

# Predição da Estrutura de Proteínas Utilizando Algoritmo Evolutivo Adaptativo

Jhennifer N. Lopes e Sandra M. Venske

Universidade Estadual do

Centro-Oeste - UNICENTRO

Departamento de Ciência da Computação

Guarapuava, Paraná 85015-430

Email: jhenniferahirnei@gmail.com, ssvenske@unicentro.br

**Resumo**—Proteínas são cadeias de aminoácidos que regulam diversas atividades em todos os organismos conhecidos. O Problema da Predição da Estrutura de Proteínas (PPEP) é um tópico importante de pesquisa em Bioinformática, cuja solução ainda está em aberto. Conhecer a estrutura de uma proteína pode ajudar a descobrir que tipo de atividade a proteína realiza na célula e criar drogas para combater doenças. Neste trabalho foi utilizado o método evolucionário da Evolução Diferencial, utilizando a técnica adaptativa *Probability Matching*, aplicado ao PPEP. Testes foram executados a fim de confrontar o método adaptativo proposto com versões não adaptativas, além de realizado o comparativo com a literatura considerando a proteína *Met-Enkephalin*.

**Palavras-chave:** Evolução Diferencial, Técnica Adaptativa, Bioinformática.

## I. INTRODUÇÃO

As proteínas são cadeias de aminoácidos (resíduos) que regulam uma variedade de atividades em todos os organismos conhecidos, da replicação do código genético ao transporte de oxigênio. As funções das proteínas são determinadas por suas estruturas, que podem ser organizadas em quatro níveis hierárquicos, com um incremento de complexidade: estrutura primária, secundária, terciária e quaternária [1]. Assim, a estrutura de uma proteína é um pré-requisito para obter uma compreensão completa da sua função em nível molecular [2]. Quando uma proteína está na sua forma terciária, sua energia potencial livre é mínima [3].

O Problema da Predição da Estrutura de Proteínas (PPEP) é um problema de minimização da energia potencial livre de uma proteína e um tópico importante de pesquisa, ainda sem solução, em Bioinformática. Conhecendo a estrutura de uma proteína, os cientistas podem descobrir que tipo de atividade a proteína realiza na célula e criar drogas para combater doenças relacionadas ao mal funcionamento de proteínas.

Há várias abordagens para modelagem computacional do PPEP [3]. Neste trabalho a modelagem utilizada é a *ab initio*, que prediz a conformação nativa de uma proteína considerando somente sua sequência de aminoácidos. A metodologia *ab initio* requer três elementos [4]: uma representação da geometria da proteína, uma função potencial e uma técnica de busca do espaço da superfície da energia livre.

A representação da geometria se dá pelo cálculo da energia potencial, a fim de procurar o espaço para o seu enovelamento. Neste trabalho foi utilizada a representação livre de látice,

em que uma proteína é representada como uma cadeia de resíduos (ou grupos de resíduos) que se deslocam em um espaço contínuo, utilizando suas coordenadas atômicas.

A função potencial retorna um valor para a energia de acordo com a sua conformação. Neste trabalho foi investigado o uso da função CHARMM (v.27) [5] para calcular a função de energia.

Nas técnicas de buscas conformacionais são calculadas as energias envolvidas no processo do enovelamento para encontrar a estrutura com a energia livre mais baixa. Neste trabalho foi utilizado o método evolucionário da Evolução Diferencial (ED) como técnica de busca.

A ED é um algoritmo evolucionário de otimização numérica [6] eficiente no contexto da otimização não-linear com variáveis contínuas. A ED tem como base alguns parâmetros como tamanho da população, constante de cruzamento e constante de mutação, além de estratégias de mutação. A adaptação de parâmetros tem como objetivo superar as desvantagens da ED clássica, em que a definição de valores adequados pode levar a elevados custos computacionais, devido a processos de tentativa-e-erro demorados e ainda pode ser ineficaz em casos em que os valores dos parâmetros ideais variam durante o processo evolutivo. Neste trabalho, a escolha da estratégia de mutação será realizada utilizando a técnica adaptativa *Probability Matching* [7].

A principal contribuição deste trabalho é a avaliação do uso ED adaptativa com *Probability Matching* para o PPEP com otimização mono-objetivo, considerando a abordagem *ab initio* e a modelagem computacional livre de látice.

