melhores valores obtidos, ou seja, o algoritmo era pouco sensível a variações dos parâmetros.

## 6.2.2 Ajuste dos Parâmetros dos Modelos Baseados na Teoria da Rede Imune

Neste trabalho foram considerados os seguintes valores para os parâmetros dos modelos baseados na rede imune:

- Limiar de Supressão = 0.85, 0.9, 0.95 e 0.98;
- Limiar de Estabilidade = 0.01, 0.005, 0.002 e 0.001;
- Percentual de Seleção = 100%, 110%, 120% e 130%;
- $M_{iterações}$  = 20, 30, 40 e 50;
- $k_{melhores}$  = 10, 20, 30 e 40;
- Tamanho mínimo da população = 20, 30, 40 e 50.

Os parâmetros comuns determinados para os modelos de Clonalg foram utilizados para as Redes Imunes. Assim, restou a determinação dos seguintes parâmetros:  $Taxa_{SUP}$ ,  $Limiar_{EST}$ ,  $Perc_{SEL}$ ,  $k_{melhores}$ ,  $M_{iterações}$  e  $Tam_{min}$ . A metodologia empregada foi a mesma utilizada para a determinação dos parâmetros dos modelos baseados no Clonalg, com a distinção de que o modelo utilizado para o ajuste foi a RedeI.

A configuração final encontrada foi:

- $Limiar_{SUP} = 0.9$ ;
- $Limiar_{EST}$  a 0.002;
- $Perc_{SEL}$  120%;
- $M_{iterações} = 50$ ;
- $k_{melhores} = 30$ ; e
- $Tamanho_{MIN} = 50$ .

Durante os testes para os ajustes dos parâmetros foi constatado que os algoritmos baseados na teoria da Rede Imune são bastante sensíveis aos parâmetros.

## 6.3 Resultados dos Modelos Baseados no Clonalg

Os resultados obtidos com os modelos baseados no Clonalg são sumarizados na Tabela 6.7 e nas Figuras 6.1 e 6.2. A Tabela 6.7 apresenta os dados referentes ao número do problema (instância) tratado ( $N^o$ ), tamanho da referida proteína (L), melhor valor de energia encontrado (Melhor), média (Média) e desvio padrão ( $\sigma$ ) obtidos para as 50 execuções e tempo médio em minutos gasto durante as 50 execuções ( $T(\text{min})$ )<sup>2</sup>.

A Figura 6.1 apresenta o valor percentual das médias obtidas pelos modelos propostos e os ótimos reportados na literatura para cada uma das instâncias, ou seja, valores superiores a 1 indicam que a média obtida pelo algoritmo é superior ao melhor valor conhecido. Já a Figura 6.2 representa o tempo médio em minutos gasto durante uma execução dos algoritmos para cada uma das instâncias (o gráfico é apresentado em escala logarítmica). Ambas as figuras são referentes aos dados coletados ao longo das 50 execuções.

Para averiguar as diferenças entre os resultados encontrados testes estatísticos com precisão de 95% foram realizados. O tipo de teste que deve ser aplicado nos dados depende do volume e da distribuição dos mesmos. Primeiramente foi realizado um teste para determinar se os dados obedeciam uma distribuição normal. Existe uma vasta gama de testes que poderiam ser aplicados para isto; contudo, pelo pequeno volume de dados (50 execuções), o teste realizado foi o Lilliefors (ABDI; MOLIN, 2007). Segundo os resultados obtidos pelo teste, os dados não seguem uma distribuição normal com 95% de confiabilidade.

Devido a isto, o teste-t que é geralmente aplicado a um conjunto pequeno de dados não pôde ser utilizado para determinar se existe ou não diferença significativa entre as médias obtidas. Para casos como este recomenda-se o uso do teste ranksum (CONOVER, 1980) (também conhecido por Mann-Whitney *ranksum* e Wilcoxon *ranksum*). Este teste verifica a hipótese de duas distribuições de dados serem diferentes.

---

<sup>2</sup>As sementes geradoras de números aleatórios são diferentes para cada uma das 50 execuções

Tabela 6.4: Comparação dos resultados obtidos pelos modelos baseados no algoritmo Clonalg

		Clonalg I				Clonalg II			
Nº	L	Melhor	Média	$\sigma$	T(min)	Melhor	Média	$\sigma$	T(min)
1	20	-11	-10.4	0.57	1.43	-11	-11	0	2.05
2	24	-13	-11.26	0.90	1.63	-13	-12.98	0.14	2.17
3	25	-9	-8.06	0.87	1.58	-9	-8.98	0.14	2.15
4	36	-18	-15.04	1.37	2.24	-18	-16.30	0.76	4.71
5	48	-29	-24.20	2.22	4.05	-29	-26.08	1.21	12.34
6	50	-27	-23.11	2.05	4.31	-29	-25.80	1.40	10.63
7	60	-48	-42.65	2.74	10.43	-47	-42.78	1.44	99.98
		Clonalg III				ClonalgI <sub>inf</sub>			
Nº	L	Melhor	Média	$\sigma$	T(min)	Melhor	Média	$\sigma$	T(min)
1	20	-11	-11	0	0.29	-11	-10.40	0.70	0.82
2	24	-13	-12.68	0.62	0.40	-13	-11.30	1.09	0.96
3	25	-9	-8.98	0.14	0.43	-9	-7.82	0.94	0.93
4	36	-18	-17.20	0.90	0.69	-18	-14.76	1.33	1.42
5	48	-29	-26.38	1.10	2.20	-28	-23.63	1.95	3.10
6	50	-30	-25.04	1.29	3.96	-27	-23.24	1.95	3.07
7	60	-47	-43.04	1.29	61.55	-47	-41.58	2.84	9.31
		ClonalgII <sub>inf</sub>				ClonalgIII <sub>inf</sub>			
Nº	L	Melhor	Média	$\sigma$	T(min)	Melhor	Média	$\sigma$	T(min)
1	20	-11	-11	0	0.99	-11	-11	0	0.36
2	24	-13	-12.96	0.20	1.19	-13	-12.90	0.36	0.46
3	25	-9	-9	0	1.21	-9	-9	0	0.50
4	36	-18	-16.04	0.86	2.60	-18	-17.28	0.88	0.89
5	48	-28	-25.49	1.12	7.76	-30	-27.02	1.17	2.09
6	50	-29	-25.48	1.22	6.87	-28	-25.06	1.15	2.34
7	60	-45	-42.20	1.20	71.71	-47	-44.02	2.00	10.05
		ClonalgIII <sub>inf-BT</sub>				E*			
Nº	L	Melhor	Média	$\sigma$	T(min)				
1	20	-11	-11	0	0.98				
2	24	-13	-12.98	0.14	1.59				
3	25	-9	-9	0	1.28				
4	36	-18	-17.76	0.59	2.75				
5	48	<b>-30</b>	-28.49	0.92	5.83				
6	50	<b>-30</b>	-26.36	1.01	11.17				
7	60	<b>-51</b>	-46.16	1.49	13.83				

Tabela 6.5: Hipóteses de que as distribuições sejam diferentes (h) com seus respectivos p-valores obtidos nas simulações dos modelos de Clonalg

	ClonalgI × ClonalgII		ClonalgII × ClonalgIII	
Proteína	h	p-valores	h	p-valores
1	V	3.3264e-010	F	idênticas
2	V	1.3912e-016	V	0.0021
3	V	2.0753e-011	F	idênticas
4	V	2.1891e-005	F	0.5054
5	V	0.0013	F	0.7952
6	V	0.0025	F	0.1562
7	V	2.2267e-07	F	0.5207
	ClonalgIII × ClonalgIII <sub>inf</sub>		ClonalgIII <sub>inf</sub> × ClonalgIII <sub>inf-BT</sub>	
Proteína	h	p-valores	h	p-valores
1	F	idênticas	F	idênticas
2	F	0.0536	F	0.1681
3	F	idênticas	F	idênticas
4	V	7.0432e-010	V	0.0014
5	V	5.0532e-005	V	1.2269e-009
6	V	0.0015	V	1.1050e-007
7	V	2.7629e-004	V	3.3655e-008

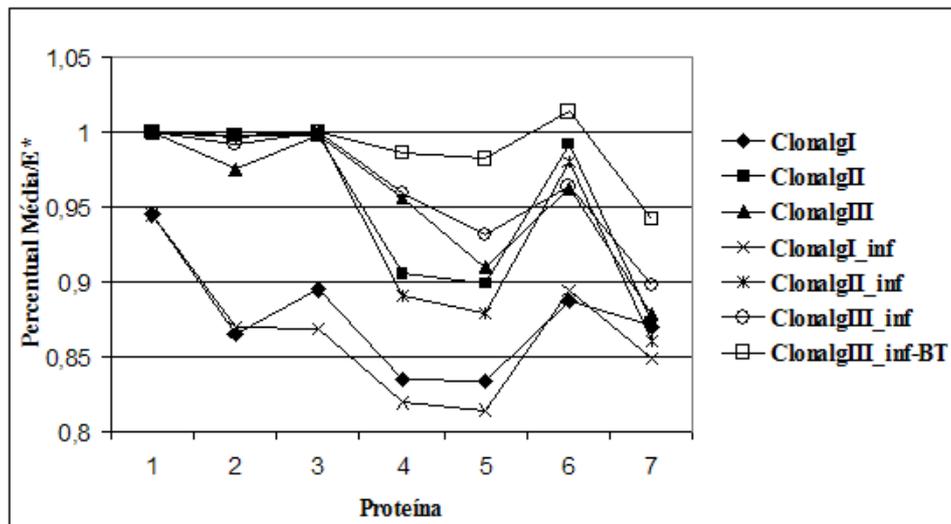


Figura 6.1: Gráfico representando a diferença percentual da média dos modelos baseados no Clonalg e do ótimo conhecido na literatura.

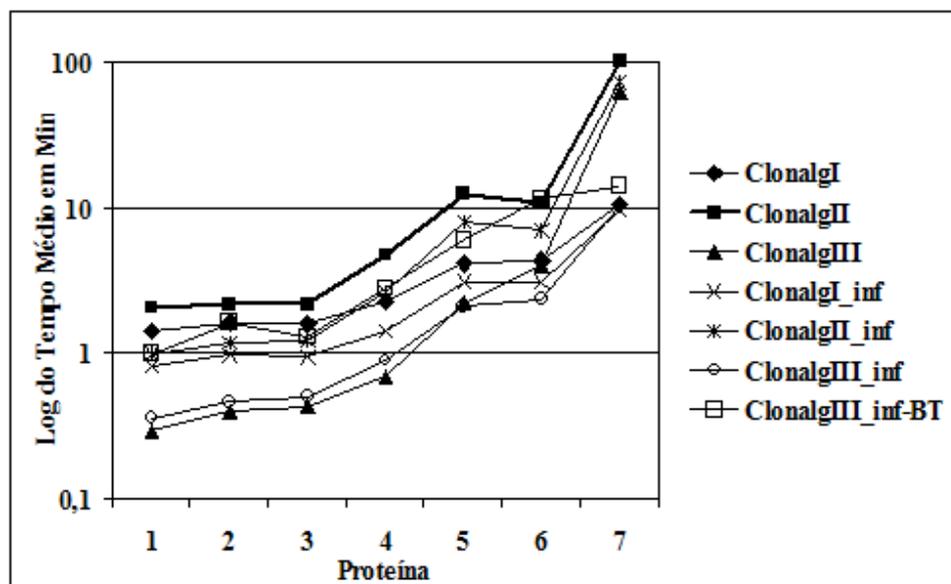


Figura 6.2: Gráfico do tempo que cada modelo baseado no Clonalg leva para uma execução.

Para aferir a eficiência do operador de *Aging* Puro foram comparados os modelos ClonalgI e ClonalgII, assim como os modelos ClonalgI<sub>inf</sub> e ClonalgII<sub>inf</sub>. As médias do ClonalgII, o qual possui o operador de *Aging* Puro, são superiores às médias do ClonalgI para todas as instâncias de teste e, de acordo com o teste ranksum, a diferença entre as médias é significativa (ver Figura 6.1). O ClonalgII<sub>inf</sub> (com operador de *Aging* Puro) também obteve melhores médias para todas as instâncias, mas a diferença entre as médias não foi estatisticamente significativa para a proteína de tamanho 60 (instância 7). Portanto, de forma geral, o uso do operador de *Aging*

Puro melhora os resultados obtidos pelos modelos sem *aging*.

Analogamente, para comparar o desempenho do *Aging Fuzzy* com o do *Aging Puro*, foram confrontados o ClonalgII (*Aging Puro*) com o ClonalgIII (*Aging Fuzzy*) e o ClonalgII<sub>inf</sub> (*Aging Puro*) com o ClonalgIII<sub>inf</sub> (*Aging Fuzzy*). Apesar do ClonalgIII possuir melhores médias para três das instâncias e o ClonalgII possuir melhores médias para duas das instâncias, segundo o teste ranksum apenas existe diferença significativa para a segunda instância. Porém, como pode ser observado na Figura 6.2, a diferença entre o tempo que o ClonalgII e o ClonalgIII levam para obter estes resultados é bastante significativa: enquanto o ClonalgIII é um dos modelos mais rápidos, o ClonalgII é o modelo mais lento, conforme pode ser visto na Figura 6.2. Já ao comparar-se o ClonalgII<sub>inf</sub> com o ClonalgIII<sub>inf</sub>, novamente tem-se que o modelo com *Aging Fuzzy* (ClonalgIII<sub>inf</sub>) vence em três instâncias e que o modelo com *Aging Puro* (ClonalgII<sub>inf</sub>) vence em duas instâncias (ver Figura 6.1). Contudo, neste caso, as instâncias onde o ClonalgIII<sub>inf</sub> vence possuem diferenças significativas enquanto que as instâncias onde o ClonalgII<sub>inf</sub> vence não possuem diferenças significativas. Com relação ao tempo também é possível perceber que o modelo com *Aging Fuzzy* é muito mais rápido do que o modelo com *Aging Puro*.

Para que fosse possível observar o impacto que a permissão de anticorpos com colisões (infectíveis) no espaço de busca iria causar, os três primeiros modelos foram contrastados com seus modelos correspondentes trabalhando com espaço de soluções infectíveis (ClonalgI x ClonalgI<sub>inf</sub>, ClonalgII x ClonalgII<sub>inf</sub> e ClonalgIII x ClonalgIII<sub>inf</sub>). Ao comparar o ClonalgI com o ClonalgI<sub>inf</sub> não existem diferenças significativas entre as médias para nenhuma das instâncias. A mesma coisa ocorre ao confrontarem-se as médias dos modelos ClonalgII e ClonalgII<sub>inf</sub>. Já o ClonalgIII<sub>inf</sub> possui melhores médias (estatisticamente significativas) para as quatro últimas instâncias. Disto, pôde-se notar que o uso de espaço de soluções infectíveis não prejudica o desempenho dos modelos e no caso do modelo que faz uso do operador de *Aging Fuzzy* há até mesmo uma melhora na qualidade dos resultados obtidos. Além disto, os modelos que trabalham com a permissão de soluções infectíveis durante a evolução são capazes de executar em um tempo menor, na maioria das vezes, do que aqueles que trabalham apenas com soluções factíveis, principalmente para as instâncias maiores (ver Figura 6.2).

A hibridização do modelo ClonalgIII<sub>inf</sub> com a metaheurística Busca Tabu, a qual resultou no modelo ClonalgIII<sub>inf</sub>-BT, foi capaz de melhorar significativamente

o desempenho médio do  $\text{ClonalgIII}_{inf}$  para as últimas quatro instâncias (tamanhos 36, 48, 50 e 60). Isso foi possível com o acréscimo do custo computacional resultando no aumento do tempo gasto pelo modelo, como pode ser visto na Tabela 6.2.

A seguir é feita uma comparação mais detalhada entre os três modelos baseados no Clonalg que obtiveram os melhores resultados.

### 6.3.1 Comparação Entre os Três Melhores Modelos de Clonalg

Nesta seção os modelos com melhores resultados (ou seja, aqueles que trabalham com espaço infactível) são analisados mais detalhadamente com um novo conjunto de simulações. De modo a entender o funcionamento dos melhores modelos propostos, foram coletados dados referentes à diversidade populacional, afinidade média, afinidade dos melhores anticorpos e percentual de anticorpos mortos a cada geração. Estes dados são apresentados na Figura 6.3. Os dados apresentados correspondem à média de 20 execuções de cada algoritmo para a proteína 48<sup>3</sup>.

Com relação às afinidades médias dos modelos pode-se notar que no início da evolução elas possuem grandes picos. Isto ocorre, pois os três modelos considerados trabalham com espaço de soluções infactíveis o que permite a existência de um grande número de anticorpos com colisões nas primeiras gerações dos algoritmos. A Figura 6.3-a mostra que os dois modelos que fazem uso do operador de *Aging Fuzzy* mantêm as médias mais estáveis, sem grandes saltos como ocorre com o modelo com *Aging Puro* ( $\text{ClonalgII}_{inf}$ ).

O comportamento geral das melhores soluções encontradas a cada geração é bem semelhante para os três modelos de acordo com a Figura 6.3-b. É perceptível que o modelo que faz uso da Busca Tabu leva um tempo maior para encontrar soluções factíveis, mas é capaz de superar os outros dois modelos ao longo de sua evolução.

A diversidade mantida por cada um dos modelos pode ser observada na Figura 6.3-c. Os modelos  $\text{ClonalgIII}_{inf}$  e  $\text{ClonalgIII}_{inf-BT}$  (*Aging Fuzzy*) mantêm a diversidade de suas populações em intervalos semelhantes enquanto o  $\text{ClonalgII}_{inf}$  permanece com diversidade inferior na maioria do tempo, mesmo tendo picos de altas

---

<sup>3</sup>As sementes geradoras de números aleatórios são diferentes para cada uma das execuções

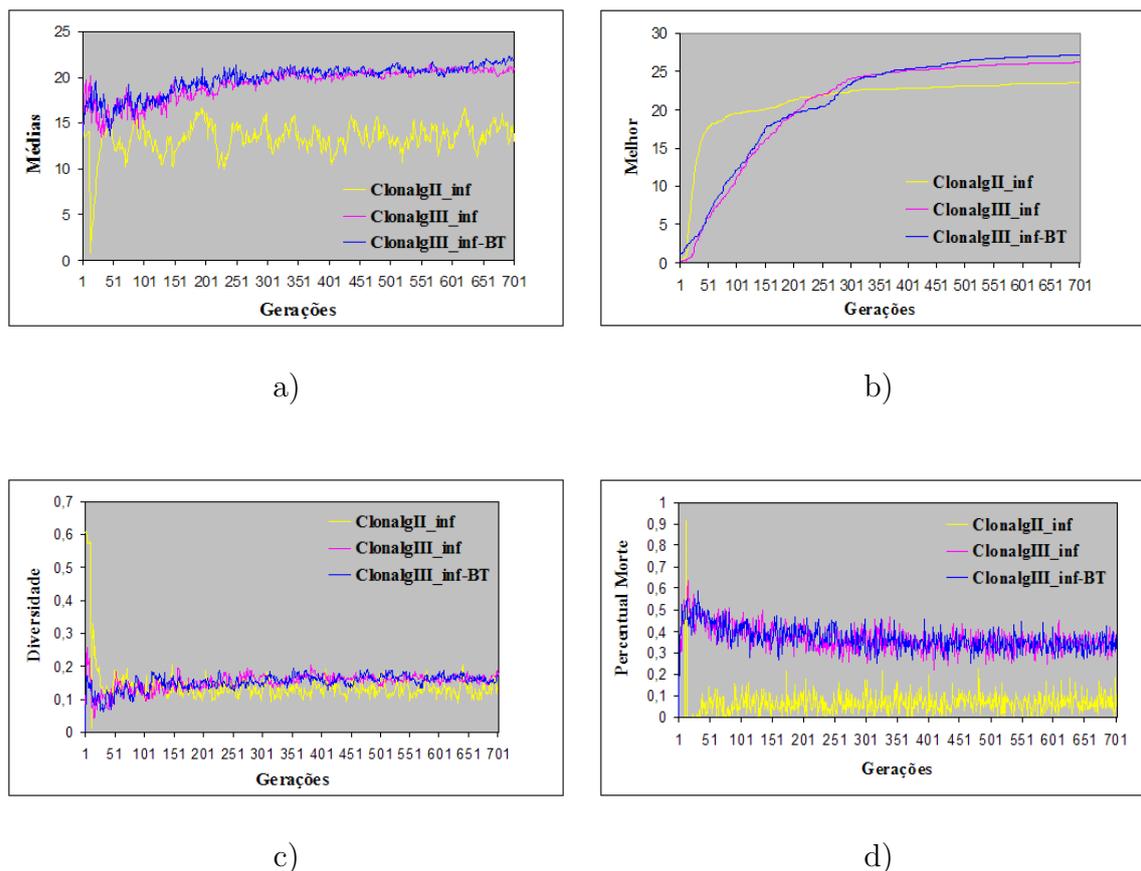


Figura 6.3: Comportamento dos três melhores clonalgs, ClonalgII<sub>inf</sub>, ClonalgIII<sub>inf</sub> e ClonalgIII<sub>inf</sub>-BT, ao longo das gerações durante uma execução para a proteína de tamanho 48: a) Média. b) Melhores. c) Diversidade. d) Percentual de Mortes.

diversidades.

Os picos de altas diversidades presentes ao longo da evolução do modelo ClonalgII<sub>inf</sub> (*Aging Puro*) são consequência dos vários momentos em que este modelo mata um grande percentual de seus anticorpos (até mesmo toda a população - ver Figura 6.3-d). Ainda com relação ao percentual de mortes por geração dos modelos, o ClonalgII<sub>inf</sub> apresenta um comportamento bem peculiar: ele mata um grande percentual de anticorpos durante uma geração, passa algumas gerações sem matar ninguém (ou matando bem poucos anticorpos) e depois torna a matar um grande percentual (próximo de 100% dos anticorpos). Já para os modelos que utilizam o operador de *Aging Fuzzy*, o percentual de morte a cada geração dos modelos concentram-se entre 20% e 60% sem muitos picos nem vales. Apesar de parecer alta, estas taxas são relativamente pequenas visto que o operador de *aging* é aplicado à população que une a população atual (POP), a população hipermutada (POP<sub>HIP</sub>) e a população hiperamacromutada (POP<sub>MACRO</sub>) cujo tamanho varia de TAM<sub>POP</sub> até

$(2 * \text{Fator}_{DUP} * \text{TAM}_{POP} + \text{TAM}_{POP})$ , o que para o caso dos parâmetros utilizados varia de 10 até 90 anticorpos (sendo mais comuns valores entre 50 e 70).