O restante deste trabalho está organizado como segue. A seção II descreve alguns conceitos e técnicas utilizadas. Na seção III é apresentado o algoritmo proposto. Na seção IV são mostrados os resultados obtidos e a comparação com a literatura. A seção V apresenta as conclusões e os trabalhos futuros.

## II. REVISÃO BIBLIOGRÁFICA

### A. Proteínas

As proteínas são moléculas formadas por aminoácidos e são responsáveis pelo funcionamento da maioria dos processos biológicos das células. Existem 20 tipos de aminoácidos, e sua combinação pode formar milhares de proteínas diferentes,

sendo que cada uma possui uma única sequência de aminoácidos. Entre as principais funções biológicas das proteínas estão a replicação de DNA, conversão de uma molécula em outra, transporte de oxigênio, catálise e formação da estrutura do citoesqueleto das células [8].

A natureza modular da síntese e dobramento de proteínas é incorporada no conceito de ordens de estruturas da proteínas, sendo elas organizadas em 4 níveis estruturais hierárquicos: estrutura primária, estrutura secundária, estrutura terciária e estrutura quaternária [9].

A estrutura primária das proteínas é formada por sua sequência linear de resíduos de aminoácidos. As estruturas secundárias são dobramentos de pequenos segmentos ligados de proteínas em unidades geometricamente ordenadas. Esses dobramentos acontecem utilizando arranjos conhecidos como folhas -  $\beta$  e hélices- $\alpha$ . As estruturas terciárias são a montagem de várias estruturas secundárias no espaço 3D. A forma 3D que a proteína assume também é chamada de estrutura nativa da proteína ou estrutura funcional. Quando a proteína assume a estrutura terciária ela desenvolve a sua funcionalidade sendo possível é prever ou analisar a função que ela exerce no organismo. A estrutura Quaternária é a junção de várias estruturas terciárias, ou seja, quando uma proteína possui mais de uma cadeia polipeptídica.

### B. Problema da Predição da Estruturas das Proteínas (PPEP)

A predição da estrutura de proteínas é um dos grandes problemas atuais da Bioinformática. Das milhares de proteínas existentes, apenas se conhece uma pequena parcela das estruturas que as compõem. Esse conhecimento das estruturas e da função das proteínas é de suma importância, pois pode ser utilizado para a descoberta de novas medicações ou até na cura contra doenças como Alzheimer, alguns tipos de câncer, Parkinson, entre outras, já que algumas doenças estão ligadas ao mal funcionamento das proteínas.

Para auxiliar no processo de predição da estrutura tridimensional de proteínas é possível usar métodos computacionais. As abordagens computacionais que podem ser utilizadas para o PPEP podem ser divididas em dois grupos [3]: as baseadas em padrões, como homologia e *threading* e os métodos que fazem a predição sem a necessidade de padrões para comparação, como *ab initio* e *de novo*. Neste trabalho foi utilizada a abordagem *ab initio* com o modelo livre de látice, que representa uma proteína como uma cadeia de resíduos (ou grupo de resíduos) que se dobram em um espaço contínuo. A abordagem *ab initio* simula o espaço conformacional da proteína utilizando uma função de energia que descreve a sua energia interna, sendo que o objetivo é encontrar um mínimo global de energia livre que corresponda ao estado nativo ou funcional da proteína [3].

O cálculo de energia livre (E) de uma dada conformação

(C) é dado pela Equação 1 [3].

$$E_C = \sum_{ligacoes} K_b(b_C - b_{eq})^2 + \sum_{angulos} K_\theta(\theta_C - \theta_{eq})^2 + \sum_{diedrais} K_\Phi[1 + \cos(n\Phi_C - \gamma)] + \sum_{atomos\ nao-ligados} \left[ \frac{A_{ij}}{r_{ij,C}^{12}} - \frac{B_{ij}}{r_{ij,C}^6} + \frac{q_i q_j}{\epsilon_0 r_{ij,c}} \right] \quad (1)$$

tal que:

- $K_b$  é a constante de ligação;
- $K_\theta$  é a constante para ângulos;
- $K_\Phi$  é a constante para ângulos de torção (diedrais);
- $b$  é a distância de ligação entre dois átomos;
- $b_{eq}$  é a distância ideal entre dois átomos;
- $\theta$  é o ângulo da ligação química;
- $\theta_{eq}$  é o ângulo ideal da ligação química;
- $n$  é o número de mínimos;
- $\gamma$  é o deslocamento angular;
- $\epsilon$  é a energia de atração/repulsão de Lennard-Jonnes;
- $A$  e  $B$  são constantes para o potencial de Lennard-Jonnes;
- $q_i$  e  $q_j$  são as cargas de átomos não-ligados;
- $r$  é a distância entre átomos não-ligados;
- $\epsilon_0$  é a constante dielétrica.

Esta função fornece o valor aproximado da energia potencial de uma conformação de proteína. Os componentes de ligação, ângulos e diedrais representam a energia dos átomos ligados, enquanto a última parte da fórmula acima representa a energia dos átomos não ligados. Numa abordagem multiobjetivo, a otimização destas energias é realizada separadamente, diferente da formulação mono-objetivo que considera a soma das energias.

A função de energia total (ou campo de força) retorna um valor para a energia de acordo com as conformações da molécula. A abordagem *ab initio* visa minimizar essas funções do PPEP. Existe muitos campos de força para simulação de proteínas, como por exemplo AMBER, CHARMM e GRO-MOS [10], [11]. Neste trabalho, foi investigado o uso do campo de força CHARMM (v.27), para calcular a função de energia interna (itens ligados) e externos (itens não ligados). Há muitos estudos que também utilizam o CHARMM como campo de força [12]–[15].

Esse trabalho adota um modelo livre de látice baseado em representação de coordenadas internas como modelo das proteínas. Cada tipo de resíduo tem um número pré-estabelecido de ângulos de torção para alcançar a conformação da proteína. A espinha dorsal de cada resíduo é representada por 3 ângulos:  $\phi$ ,  $\psi$ ,  $\omega$ . As cadeias laterais são representadas pelos ângulos

$\chi$ . O número total de ângulos  $\chi$  depende do tipo de resíduo (conforme apresentado na Tabela I). Os valores dos ângulos podem variar entre  $-180^\circ$  a  $180^\circ$ .

Tabela I. NÚMERO DE ÂNGULOS  $\chi$  EM CADA RESÍDUO.

Resíduo	Número de ângulos $\chi$
GLY, ALA, PRO	cadeia principal
SER, CYS, THR, VAL	$\chi_1$
ILE, LEU, ASP, ASN, PHE, TYR, TRP	$\chi_1, \chi_2$
MET, GLU, GLN	$\chi_1, \chi_2, \chi_3$
LYS, ARG	$\chi_1, \chi_2, \chi_3, \chi_4$

### C. Evolução Diferencial (ED)

Algoritmos Evolucionários são algoritmos de otimização, baseados em população, que têm como objetivo a resolução de problemas por meios iterativos sendo muitos deles baseados na teoria evolucionária de Charles Darwin. A cada iteração novos indivíduos (chamados de vetores da população) são gerados e avaliados (função de avaliação - *fitness*) a partir de indivíduos usados nas iterações anteriores, assim encontrando soluções aproximadas para problemas específicos [16].

A Evolução Diferencial é um algoritmo evolutivo para otimização numérica tal que é adicionada a diferença ponderada entre vetores da população (vetores de diferença) a um terceiro vetor (vetor alvo) [6].

A ED possui basicamente três parâmetros de controle: tamanho da população (NP), constante de mutação (F) e constante de cruzamento (CR). O desempenho da ED é sensível a esses parâmetros pois diferentes valores apresentarão desempenhos diferentes [17].

Existem diferentes estratégias para a aplicação da ED, representadas pela notação “ED/x/y/z”. Estas estratégias diferem na forma em que o vetor alvo é encontrado (x), no número de vetores de diferença utilizados (y) e, por fim, no tipo de cruzamento a ser utilizado (z) [18]. Neste trabalho foram utilizadas as estratégias apresentadas na Tabela II.

Tabela II. ESTRATÉGIAS DE MUTAÇÃO DA ED UTILIZADAS NESTE TRABALHO.