Dos comportamentos descritos para cada um destes modelos é possível tirar algumas conclusões a respeito do tempo computacional de cada modelo. No caso do  $\text{ClonalgII}_{inf}$  (*Aging Puro*), devido aos altos picos de mortes, é constante a necessidade de inserção de novos anticorpos gerados de maneira aleatória, o que aumenta o tempo computacional do algoritmo e provoca os picos de diversidade e queda dos valores médios de afinidade. Além disto, como o operador de *Aging Puro* submete todos os anticorpos à etapa de maturação fraca de afinidade antes de matá-los, isto implica que vários anticorpos passam pela busca local (que consome bastante tempo computacional) e morrem logo em seguida. Já o *Aging Fuzzy* seleciona alguns anticorpos para morrerem direto sem nem passar pela etapa de maturação fraca de afinidade, evitando gastar recursos com buscas locais desnecessárias.

O modelo  $\text{ClonalgIII}_{inf-BT}$  é considerado o melhor entre todos os modelos baseado no princípio da seleção clonal proposto. Isto se deve ao fato de que o  $\text{ClonalgIII}_{inf-BT}$  é capaz de atingir melhores valores de energia, de manter uma boa diversidade populacional e de selecionar melhor os anticorpos que devem morrer mantendo a afinidade média melhor e mais constante.

## 6.4 Resultados dos Modelos Baseados na Rede Imune

Os resultados obtidos com os modelos baseados na Teoria da Rede Imune são sumarizados na Tabela 6.6 e nas Figuras 6.4 e 6.5. A Tabela 6.6 apresenta os dados referentes ao número do problema (instância) tratado ( $N^o$ ), tamanho da referida proteína (L), melhor valor de energia encontrado (Melhor), média (Média) e desvio padrão ( $\sigma$ ) obtidos para as 50 execuções e tempo médio em minutos gasto durante as 50 execuções ( $T(\text{min})$ )<sup>4</sup>.

A Figura 6.4 apresenta o valor percentual das médias obtidas pelos modelos propostos e os ótimos reportados na literatura para cada uma das instâncias. Já a Figura 6.5 representa o tempo médio em minutos gasto durante uma execução

<sup>4</sup>As sementes geradoras de números aleatórios são diferentes para cada uma das 50 execuções

dos algoritmos para cada uma das instâncias (o gráfico é apresentado em escala logarítmica). Ambas as figuras são referentes aos dados coletados ao longo das 50 execuções.

Tabela 6.6: Comparação entre os modelos de rede implementados

		Rede I				Rede II			
Nº	L	Melhor	Média	$\sigma$	T(min)	Melhor	Média	$\sigma$	T(min)
1	20	-10	-9.40	0.49	0.60	-10	-9.36	0.48	0.12
2	24	-12	-11.56	0.54	0.50	-11	-9.94	0.51	0.24
3	25	-9	-8.96	0.28	1.91	-9	-7.61	0.94	0.25
4	36	-18	-16.02	1.13	1.62	-18	-13.86	1.15	1.90
5	48	-28	-25.94	1.07	34.01	-21	-18.08	1.35	25.78
6	50	-26	-23.75	1.68	23.18	-19	-16.52	1.15	100.75
7	60	-49	-43.98	2.22	38.14	-38	-33.29	2.36	285.91
		Rede III				Rede I <sub>inf</sub>			
Nº	L	Melhor	Média	$\sigma$	T(min)	Melhor	Média	$\sigma$	T(min)
1	20	-10	-8.30	0.61	0.17	-10	-9.52	0.50	0.03
2	24	-10	-7.94	0.62	0.64	-12	-11.62	0.57	0.03
3	25	-7	-6.15	0.57	0.49	-9	-8.95	0.30	0.04
4	36	-13	-10.32	0.84	0.14	-18	-15.64	1.41	0.10
5	48	-17	-14.38	1.12	3.16	-28	-25.38	1.70	0.45
6	50	-16	-13.02	1.00	7.45	-26	-23.24	1.29	0.57
7	60	-29	-26.57	1.68	17.50	-48	-43.54	2.17	3.18
		RedeI <sub>inf-Rec</sub>				RedeII <sub>inf</sub>			
Nº	L	Melhor	Média	$\sigma$	T(min)	Melhor	Média	$\sigma$	T(min)
1	20	-10	-9.54	0.50	0.15	-10	-9.40	0.49	0.06
2	24	-12	-10.70	0.86	0.23	-12	-11.64	0.53	0.09
3	25	-9	-7.80	0.99	0.25	-9	-8.34	0.94	0.08
4	36	-18	-16.00	1.21	0.75	-18	-15.92	1.05	0.15
5	48	-28	-24.02	2.05	1.96	-28	-25.00	1.83	0.28
6	50	-27	-22.40	2.21	2.56	-27	-23.25	1.61	1.91
7	60	-47	-43.18	2.00	7.34	-48	-43.30	2.36	4.48
		RedeIII <sub>inf</sub>				RedeI <sub>inf-BT</sub>			
Nº	L	Melhor	Média	$\sigma$	T(min)	Melhor	Média	$\sigma$	T(min)
1	20	-9	-8.32	0.51	0.04	-10	-9.74	0.44	0.06
2	24	-10	-8.34	0.69	0.13	-12	-11.90	0.30	0.10
3	25	-9	-6.34	0.69	0.13	-9	-8.96	0.28	0.07
4	36	-18	-16.86	1.07	0.37	-18	-16.86	1.07	0.37
5	48	-19	-10.78	2.33	1.66	-28	-26.32	0.94	1.50
6	50	-15	-9.93	2.59	1.65	-29	-25.12	1.61	1.42
7	60	+1	+1	-	0.75	-49	-45.16	2.64	6.43

Para averiguar a distinção entre os diferentes modelos foram realizados testes estatísticos com significância de 95%. Os testes realizados foram os mesmos efetuados nos modelos baseados no Clonalg.

Tabela 6.7: Hipóteses de que as distribuições sejam diferentes (h) com seus respectivos p-valores obtidos nas simulações dos modelos de Rede Imune

	<b>RedeI × RedeII</b>		<b>RedeII × RedeIII</b>	
<b>Proteína</b>	<b>h</b>	<b>p-valores</b>	<b>h</b>	<b>p-valores</b>
1	F	0.6818	V	1.1133e-012
2	V	0	V	0
3	V	5.8196e-012	V	4.2188e-015
4	V	7.3719e-013	V	0
5	V	1.0125e-012	V	3.5281e-012
6	V	3.9297e-018	V	0
7	V	1.9778e-005	V	0.0025
	<b>RedeI × RedeI<sub>inf</sub></b>		<b>RedeI<sub>inf</sub> × RedeI<sub>inf-Rec</sub></b>	
<b>Proteína</b>	<b>h</b>	<b>p-valores</b>	<b>h</b>	<b>p-valores</b>
1	F	0.2310	F	0.8420
2	F	0.4768	V	1.2437e-007
3	F	idênticas	V	3.4508e-009
4	F	0.1662	F	0.2512
5	F	0.1251	V	6.4509e-004
6	V	0.0400	V	0.0194
7	F	0.2620	F	0.4537
	<b>RedeI<sub>inf</sub> × RedeI<sub>inf-BT</sub></b>			
<b>Proteína</b>	<b>h</b>	<b>p-valores</b>		
1	V	0.0234		
2	V	0.0035		
3	F	0.9275		
4	V	1.1882e-005		
5	V	0.0018		
6	V	6.6516e-008		
7	V	5.0031e-004		

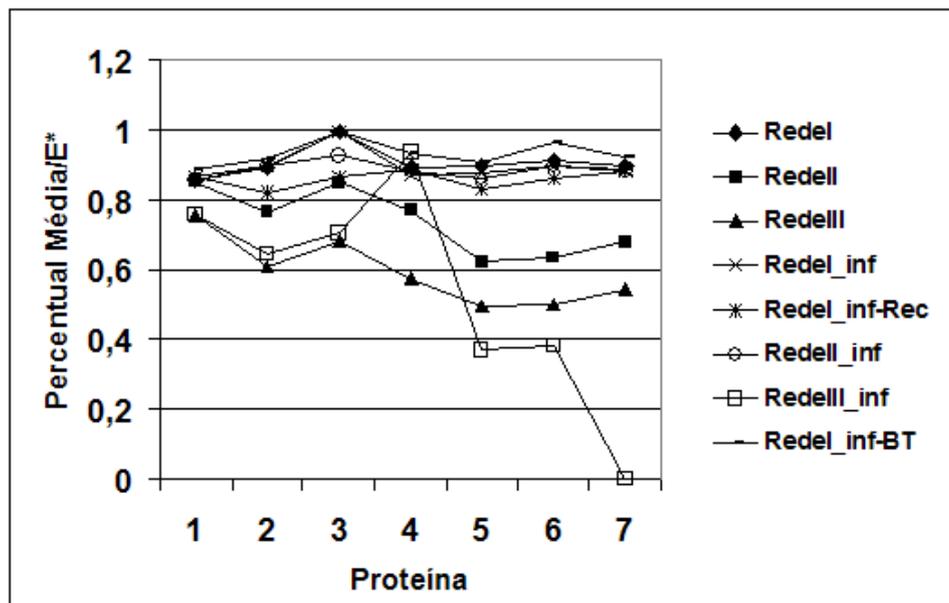


Figura 6.4: Gráfico representando a diferença percentual da média dos modelos baseados na teoria da Rede Imune e do ótimo conhecido na literatura

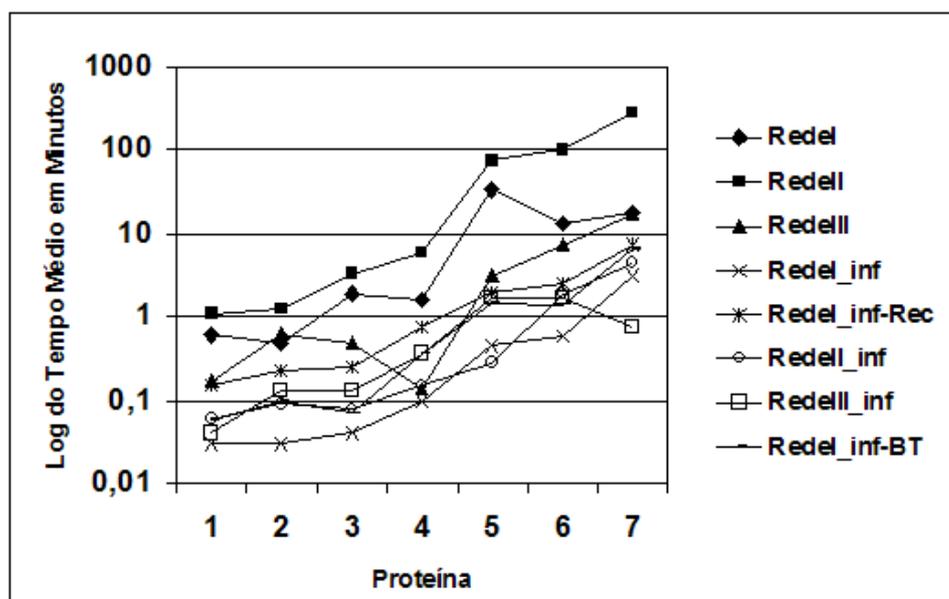


Figura 6.5: Gráfico do tempo que cada modelo baseado na teoria da Rede Imune leva para uma execução

Para aferir a eficiência do operador de *Aging* Puro foram comparados os modelos RedeI e RedeII, assim como os modelos RedeI<sub>inf</sub> e RedeII<sub>inf</sub> (ver Figura 6.4). Para todas as instâncias, a RedeI foi capaz de obter melhores médias do que a RedeII, com exceção da primeira instância na qual os dois modelos não possuem diferença significativa. Analisando os modelos que trabalham em espaços de soluções ineficazes (RedeI<sub>inf</sub> x RedeII<sub>inf</sub>) não existem diferenças significativas nas médias. Com isto, pode-se afirmar que o uso do operador de *Aging* Puro não traz ganhos de desempenho aos modelos. Os dois modelos sem *aging* foram mais rápidos, exceto para a instância 5 conforme observado na Figura 6.5.

Quando os modelos que trabalham com operadores de *Aging* Puro foram contrastados com os modelos que trabalham com operadores de *Aging Fuzzy* houve uma predominância de melhores resultados nos modelos com operador de *Aging* Puro. A Tabela 6.6 mostra que apenas para a quarta instância uma rede com *Aging Fuzzy* (RedeIII<sub>inf</sub>) é significativamente superior à sua contrapartida com *Aging* Puro (RedeII<sub>inf</sub>).

Para que fosse possível observar o impacto que a permissão de anticorpos com colisões (ineficazes) no espaço de busca iria causar os três primeiros modelos foram contrastados com seus modelos correspondentes trabalhando com espaço de soluções ineficazes (RedeI x RedeI<sub>inf</sub>, RedeII x RedeII<sub>inf</sub> e RedeIII x RedeIII<sub>inf</sub>). Ao comparar a RedeI com a RedeI<sub>inf</sub> não existem diferenças significativas entre as médias para nenhuma das instâncias, exceto para a sexta instância. O confronto entre os modelos RedeII e RedeII<sub>inf</sub> demonstra que a RedeII<sub>inf</sub> possui médias significativamente melhores do que a RedeII, exceto para a primeira instância. Para as quatro primeiras instâncias o modelo RedeIII<sub>inf</sub> possui melhores valores médios de energia do que a RedeIII, porém para as últimas três instâncias o desempenho deste modelo cai drasticamente, sendo que para a última instância a RedeIII<sub>inf</sub> é incapaz de encontrar um dobramento eficaz durante todas as 50 execuções (ver Figura 6.4). Os modelos de rede trabalhando com espaço de soluções ineficazes demandam um tempo computacional menor do que aquelas redes que trabalham com espaço apenas de soluções eficazes, conforme pode ser observado na Figura 6.5.

As Redes I<sub>inf</sub> e I<sub>inf-Rec</sub> foram comparadas para determinar o efeito do uso da recombinação nos modelos de Rede. A Tabela 6.6 mostra que o modelo que utiliza a geração aleatória de anticorpos possui melhor desempenho para 5 das instâncias, sendo 4 estatisticamente significativas, e perde em 2 instâncias (sem diferenças sig-

nificativas). Isto leva a crer que o uso de recombinação não melhora o desempenho dos modelos baseados em Rede.

A hibridização do modelo  $\text{RedeI}_{inf}$  com a metaheurística Busca Tabu, a qual resultou no modelo  $\text{RedeI}_{inf-BT}$ , foi capaz de melhorar os resultados deste modelo. A  $\text{RedeI}_{inf-BT}$  possui melhores valores de energia (estatisticamente significativos) para todas as instâncias, exceto para a terceira instância. A Figura 6.5 mostra que, no geral, o modelo executa em um tempo computacional maior do que a  $\text{RedeI}_{inf}$ .

Portanto, o modelo que não faz uso de nenhum operador de *aging* e possui as etapas de maturação fraca da afinidade e maturação intensiva da afinidade apresenta os melhores resultados dentre os modelos baseados na teoria da rede imune.

## 6.5 Clonalg versus Rede Imune

Esta seção apresenta uma comparação entre os melhores modelos baseados no Clonalg e na Rede Imune:  $\text{ClonalgIII}_{inf-BT}$  e  $\text{RedeI}_{inf-BT}$ , respectivamente. Por conveniência os resultados obtidos por ambos os modelos são rerepresentados na Tabela 6.8.

Tabela 6.8: Comparação entre  $\text{ClonalgIII}_{inf-BT}$  e a  $\text{RedeI}_{inf-BT}$

N.	$E^*$	$\text{ClonalgIII}_{inf-BT}$				$\text{RedeI}_{inf-BT}$				p-valores
		Melhor	Média	$\sigma$	T(min)	Melhor	Média	$\sigma$	T(min)	
1	-11	-11	<b>-11</b>	0	0.98	-10	-9.74	0.44	<b>0.06</b>	2.0342e-021
2	-13	-13	<b>-12.98</b>	0.14	1.59	-12	-11.90	0.30	<b>0.10</b>	1.2827e-021
3	-9	-9	<b>-9</b>	0	1.28	-9	-8.96	0.28	<b>0.07</b>	0.3173
4	-18	-18	<b>-17.76</b>	0.59	2.75	-18	-16.86	1.07	<b>0.37</b>	4.4283e-006
5	-29	<b>-30</b>	<b>-28.49</b>	0.92	5.83	-28	-26.32	0.94	<b>1.50</b>	2.0088e-015
6	-26	<b>-30</b>	<b>-27.36</b>	1.01	11.17	-29	-25.12	1.62	<b>1.42</b>	4.0379e-005
7	-49	<b>-51</b>	<b>-46.16</b>	1.49	13.83	-49	-45.16	2.64	<b>6.43</b>	0.1103

Conforme pode ser observado na tabela, todas as médias obtidas pelo modelo baseado no Clonalg são melhores do que aquelas obtidas pelo modelo baseado na teoria da rede imune, sendo que apenas os resultados obtidos nas instâncias 3 e 7 não são significativamente diferentes.

No que diz respeito ao tempo computacional despendido na execução dos modelos, a  $\text{RedeI}_{inf-BT}$  é mais rápida do que o  $\text{ClonalgIII}_{inf-BT}$ .

Portanto, o modelo proposto que, no geral, apresentou melhores resultados, ao

custo de um tempo computacional maior, foi o  $\text{ClonalgIII}_{inf-BT}$  e na seção seguinte este modelo será confrontado com outros modelos reportados na literatura.

## 6.6 Comparação Entre o Modelo Proposto e Trabalhos da Literatura

Dentre todos os modelos desenvolvidos o que apresentou melhores resultados foi aquele que uniu as técnicas de SIA, SIF e BT, trabalhado com a permissão de anticorpos factíveis na população<sup>5</sup> - o  $\text{ClonalgIII}_{inf-BT}$ . O modelo  $\text{ClonalgIII}_{inf-BT}$  pode ser favoravelmente comparado com trabalhos reportados na literatura, como mostra a Tabela 6.9. As estruturas geradas por este modelo podem ser vistas no Apêndice B. Não foi possível a realização de testes estatísticos para verificar a significância dos resultados encontrados com relação ao encontrados pelo Aging-SIA (CUTELLO et al., 2005) e pelo Backtracking-EA (COTTA, 2003), pois a distribuição dos dados gerados por estes autores não é conhecida.

Tabela 6.9: Comparação entre  $\text{ClonalgIII}_{inf-BT}$ , Backtracking-EA (COTTA, 2003) e Aging-SIA (CUTELLO et al., 2005)

N.	$E^*$	$\text{ClonalgIII}_{inf-BT}$				Backtracking-EA			Aging-SIA		
		Pior	Melhor	Média	$\sigma$	Melhor	Média	$\sigma$	Melhor	Média	$\sigma$
1	-11	-11	-11	-11	0	-11	-10.32	0.61	-11	-11	0
2	-13	-12	-13	-12.98	0.14	-13	-10.90	0.98	-13	<b>-13</b>	0
3	-9	-9	-9	-9	0	-9	-7.98	0.71	-9	-9	0
4	-18	-16	-18	<b>-17.76</b>	0.59	-18	-14.38	1.26	-18	-16.76	1.02
5	-29	-26	<b>-30</b>	<b>-28.49</b>	0.92	-25	-20.80	1.61	-29	-25.16	0.45
6	-26	-25	<b>-30</b>	<b>-27.36</b>	1.01	-23	-20.20	1.50	-23	-22.60	0.40
7	-49	-42	<b>-51</b>	<b>-46.16</b>	1.49	-39	-34.18	2.31	-41	-39.28	0.24

Mesmo assim, pode-se notar que o modelo  $\text{ClonalgIII}_{inf-BT}$  encontra melhores valores de energia (Melhor) iguais ou superiores àqueles encontrados pelos algoritmos Backtracking-EA e Aging-SIA. Com relação aos valores médios de energia o modelo  $\text{ClonalgIII}_{inf-BT}$  tem melhores valores do que os obtidos pelo algoritmo Backtracking-EA para todas as instâncias e é capaz de produzir melhores valores do que o algoritmo Aging-SIA para quatro instâncias (4, 5, 6 e 7) e piores valores

<sup>5</sup>Vale lembrar que todas as soluções encontradas correspondem a conformações factíveis.

apenas para a segunda instância.

Cabe salientar ainda que para as instâncias 5, 6 e 7 o modelo proposto foi capaz de melhorar os valores de  $E^*$  (ver Tabela 6.1). Além disso, os piores valores de energia encontrados para as proteínas de tamanho 50 e 60 (-25 e -42, respectivamente) são superiores aos melhores valores de energia encontrados pelos dois trabalhos utilizados para comparação, com pode ser observado na Figura 6.6.

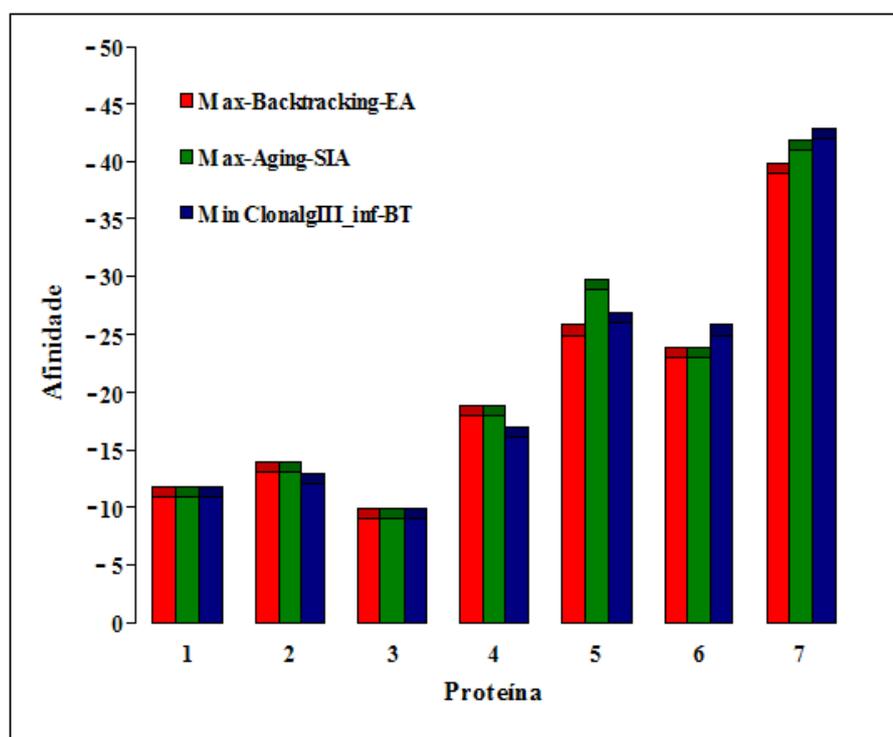


Figura 6.6: Relação entre as melhores afinidades encontradas pelos algoritmos Backtracking-EA (COTTA, 2003) e Aging-SIA (CUTELLO et al., 2005) e as piores afinidades encontradas pelo melhor modelo proposto neste trabalho -  $ClonalgIII_{inf-BT}$

---

## *Conclusões e Trabalhos Futuros*

---

---

### 7.1 Conclusões

Este trabalho apresentou diversos modelos computacionais para o Problema da Predição da Estrutura de Proteínas no modelo Hidrofóbico-Polar Tri-Dimensional. Os modelos foram baseados em dois importantes aspectos dos sistemas imunológicos: o princípio da seleção clonal e a teoria da rede imune. Para cada uma das inspirações biológicas (seleção clonal e teoria da rede imune), foram desenvolvidos modelos com e sem operadores de *aging* (que podem ser Puros ou *Fuzzy*), com e sem o uso de penalidades para anticorpos inactiváveis e com e sem estágio de maturação intensiva da afinidade, com o intuito de aferir o impacto de cada uma destas possíveis melhorias no algoritmo proposto.