Estratégia	Notação
ED/rand/1	$\omega = x_1 + F * (x_2 - x_3)$
ED/current-to-rand/1	$\omega = x_t + F * (x_1 - x_t) + F * (x_2 - x_3)$
ED/rand/2	$\omega = x_1 + F * (x_2 + x_3 - x_4 - x_5)$
ED/Non-Linear	$\omega = w^2 \mathbf{c}_a + w \mathbf{c}_b + \mathbf{c}_c$

As estratégias ED/rand/1 e ED/rand/2 selecionam aleatoriamente o vetor alvo ( $x_1$ ) e utilizam um ou dois pares de vetores diferença, respectivamente. Na estratégia ED/current-to-rand/1 o indivíduo  $x_t$  é o indivíduo atual da população e é utilizado com um par de vetores aleatórios. A estratégia ED/Non-Linear foi proposta em [19] e é definida pela Equação 2.

$$\mathbf{y} = w^2 \mathbf{c}_a + w \mathbf{c}_b + \mathbf{c}_c, \quad (2)$$

sendo que  $w$  é dado pela Equação 3.

$$w \in \begin{cases} U[0, 2], & \text{se } rand \leq P_{inter} \\ U[2, 3], & \text{caso contrário.} \end{cases} \quad (3)$$

com  $rand \in U[0, 1]$  (i.e.,  $rand$  é um valor entre 0 e 1 aleatoriamente gerado por uma distribuição uniforme) e a probabilidade de interpolação,  $P_{inter}$  definida como em [19].

Os indivíduos  $\mathbf{c}_a$ ,  $\mathbf{c}_b$  e  $\mathbf{c}_c$  são definidos de acordo com a Equação 4.

$$\begin{aligned} \mathbf{c}_a &= \frac{\mathbf{x}_c - 2\mathbf{x}_b + \mathbf{x}_a}{2}, \\ \mathbf{c}_b &= \frac{4\mathbf{x}_b - 3\mathbf{x}_c - \mathbf{x}_a}{2}, \\ \mathbf{c}_c &= \mathbf{x}_c \end{aligned} \quad (4)$$

tal que  $\mathbf{x}_a$ ,  $\mathbf{x}_b$  e  $\mathbf{x}_c$  são indivíduos aleatoriamente selecionados da população atual.

### D. Técnica adaptativa: Probability Matching (PM)

Probability Matching é uma técnica em que se calcula a probabilidade de uso de cada estratégia de um determinado conjunto de estratégias. A PM faz esse cálculo considerando quatro valores base: médias das melhorias dos *fitness* dos indivíduos, média de qualidade, taxa de adaptação e um valor de probabilidade [7].

A PM avalia o desempenho das estratégias de um conjunto com uma medida de *recompensa*, onde os valores bases são testados com uma variável que repete o grau de melhoria da medida do *fitness*. Se houver alguma melhoria é atribuído um valor numérico como recompensa.

Seja um conjunto  $\zeta = \{s_1, \dots, s_S\}$  e  $S > 1$  em um vetor de probabilidades  $\mathbf{p}(g) = (p_1(g), \dots, p_S(g))$ , tal que  $p_s(g)$  é a probabilidade de que a estratégia  $s^{th}$  seja utilizada na geração  $g$ , existem vários métodos para utilizar adaptativamente a probabilidade  $p_s(g)$  da estratégia  $s^{th}$  [20]. Em todos os casos tem-se que  $\sum_{s=1}^S p_s(g) = 1; \forall g : p_{min} \leq p_s(g) \leq 1$ , sendo  $p_{min}$  o valor de probabilidade mínimo que cada estratégia possui, este valor é utilizado para que nenhuma estratégia se perca [20].

O método PM segundo [20] atualiza-se  $p_s(g)$  seguindo a Equação 5 :

$$p_s(g+1) = p_{min} + (1 - S * p_{min}) * \frac{q_s(g+1)}{\sum_{s=1}^S q_s(g+1)} \quad (5)$$

na qual

$$q_s(g+1) = q_s(g) + \alpha * [r_s(g) - q_s(g)] \quad (6)$$

tal que  $\alpha \in (0, 1]$  é a taxa de adaptação,  $q_s(g)$  é a qualidade empírica e  $r_s(g)$  é a recompensa que a estratégia  $s^{th}$  recebe.