Os primeiros modelos desenvolvidos, que foram baseados no princípio da seleção clonal, são variações do algoritmo Clonalg enquanto os últimos modelos, mais complexos e baseados na teoria da rede imune, possuem semelhanças tanto com a opt-Ainet quanto com a copt-Ainet.

Os resultados obtidos pelos modelos baseados no Clonalg foram melhores do que

os obtidos pelos modelos análogos baseados na teoria da rede. Existem duas possíveis razões principais para este comportamento: menor poder computacional do modelo baseado na teoria da rede ou um ajuste não adequado dos parâmetros. Visto que o modelo baseado na rede é mais complexo é difícil de acreditar que os resultados sejam decorrentes de uma falta de poder computacional, sendo a segunda hipótese muito mais provável. Como a rede imune proposta possui muitos parâmetros e é bastante sensível a estes (assim como o que ocorre com a maioria das redes imunes propostas na literatura) é bastante compreensível que os parâmetros encontrados durante a fase de ajuste de parâmetros sejam bem distantes dos ótimos. Contudo, o problema de encontrar valores ótimos para os parâmetros é um problema bastante complexo e que deve ser considerado na escolha de um modelo computacional, fazendo com que os modelos baseados no Clonalg sejam preferidos em detrimento aos modelos baseados na teoria da rede imune para resolver o Problema da Predição da Estrutura de Proteínas (PPEP).

No que tange ao uso de espaço de busca factível ou infactível pode-se afirmar que o uso do espaço de busca infactível é benéfico, pois, com exceção da RedeIII<sub>inf</sub>, resulta em soluções de qualidade comparável (em muitos dos casos superior) a sua contrapartida em espaço apenas com anticorpos factíveis e realiza isto num tempo computacional bastante inferior.

O uso de um estágio de maturação intensiva da afinidade, implementado na forma de uma Busca Tabu, colabora para a melhoria da qualidade dos resultados obtidos, mas a custa de um acréscimo no tempo despendido para encontrá-los. Portanto, para os casos onde a qualidade da solução é mais importante do que o tempo despendido para obtê-la, recomenda-se fortemente o uso deste estágio.

Os operadores de *aging*, em particular o operador de *Aging Fuzzy*, representam um ponto de dissonância entre os resultados obtidos pelos modelos baseados no Clonalg e os modelos baseados na teoria da rede imune. Enquanto nos modelos baseados no Clonalg o uso de operadores de *aging*, e mais especificamente de *Aging Fuzzy*, são benéficos do ponto de vista estatístico com uma confiabilidade de 95%, no caso dos modelos baseados na teoria da rede ocorre exatamente o contrário: os operadores de *aging* pioram os resultados, sendo que os modelos que utilizam o operador de *Aging Fuzzy* obtêm valores ainda inferiores àqueles obtidos pelos modelos que fazem uso do operador de *Aging Puro*. Isto se deve a dois fatores fundamentais: os modelos baseados na teoria da rede já possuem um mecanismo

de eliminação de anticorpos não interessantes (a supressão) mesmo na ausência dos operadores de *aging* e o operador de *Aging Fuzzy* teve seus parâmetros ajustados para funcionar nos modelos baseados no algoritmo Clonalg. Pode ser que o ajuste destes parâmetros leve a resultados melhores, contudo acredita-se que o desenvolvimento de um mecanismo eficiente de supressão para os modelos baseados na teoria da rede imune possa contribuir muito mais para a melhoria dos resultados das redes.

O melhor modelo encontrado foi o ClonalgIII<sub>inf-BT</sub> que agrega as três técnicas propostas neste trabalho: espaço de busca inactível com penalidades, operador de *Aging Fuzzy* e estágio de maturação intensiva da afinidade (através de uma Busca Tabu). Quando comparado a trabalhos reportados na literatura, este modelo demonstrou resultados semelhantes ou superiores, posicionando-o como uma boa opção para o dobramento de proteínas no modelo 3D HP. Em alguns casos este modelo foi capaz de aprimorar os resultados conhecidos na literatura.

## 7.2 Trabalhos Futuros

Embora os resultados obtidos pelos modelos propostos sejam satisfatórios, melhorias podem ser realizadas de forma a explorar com mais eficiência as qualidades apontadas e superar as deficiências remanescentes. Sendo que superar as deficiências é considerada prioridade em trabalhos futuros.

Como o principal problema enfrentado neste trabalho diz respeito à sensibilidade dos parâmetros dos modelos baseados na teoria da rede imune, uma auto-adaptação eficiente destes parâmetros poderia não somente facilitar o uso dos modelos de redes, como também resultar numa melhor qualidade das estruturas tri-dimensionais encontradas pelos mesmos. Apesar desta deficiência ser mais acentuada nos modelos de rede, os modelos baseados no Clonalg também poderiam se beneficiar da auto-adaptação que, na pior das hipóteses, facilitaria o uso da técnica. Além disso, pode-se considerar a co-evolução das populações do SIA e do SIF.

No que tange às qualidades já identificadas, um ponto importante a ser explorado em trabalhos futuros diz respeito ao ajuste do Sistema de Inferência Fuzzy utilizado no operador de *Aging Fuzzy*. Este Sistema poderia ser gerado com o auxílio de uma técnica computacional eficiente como as propostas em (DELGADO, 2002) e (ALVES et al., 2004b) ou, ainda melhor, auto-ajustados ao longo do processo evolucionário,

permitindo que eles se adequassem não somente ao problema sendo resolvido mas também as características apresentada pela população corrente.

Um outro ponto que merece ser aprofundado em trabalhos futuros é a melhoria da metaheurística utilizada no estágio de maturação intensiva da afinidade. Neste quesito a Busca Tabu utilizada poderia incorporar técnicas mais sofisticadas como, por exemplo, o uso de memórias de médio e longo prazo, construção de vocabulário, reatividade, entre outras (GLOVER, 1995). Além disto, outras metaheurísticas poderiam ser exploradas, como por exemplo a Filter-and-Fan, que foi recentemente utilizada no PPEP no modelo 2D HP e obteve resultados bastante impressionantes tanto na qualidade das soluções quanto, principalmente, no tempo gasto para obtê-las, tornando a técnica muito atrativa como mecanismo de intensificação da busca (REGO; LI; GLOVER, ).

Em trabalhos futuros também deveriam ser considerados modelos de representação mais próximo das proteínas reais; como, por exemplo, os modelos livres de *lattices*.

Além disso, poderiam ser consideradas outras formas de distribuição para as variáveis aleatórias. Também pode ser considerada a utilização de outras equações para o operador de *aging*.

---

## *Glossário*

---

---

**Afinidade:** é uma medida de quão apto um anticorpo está para detectar e combater um antígeno. Em Sistemas Imunológicos Artificiais para otimização a afinidade corresponde a minimização ou maximização de uma função.

**Aminoácido:** classe de compostos orgânicos que contêm um grupamento carboxila e um grupamento amina.

**Anticorpo:** gamaglobulina formada como resposta a estímulo imunogênico e capaz de interagir com o antígeno que levou à sua síntese, ou com outro estreitamente relacionado com ele.

**Antígeno:** qualquer substância a que, em circunstâncias favoráveis, um indivíduo é exposto, e que pode produzir uma resposta imunológica específica, com a formação de anticorpos específicos ou de linfócitos T especificamente sensibilizados, ou ambos.

**Atividade catalítica:** é uma atividade que, em geral, aumenta a velocidade de uma reação química.

**Bioinformática:** é uma área da ciência que une conceitos biológicos e ferramentas computacionais. Essas ferramentas tem por objetivo analisar, armazenar, organizar, visualizar e simular dados ou eventos biológicos.

**Clone:** conjunto de células ou organismos originários de outros por algum tipo de multiplicação assexual (divisão, enxertia, apomixia, etc.).

**Cristalografia:** é uma ciência experimental para a determinação do arranjo de átomos em sólidos. É o método primário utilizado na determinação da estrutura de proteínas.

**Desnaturar:** é modificar a estrutura de uma proteína pelo uso de calor ou outros agentes, físicos e químicos, destruindo ou alterando suas propriedades fisiológicas e físicas, mas sem alterar a seqüência de seus aminoácidos constituintes.

**Energia livre:** é a diferença entre a energia interna de um sistema e o produto da sua temperatura absoluta e entropia. Mede o nível de desordem de uma molécula, quanto menor a energia livre, menor é a desordem.

**Espectroscopia:** é uma técnica utilizada para determinar a presença e a magnitude de cargas elétricas em um sólido.

**Genoma:** é a constituição genética total de um indivíduo.

**Hidrofobicidade:** é o grau de dificuldade que uma molécula tem de dissolver-se ou ser dissolvida na água.

**Lattice:** é um grupo onde os elementos estão regularmente espaçados. Um *lattice* pode ser formado pela repetição de formas geométricas regulares e, geralmente, são essas formas que dão nome ao *lattice*, por exemplo, um *lattice* triangular é formado pela justaposição de triângulos equiláteros, um *lattice* cúbico é formado pela justaposição de cubos e assim por diante. Qualquer forma regular pode dar origem a um *lattice*.

**Ligação Covalente:** é uma ligação formada pelo compartilhamento de um ou mais pares de electrons entre átomos.

**Ligação Iônica:** é uma ligação química entre dois íons de cargas opostas.

**Microscopia:** é uma técnica que produz imagens visíveis de estruturas ou detalhes que de outra forma não seriam visíveis ao olho humano. Faz uso de microscópios para isso.

**Molécula:** grupamento estável de dois ou mais átomos, que caracteriza quimicamente uma certa substância.

**Patógeno:** é um agente causador de doenças, geralmente um microorganismo

vivo como uma bactéria ou um fungo.

**Proteína:** classe de compostos orgânicos de carbono, nitrogênio, oxigênio e hidrogênio, que constituem o principal componente dos organismos vivos.

**Polímero:** Composto formado por sucessivas aglomerações de grande número de moléculas fundamentais.

**Resíduo de aminoácido:** é uma unidade de uma proteína. Recebe esse nome porque um aminoácido perde alguns átomos para fazer parte de uma proteína.

**Resposta Imune:** é o conjunto de fenômenos com que o organismo reage ao contato com antígenos.

---

## *Referências Bibliográficas*

---

---

- ABDI, H.; MOLIN, P. Lilliefors test of normality. In: SALKIND, N. (Ed.). *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA, USA: Sage Publications, 2007.
- ALBERTS, B. et al. *Biologia Molecular da Célula*. 3. ed. Porto Alegre, BR: Artes Médicas, 1997.
- ALBERTS, B. et al. *Molecular Biology of the Cell*. 4. ed. New York, USA: Garland Publishing, 2002.
- ALBERTS, B. et al. *Molecular Biology of the Cell, Fourth Edition*. 5. ed. USA: Garland Publishing, 2005.
- ALMEIDA, C. P.; GONÇALVES, R. A.; DELGADO, M. R. A hybrid immune-based system for the protein folding problem. In: COTTA, C.; HEMER, J. V. (Ed.). *Aceito para publicação em Seventh European Conference on Evolutionary Computation in Combinatorial Optimisation - EvoCOP 2007*. Valencia, Spain: Springer, 2007. (Lecture Notes in Computer Science).
- ALVES, R. T. et al. An artificial immune system for fuzzy-rule induction in data mining. In: YAO, X. et al. (Ed.). *8th International Conference on Parallel Problem Solving from Nature*. Birmingham, UK: Springer-Verlag, 2004a. (Lecture Notes in Computer Science, v. 3242), p. 1011–1020.
- ALVES, R. T. et al. An artificial immune system for multimodality image alignment. In: SBRN. *8º Simpósio Brasileiro de Redes Neurais*. Maranhão, BR, 2004b. p. 11–21.
- ANCHOR, K. P. et al. Extending the computer defense immune system: Network intrusion detection with a multiobjective evolutionary programming approach. In: ICARIS. *1st International Conference on Artificial Immune Systems*. University of Kent at Canterbury, 2002. p. 12–21.

- ANILE A. M. CUTELLO, V. et al. Direct and immunological computing with quasi-newton method for the protein folding problem. *International Journal Natural Computing*, 2006.
- BERGER, B.; LEIGHTON, T. Protein folding in the hydrophobic-hydrophilic model is np complete. *Journal of Computational Biology*, v. 5, p. 27–40, 1998.
- BERGERON, B. *Bioinformatics Computing*. San Francisco, CA, USA: Prentice Hall, 2003.
- BERSINI, H. Artificial immune systems and their applications. In: DASGUPTA, D. Berlin, German: Springer-Verlag, 1999. cap. The Endogenous Double Plasticity of the Immune Network and the Inspiration to the Drawn for Engineering Aircrafts, p. 22–44.
- BERSINI, H.; VARELA, F. J. Hint for adaptive problem solving gleaned from immune networks. In: PPSN. *Parallel Problem Solving from Nature*. Dortmund, FRG, 1990. p. 343–354.
- BLAZEWICZ, J.; LUKASIAK, P.; MILOSTAN, M. Application of tabu search strategy for finding low energy structure of protein. *Artificial Intelligence in Medicine*, v. 35, p. 135–145, 2005.
- BRADLEY, D. W.; TYRRELL, A. M. Immunotronics: Hardware fault tolerance inspired by the immune system. In: ICES. *International Conference on Evolvable Systems*. Edinburgh, Scotland, UK, 2000. p. 11–20. Disponível em: <citeseer-ist.psu.edu/bradley00immunotronics.html>.
- CAMPELO, F. et al. A modified immune network algorithm for multimodal electromagnetic problems. *IEEE Transactions on Magnetics*, v. 42, n. 4, p. 1111–1114, 2006.
- CANHAM, R.; JACKSON, A.; A., T. Robot error detection using an artificial immune system. In: EH. *NASA/DoD Conference Evolvable Hardware*. Illinois, USA, 2003. p. 199–207.
- CHU, D.; TILL, M.; ZOMAYA, A. Y. Parallel ant colony optimization for 3d protein structure prediction using the hp lattice model. In: IPDPS. *19th International Parallel and Distributed Processing Symposium*. Nice, France, 2005.
- COELLO, C. A. C.; CORTES, N. C. Solving multiobjective optimization problems using an artificial immune system. *Genetic Programming and Evolvable Machines*, v. 6, n. 2, p. 163–190, Junho 2005.
- CONOVER, W. J. *Practical Nonparametric Statistics*. USA: John Wiley & Sons, 1980.
- COSTA, A. M. et al. Makespan minimization on parallel processors: An immune-based approach. In: TIMMIS, J.; BENTLEY, P.; HART, E. (Ed.). *Special Sessions on Artificial Immune Systems in the Congress on Evolutionary Computation*. Honolulu, Hawaii, 2002. (Lecture Notes in Computer Science, v. 2787), p. 11–21.

- COTTA, C. Protein structure prediction using evolutionary algorithms hybridized with backtracking. *7th International Work-Conference on Artificial and Natural Neural Networks*, v. 2687, p. 321–328, 2003.
- CRESCENZI, P. et al. On the complexity of protein folding. *Journal of Computational Biology*, v. 5, n. 3, p. 423–466, 1998.
- CUTELLO, V. et al. Immune algorithms with aging operators for the string folding problem and the protein folding problem. In: EVOCOP. *Evolutionary Computation in Combinatorial Optimization*. Lausanne, Switzerland, 2005. p. 80–90.
- CUTELLO, V.; NARZISI, G.; NICOSIA, G. A multi-objective evolutionary approach to the protein structure prediction problem. *Journal of the Royal Society Interface*, v. 3, n. 6, p. 139–151, 2006.
- CUTELLO, V.; NICOSIA, G.; PAVONE, M. A hybrid immune algorithm with information gain for the graph coloring problem. In: GECCO. *Genetic and Evolutionary Computation Conference*. Illinois, USA, 2003. (Lecture Notes in Computer Science, v. 2723), p. 171 ff.
- CUTELLO, V.; NICOSIA, G.; PAVONE, M. Exploring the capability of immune algorithms: A characterization of hypermutation operators. In: ICARIS. *Third International Conference on Artificial Immune Systems*. Catania, Italy, 2004. p. 263–276.
- CUTELLO, V. et al. An immune algorithm for protein structure prediction on lattice models. *IEEE Transactions on Evolutionary Computation*, v. 10, 2006.
- DASGUPTA, D. *Artificial Immune Systems and Their Applications*. Secaucus, NJ, USA: Springer, 1998.
- DE CASTRO, L.; TIMMIS, J. An artificial immune network for multimodal function optimization. In: CEC. *IEEE Congress on Evolutionary Computation*. Kent University, 2002.
- DE CASTRO, L.; TIMMIS, J. *Artificial Immune Systems: A New Computational Approach*. London. UK.: Springer-Verlag, 2002. 357 p. ISBN 1852335947. Disponível em: <<http://www.cs.ukc.ac.uk/pubs/2002/1507>>.
- DE CASTRO, L. N. *Engenharia Imunológica: Desenvolvimento e Aplicação de Ferramentas Computacionais Inspiradas em Sistemas Imunológicos Artificiais*. Tese (Doutorado) — Universidade Estadual de Campinas - Faculdade de Engenharia Elétrica e de Computação, 2001.
- DE CASTRO, L. N. *Fundamentals of Natural Computing: Basic Concepts, Algorithms, and Applications*. 1. ed. USA: Chapman Hall/CRC, 2006.
- DE CASTRO, L. N.; TIMMIS, J. *Artificial Immune Systems: A New Computational Intelligence Approach*. Great Britain: Springer, 2002.
- DE CASTRO, L. N.; TIMMIS, J. A simple model for the immune network. In: CEC. *Congress on Evolutionary Computation*. Honolulu, Hawaii, 2002. v. 1, p. 699–704.

- DE CASTRO, L. N.; VON ZUBEN, F. J. *Artificial Immune Systems: Part II - A Survey of Applications*. Campinas, SP, BR, Fevereiro 2000.
- DE CASTRO, L. N.; VON ZUBEN, F. J. The clonal selection algorithm with engineering applications. In: GECCO. *Workshop on Artificial Immune Systems and Their Applications*. Las Vegas, Nevada, USA, 2000. p. 36–37.
- DE CASTRO, L. N.; VON ZUBEN, F. J. Learning and optimization using the clonal selection principle. *IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems*, v. 6, n. 3, p. 239–251, 2002.
- DE PAULA, F. S.; DE CASTRO, L. N.; GEUS, P. L. de. An intrusion detection system using ideas from the immune system. In: CEC. *International Conference on Evolutionary Computation*. San Diego, CA, USA, 2004. p. 1059–1066.
- DELGADO, M. R. *Projeto Automático de Sistemas Nebulosos: Uma Abordagem Co-Evolutiva*. Tese (Doutorado) — Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, Campinas, BR, 2002.
- DING, Y.; REN, L. Fuzzy self-tuning immune feedback controller for tissue hyperthermia. In: FUZZ. *IEEE International Conference on Fuzzy Systems*. San Antonio, Texas, 2000. v. 1, p. 534–538.
- FERREIRA, A. B. de H. *Dicionário Aurélio Eletrônico - Século XXI*. 1999. Editora Nova Fronteira.
- FRANÇA, F. O.; GOMES, L. C. T. Handling time-varying tsp instances. In: CEC. *Congress on Evolutionary Computation*. Vancouver, BC, Canada, 2006. p. 9735–9742.
- FRANÇA, F. O.; VON ZUBEN, F. J.; DE CASTRO, L. N. An artificial immune network for multimodal function optimization on dynamic environments. In: GECCO. *Genetic and Evolutionary Computation*. Washington DC, USA, 2005. p. 289–296.
- GALEANO, J. C.; VELOZA-SUAN, A.; GONZÁLEZ, F. A. A comparative analysis of artificial immune network models. In: GECCO. *Genetic and Evolutionary Computation Conference*. Washington DC, USA, 2005. p. 361–368.
- GARRETT, S. Parameter-free, adaptive clonal selection. In: CEC. *Congress on Evolutionary Computation*. San Diego, CA, USA, 2004. v. 1, p. 1052–1058.
- GENDREAU, M. *An Introduction to Tabu Search*. Montréal, Canada: [s.n.], 2002. Disponível em: <[http://www.ifi.uio.no/infheur/Bakgrunn/Intro\\_to\\_TS\\_Gendreau.htm](http://www.ifi.uio.no/infheur/Bakgrunn/Intro_to_TS_Gendreau.htm)>.
- GLOVER, F. *Tabu Search Fundamentals and Uses*. 1995. University of Colorado at Boulder. Disponível em: <<http://leeds-faculty.colorado.edu/glover/TS%20-%20Fundamentals&Uses.pdf>>.
- GLOVER, F.; LAGUNA, M. Modern heuristic techniques for combinatorial problems. In: REEVES, C. R. London, UK: John Wiley & Sons, 1993. cap. Tabu Search, p. 70–150.