Na Equação 5 percebe-se que  $\sum_{s=1}^S p_s(g+1) = 1$ . Quando apenas uma estratégia obtém recompensa por um longo tempo, a sua probabilidade de seleção converge  $p_{max} = p_{min} + (1 - S * p_{min})$ .

O mecanismo de recompensa funciona com o melhoramento do indivíduo da população durante as avaliações, como mostra a Equação 7.

$$\eta_s^u(g) = \begin{cases} f_{best}/f_c(g) * |f_p(g) - f_c(g)|, & \text{Se } f_c(g) \text{ melhor que } f_p(g) \\ 0 & \text{caso contrário,} \end{cases} \quad (7)$$

sendo que  $f_{best}$  é o *fitness* da melhor solução da população,  $f_p(g)$  e  $f_c(g)$  são os valores de *fitness* do vetor pai e seu filho na geração  $g$ , ambos associados com um melhoramento. Caso não ocorra nenhum melhoramento, ou seja, o filho é igual ou pior que o pai, o impacto da estratégia é considerado nulo ( $\eta_s^u(g) = 0$ ).

No final de cada geração  $g$ , um valor de recompensa  $R_s(g)$  é usado para atualizar a medida de qualidade  $q_s(g+1)$  da estratégia. Esse valor obtém-se pela Equação 8.

$$r_s(g) = \eta_s^*(g) = \max [\eta_s^u(g)]_{u=1, \dots, |R_s|}, \quad (8)$$

### III. ALGORITMO PROPOSTO

O Algoritmo 1 representa o pseudocódigo da abordagem proposta para o PPEP.

**Algorithm 1** Pseudocódigo do algoritmo proposto.

- 1: Gerar a população inicial  $P^0 = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ ,  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})$
- 2: Avaliar cada indivíduo da população inicial  $P^0$
- 3:  $g = 1$ . Para todas as estratégias  $s = 1, \dots, S$ , atribuir  $q_s(g) = 0$  e  $p_s(g) = 1/S$
- 4: **repita**
- 5:     **para** cada vetor pai  $\mathbf{x}_i$ ,  $i = 1, \dots, N$  **faça**
- 6:         Selecionar a estratégia  $s$  do conjunto  $\zeta$  de acordo com  $p_s(g)$
- 7:         Gerar nova solução  $\mathbf{y}$  pela estratégia  $s$  (reparar se necessário)
- 8:         Aplicar mutação polinomial e produzir  $\mathbf{y}'$  (reparar se necessário)
- 9:         **se**  $y' < x_i$  **então**
- 10:             Substituir  $\mathbf{x}_i$  por  $\mathbf{y}'$
- 11:         Calcular o melhoramento  $\eta_s^u(g)$  da estratégia e agrupá-la em um conjunto  $R_s$
- 12:         **senão**
- 13:              $\eta_s^u(g) = 0$
- 14:         **fim se**
- 15:         **fim para**
- 16:         Calcular (com base no ExtAbs) a recompensa  $r_s(g)$  para cada estratégia
- 17:         Atualizar a qualidade  $q_s(g)$  para cada estratégia com base em  $r_s(g)$
- 18:         Atualizar a probabilidade  $p_s(g)$  para cada estratégia
- 19:          $g = g + 1$ ;
- 20: **até** que  $MAX - AV$  seja atingido

Nos primeiros passos do Algoritmo 1, nas linhas 1 a 3 ocorrem a inicialização da população, avaliação e também são inicializados os parâmetros do método adaptativo ( $q_s(g) = 0$  e  $p_s(g)$ ). Em seguida, o algoritmo entra no laço principal (linha 4 a 20), que seleciona a estratégia a ser usada na geração  $g$ . O conjunto das estratégias candidatas que podem ser escolhidas está incluído na Tabela II.

No passo 6 é selecionada a estratégia  $s^{th}$  a ser usada para gerar um novo indivíduo  $\mathbf{y}$ . A seleção é realizada na geração atual  $g$ , baseada na medida de probabilidade  $p_s(g)$ . Inicialmente, todas as estratégias candidatas do conjunto têm a mesma probabilidade de serem selecionadas.