- GOMES, L. C. T. *Inteligência Computacional na Síntese de Meta-Heurísticas para Otimização Combinatória e Multimodal*. Tese (Doutorado) — Universidade Estadual de Campinas - Faculdade de Engenharia Elétrica e de Computação, 2006.
- GONZÁLEZ, F. A.; DASGUPTA, D. Anomaly detection using real-valued negative selection. *Genetic Programming and Evolvable Machines*, v. 4, n. 4, 2003.
- GREENSMITH, J.; AICKELIN, U.; TWYXCROSS, J. Detecting danger: Applying a novel immunological concept to intrusion detection systems. In: ACDM. *6th International Conference in Adaptive Computing in Design and Manufacture*. Clifton, Bristol, UK, 2004.
- HART, W.; ISTRAIL, S. *HP Benchmarks*. visitado em 08/2005. Disponível em: [http://www.cs.sandia.gov/tech\\_reports/compbio/tortilla-hp-benchmarks.html](http://www.cs.sandia.gov/tech_reports/compbio/tortilla-hp-benchmarks.html).
- HERTZ, A.; TAILLARD, E.; WERRA, D. de. A tutorial on tabu search. In: AIRO. *Enterprise Systems: Management of Technological and Organizational Changes*. Lisboa, Portugal, 1995. p. 13–24.
- HSU, H. P. et al. Growth algorithm for lattice heteropolymers at low temperatures. *Journal of Chemical Physics*, v. 118, p. 444–451, 2003.
- ISHAK, S.; ABIDIN, A.; RAHMAN, T. Static var compensator planning using artificial immune system for loss minimisation and voltage improvement. In: PECON. *National Power and Energy Conference*. [S.l.], 2004. p. 41–45.
- ISHIGURO, A. et al. Moderationism in the immune system: Gait acquisition of a legged robot using the metadynamics function. In: SMC. *IEEE System, Man, and Cybernetics Conference*. San Diego, USA, 1998. p. 3827–3832.
- JANG, J. S.; SUN, C. T.; MIZUTANI, E. *Neuro-fuzzy and soft computing: A Computational Approach to Learning and Machine Intelligence*. USA: Prentice Hall, 1997.
- JERNE, N. K. Towards a network theory of the immune system. *Immunology*, v. 125C, p. 373–389, 1974.
- JUNQUEIRA, C. et al. Immune-inspired dynamic optimization for blind spatial equalization in undermodeled channels. In: CEC. *Congress on Evolutionary Computation*. Vancouver, BC, Canada, 2006. p. 9801–9808.
- KIM, D. H. Tuning of a pid controller using an artificial immune network model and local fuzzy set. In: IFA/NAFIPS. *International Conference IFSA World Congress and 20th NAFIPS*. Vancouver, Canada, 2001. v. 5, p. 2698–2703.
- KOLINSKI, A.; SKOLNICKB, J. Reduced models of proteins and their applications. *Polymer*, n. 45, p. 511–524, 2004.
- LAU, K. F.; DILL, K. A. Lattice statistical mechanics model of the conformation and sequence space of proteins. *Macromolecules*, v. 22, p. 3986–3997, 1989.

- LAW, H. Y. K.; WONG, V. W. K. Immunologic control framework for automated material handling. In: TIMMIS, J.; BENTLEY, P.; HART, E. (Ed.). *Second International Conference on Artificial Immune Systems*. Edinburgh, UK, 2004. (Lecture Notes in Computer Science, v. 2787), p. 57–68.
- LEE, C. C. Fuzzy logic in control systems - part i. *IEEE Transactions on Systems, Man and Cybernetics*, v. 20, n. 2, p. 404–418, 1990.
- LEE, C. C. Fuzzy logic in control systems - part ii. *IEEE Transactions on Systems, Man and Cybernetics*, v. 20, n. 2, p. 419–435, 1990.
- LEE, D. et al. Immune-based framework for exploratory bio-information retrieval from the semantic web. In: TIMMIS, J.; BENTLEY, P.; HART, E. (Ed.). *Second International Conference on Artificial Immune Systems*. Edinburgh, UK, 2003. (Lecture Notes in Computer Science 2787), p. 128–135.
- LEE, S. et al. A new immunotronic approach to hardware fault detection using symbiotic evolution. In: IWINAC. *First International Work-Conference on the Interplay Between Natural and Artificial Computation*. Canary Islands, Spain, 2005. (Lecture Notes in Computer Science, v. 3562), p. 133–142.
- LOPES, H. S.; SCAPIN, M. P. An enhanced genetic algorithm for protein structure prediction using the 2d hydrophobic-polar model. In: EA. *7th International Conference on Artificial Evolution*. University of Exeter, England, 2005. p. 238–246.
- MACKERELL, A. D. J. et al. The encyclopedia of computational chemistry. In: \_\_\_\_\_. [S.l.]: John Wiley Sons, 1998. v. 1, cap. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program, p. 271–277.
- MELO, A. R. et al. *Principais Constituintes dos Microorganismos*. visitado em 04/2006. [Http://www.enq.ufsc.br/labs/probio/disc\\_eng\\_bioq/trabalhos\\_pos2003/const\\_microorg/engenharia\\_bioquimica.htm](http://www.enq.ufsc.br/labs/probio/disc_eng_bioq/trabalhos_pos2003/const_microorg/engenharia_bioquimica.htm).
- PARISI, G. A simple model for the immune network. *Immunology*, v. 87, p. 429–433, 1990.
- PEDRYCZ, W.; GOMIDE, F. *An Intruction to Fuzzy Sets: Analysis and Design*. London, UK: Cambridge: MIT Press, 1998.
- PERELSON, A. S.; OSTER, G. F. Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-nonsel self discrimination. *Journal of Theoretical Biology*, v. 81, p. 645–670, 1979.
- RAHMAN, T. A.; YASIN, Z.; ABDULLAH, W. Artificial-immune-based for solving economic dispatch in power system. In: PECON. *National Power and Energy Conference*. [S.l.], 2004. p. 31–35.
- REGO, C.; LI, H.; GLOVER, F. *A Filter-and-Fan Approach to the 2D HP Model of the Protein Folding Problem*. Disponível em: <http://faculty.bus.olemiss.edu/crego/Papers.html>.

- RICHARDS, F. M. Areas, volumes, packing and protein structures. *Annual Review of Biophysics and Bioengineering*, v. 6, p. 151–176, 1977.
- SHMYGELSKA, A.; HOOS, H. H. An improved ant colony optimisation algorithm for the 2d hp protein folding problem. v. 2671, n. 2671, p. 400–417, 2003.
- SHMYGELSKA, A.; HOOS, H. H. An ant colony optimisation algorithm for the 2d and 3d hidrofobic polar protein folding problem. *BMC Bioinformatics*, v. 6, p. 1–22, 2005.
- TIMMIS, J. *Challenges for Artificial Immune Systems*. Março 2005. Disponível em: <[www.cs.york.ac.uk/nature/workshop/papers/Timmis.pdf](http://www.cs.york.ac.uk/nature/workshop/papers/Timmis.pdf)>.
- TIMMIS, J.; EDMONDS, C. A comment on opt-ainet: An immune network algorithm for optimisation. In: KALYANMOY, D. (Ed.). *Genetic and Evolutionary Computation*. Seattle, WA, USA: Springer, 2004. (Lecture Notes in Computer Science, v. 3102), p. 308–317.
- UNGER, R.; MOULT, J. Genetic algorithms for protein folding simulations. *Journal of Molecular Biology*, v. 231, n. 1, p. 75–81, 1993.
- VARGAS, P. A. et al. Implementation of an immuno-genetic network on a real khepera ii robot. In: CEC. *Congress on Evolutionary Computation*. Newport Beach, CA, USA, 2003. v. 1, p. 420–426.
- WALKER, J. H.; GARRETT, S. M. Dynamic function optimisation: Comparing the performance of clonal selection and evolution strategies. In: TIMMIS, J.; BENTLEY, P.; HART, E. (Ed.). *Proc. Of the Second International Conference on Artificial Immune Systems*. Edinburgh, UK, 2003. (Lecture Notes in Computer Science, v. 2787), p. 273–284.
- WANG, L.; HIRSBRUNNER, B. Immune mechanism based computer security design. In: ICMLC. *International Conference on Machine Learning and Cybernetics*. Beijing, China, 2002. v. 4, p. 1887–1893.
- WATKINS, A. *Exploiting Immunological Metaphors in the Development of Serial, Parallel and Distributed Learning Algorithms*. Tese (Doutorado) — Computing Laboratory, University of Kent, Canterbury, UK, March 2005. Disponível em: <<http://www.cs.kent.ac.uk/pubs/2005/2178>>.
- WATKINS, A.; TIMMIS, T.; BOGGESE, L. Artificial immune recognition system (airs): An immune inspired supervised machine learning algorithm. *Genetic Programming and Evolvable Machines*, v. 5, n. 1, p. 291–317, 2004.
- ZADEH, L. A. Fuzzy sets. *Information and Control*, v. 8, p. 338–352, 1965.
- ZHOU, J.; DASGUPTA D.; ZHILING, Y.; HONGMEI, T. Analysis of dental images using artificial immune systems. In: CEC. *IEEE Congress on Evolutionary Computation*. [S.l.], 2006. p. 528–535.

## *APÊNDICE A -- Aminoácidos*

Os aminoácidos se diferem em dois grupos: hidrofóbicos e hidrofílicos. Os hidrofóbicos são não-polares e os hidrofílicos são polares. Os hidrofílicos podem ser carregados positivamente, negativamente ou eletricamente neutros. O nome dos aminoácidos são apresentados na tabela A.1 com seus respectivos símbolos e tipos de cadeias laterais.

Tabela A.1: Tabela de aminoácidos.

Hidrofílicos ou Polares				Hidrofóbicos ou não-Polares			
Aminoácido		Cadeia lateral		Aminoácido		Cadeia lateral	
Ácido Aspártico	Asp	D	negativa	Alanina	Ala	A	não-polar
Ácido Glutâmico	Glu	E	negativa	Glicina	Gly	G	não-polar
Arginina	Arg	R	positiva	Valina	Val	V	não-polar
Lisina	Lys	K	positiva	Leucina	Leu	L	não-polar
Histidina	Hys	H	positiva	Isoleucina	Ile	I	não-polar
Asparagina	Asn	N	neutra	Prolina	Pro	P	não-polar
Glutamina	Gln	Q	neutra	Fenilalanina	Phe	F	não-polar
Serina	Ser	S	neutra	Metionina	Met	M	não-polar
Treonina	Thr	T	neutra	Triptofano	Trp	W	não-polar
Tirosina	Tyr	Y	neutra	Cisteína	Cys	C	não-polar

## *APÊNDICE B -- Estruturas das Proteínas*

As estruturas foram geradas pelo software para desenvolvimento de moléculas Tinker na versão 4.2, disponível em <http://dasher.wustl.edu/tinker/>. Os resíduos vermelhos correspondem a resíduos hidrofóbicos e os resíduos azuis correspondem a resíduos polares. Estes dobramentos foram realizados pelo melhor modelo proposto neste trabalho, o  $\text{ClonalgIII}_{inf-BT}$ , e correspondem aos anticorpos que alcançaram as melhores afinidades durante as 50 execuções.

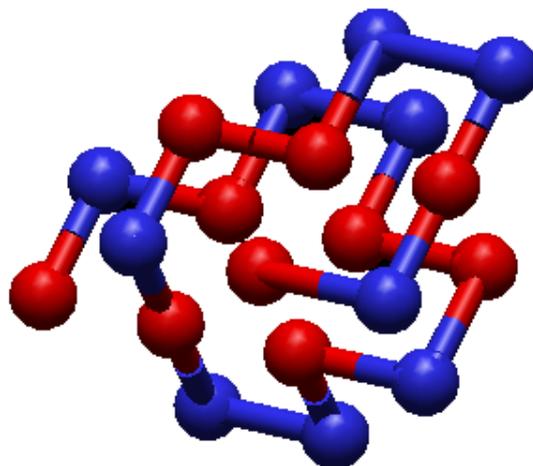


Figura B.1: Estrutura da proteína de tamanho 20.

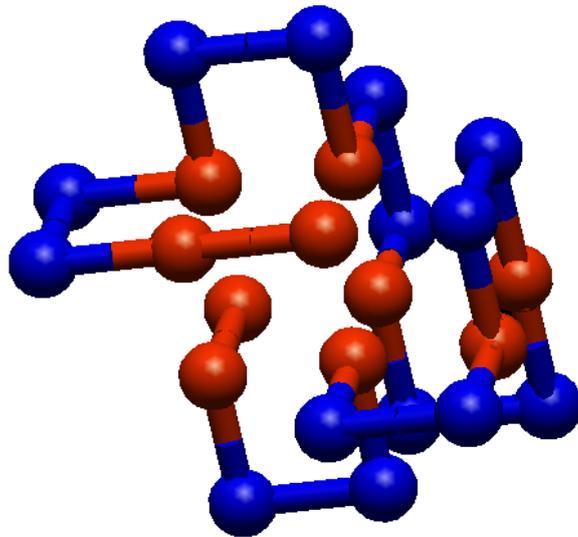


Figura B.2: Estrutura da proteína de tamanho 24.

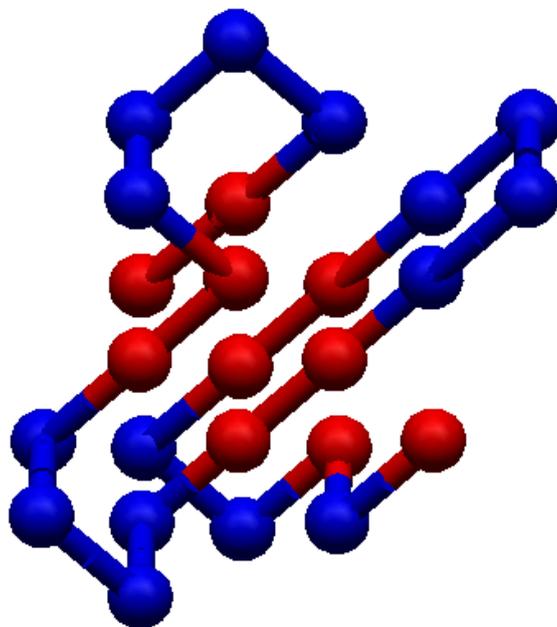


Figura B.3: Estrutura da proteína de tamanho 25.

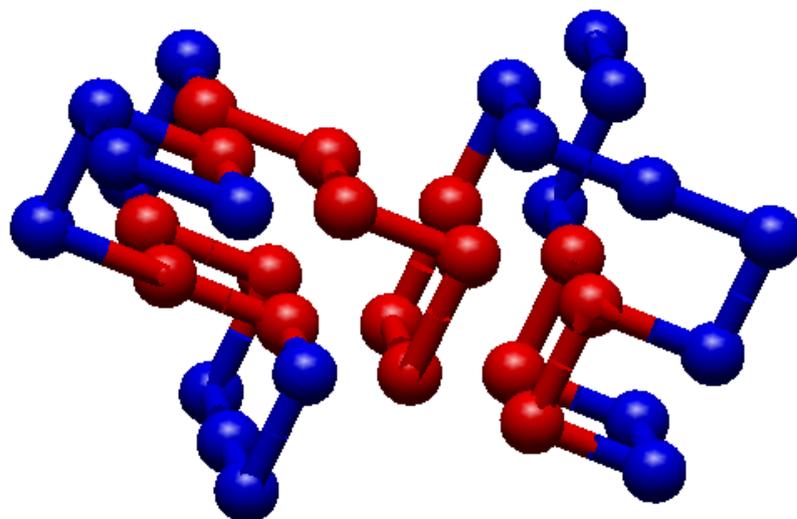


Figura B.4: Estrutura da proteína de tamanho 36.

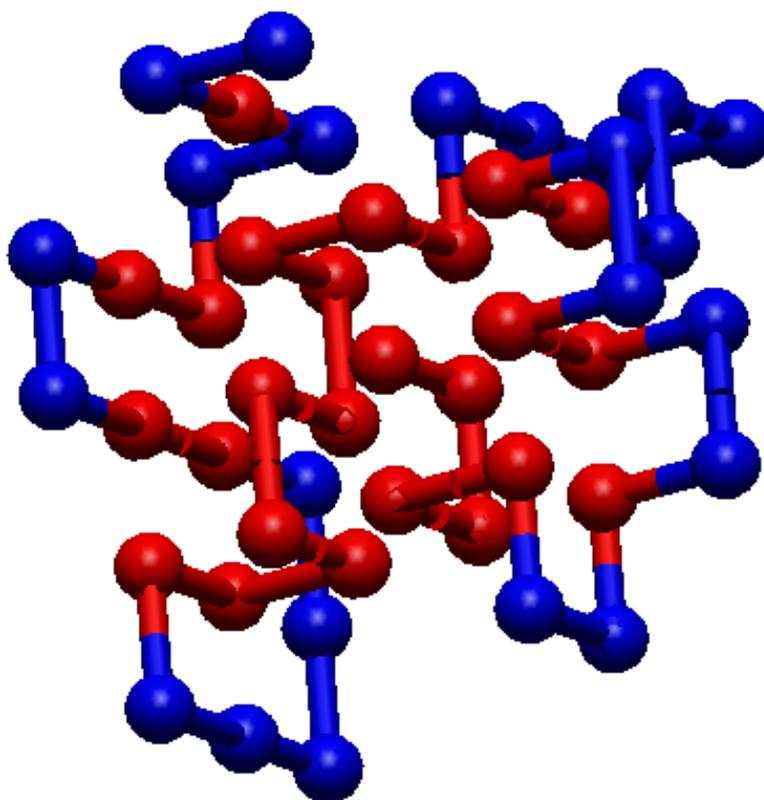


Figura B.5: Estrutura da proteína de tamanho 48, esta estrutura possui uma energia mínima (-30) melhor do que aquelas encontradas na literatura (-29).

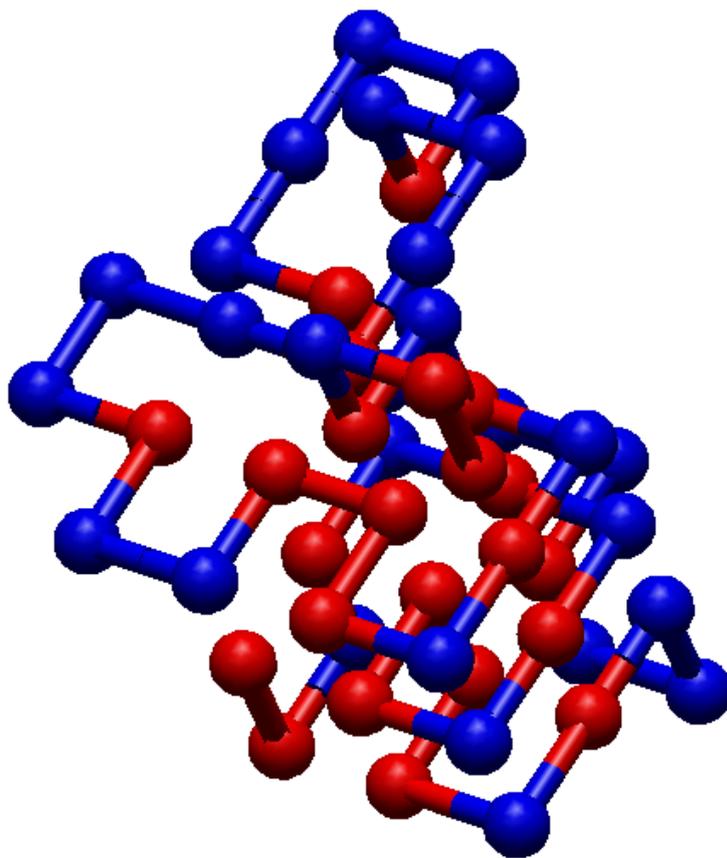


Figura B.6: Estrutura da proteína de tamanho 50, esta estrutura possui uma energia mínima (-30) melhor do que aquelas encontradas na literatura (-26).

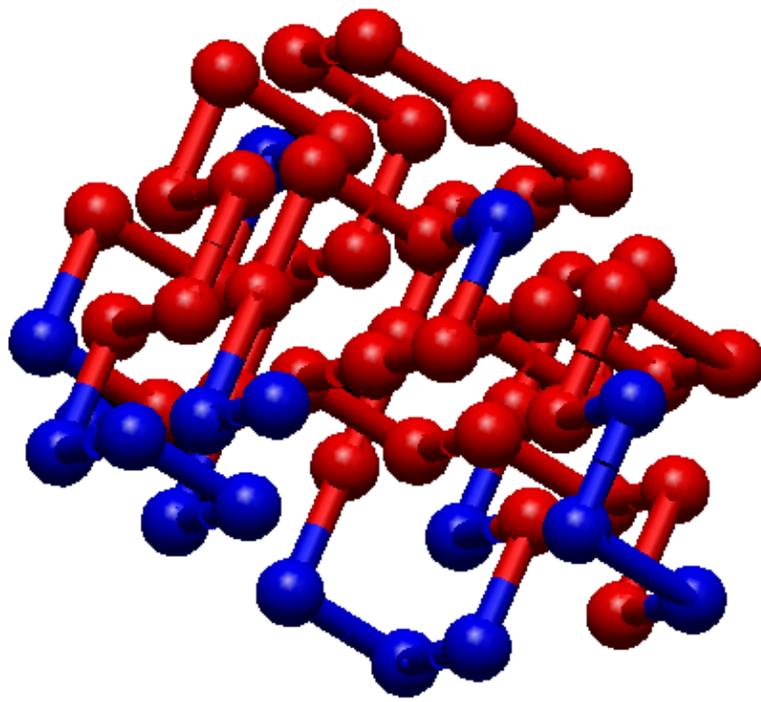


Figura B.7: Estrutura da proteína de tamanho 60, esta estrutura possui uma energia mínima (-51) melhor do que aquelas encontradas na literatura (-49).

## RESUMO:

Este trabalho apresenta modelos computacionais baseados em Sistemas Imunológicos Artificiais para resolver o problema da Predição da Estrutura de Proteínas no modelo Tri-dimensional Hidrofóbico-Polar. A Predição da Estrutura de Proteínas consiste em encontrar o arranjo espacial dos aminoácidos de uma proteína que possua energia mínima. A metodologia proposta tem foco nos Sistemas Imunológicos Artificiais mas utiliza como técnicas auxiliares os Sistemas de Inferência *Fuzzy* e a Busca Tabu. São aplicados dois tipos de algoritmos imunológicos: o algoritmo Clonalg e a Rede Imune. Em ambos os casos são comparados métodos baseados em penalidade com métodos que trabalham somente com soluções factíveis. Em alguns modelos implementados, um operador de *Aging* (que pode ser *Fuzzy* ou Puro) é usado para decidir quais anticorpos serão eliminados da população antes da etapa de seleção. Além disso, são implementados dois estágios de maturação da afinidade: fraco e intensivo, sendo o primeiro baseados nos operadores de hipermutação e hipermacromutação e o segundo implementado através da Busca Tabu. Para validação dos resultados os algoritmos foram aplicados a algumas instâncias do *benchmark* Tortilla. Os resultados mostram que os modelos baseados no Clonalg superam os baseados na Rede Imune. Os melhores resultados foram obtidos com o modelo de Clonalg que uniu as três técnicas supracitadas e são comparáveis àqueles encontrados na literatura. Em alguns casos inclusive, novos valores de mínimo de energia foram encontrados pela metodologia proposta.

## PALAVRAS-CHAVE

Sistemas Imunológicos Artificiais, Sistema de Inferência *Fuzzy*, Busca Tabu, Predição da Estrutura de Proteínas.

## ÁREA/SUB-ÁREA DE CONHECIMENTO

- 1.03.03.00 – 7 Ciência da Computação
  - 1.03.03.04 – 9 Sistemas de Informação
    - 1.03.04.03 – 7 Software Básico

Ano 2007

Nº: 438

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)