Após selecionar a estratégia (passo 6), no passo 7 será gerada uma nova solução  $\mathbf{y}$ , que pode ser reparada, se necessário. No passo 8 é aplicada a mutação polinomial em  $\mathbf{y}$  e gerado

o indivíduo modificado  $\mathbf{y}' = (y'_1, \dots, y'_n)$  (que também pode sofrer reparação).

Se  $\mathbf{y}'$  (indivíduo filho gerado) é melhor que  $\mathbf{x}_i$  (indivíduo pai),  $\mathbf{y}'$  irá substituir  $\mathbf{x}_i$  na população. A avaliação do quanto o indivíduo  $\mathbf{y}'$  melhorou (passo 11) é obtida pela Equação 7. Se nenhuma melhoria for alcançada,  $\eta_s^u(g)$  é considerado nulo.

Ao fim da geração  $g$ , um valor de recompensa é calculado (passo 16) pela Equação 8. Este valor de recompensa é realizado para atualizar a medida de qualidade da estratégia. Em seguida a medida de qualidade é atualizada pela Equação 6 (passo 17). No passo 18, os valores de probabilidade das diferentes estratégias são atualizados (Equação 5) a fim de serem utilizados na próxima geração.

O critério de parada do algoritmo é atingir o número máximo de avaliações (passo 20).

### IV. ANÁLISE DOS RESULTADOS OBTIDOS

Esta seção relata a análise dos resultados obtidos com os testes do algoritmo proposto. Inicialmente, foram feitos testes da versão adaptativa e não adaptativas implementadas, a fim de observar o impacto nos resultados, do uso da abordagem adaptativa nas estratégias de mutação. Foram testadas quatro versões não adaptativas, cada uma considerando o uso de uma das seguintes estratégias: *ED/Rand/1*, *ED/Current-to-Rand/1*, *ED/Rand/2* e *ED/Non-Linear*.

Para cada versão foram consideradas 30 execuções independentes. A instância real escolhida para os testes foi a proteína *Met-Enkephalin* (1PLW), formada por 5 aminoácidos (TYR-GLY-GLY-PHE-MET) e 75 átomos. A *Met-Enkephalin* foi utilizada por ser a proteína que tem sido escolhida como caso base para teste em diversos estudos para o PPEP, tanto considerando otimização mono-objetivo, quanto multiobjetivo [12], [14], [21]–[25]. A Tabela III apresenta os valores dos parâmetros da ED e PM utilizados nos testes realizados.

Tabela III. PARÂMETROS PARA O ALGORITMO PROPOSTO PARA O PPEP.

	Valores	Descrição
Parâmetros da ED		
N	100	Tamanho da População.
CR	1,0	Taxa de Cruzamento.
F	0,5	Fator de Mutação.
MAX-AV	500.000	Número Máximo de Avaliações.
Parâmetros do PM		
$\alpha$	0,3	Taxa de Adaptação.
$p_{min}$	0,05	Probabilidade Mínima de cada Estratégia.
K	4	Número de Estratégias do Conjunto.

A energia potencial livre (Equação 1) e a métrica RMSD foram os critérios de avaliação utilizados para a comparação entre proteína real e a proteína predita obtida pelo algoritmo. O RMSD (*Root Mean Square Deviation*) é uma medida utilizada para avaliar a semelhança entre a estrutura predita e a estrutura real da proteína [3]. A Equação 9 mostra como o RMSD é calculado.

$$RMSD_{C_\alpha}(a, b) = \sqrt{\frac{\sum_{i=1}^{N_a} |\mathbf{r}_{ai} - \mathbf{r}_{bi}|^2}{N_a}} \quad (9)$$

sendo que  $\mathbf{r}_{ai}$  e  $\mathbf{r}_{bi}$  são as posições dos átomos  $i$  nas estruturas  $a$  e  $b$  [18].

A Tabela IV apresenta os resultados obtidos pelo algoritmo adaptativo proposto comparado às quatro versões não adaptativas. A tabela apresenta os valores de energia obtidos referentes ao melhor indivíduo (que possui o menor valor de energia) das 30 execuções realizadas.

Tabela IV. RESULTADOS OBTIDOS PELO ALGORITMO PROPOSTO E VERSÕES NÃO ADAPTATIVAS.

Estratégia	Energia (kcal mol <sup>-1</sup> )
Adaptativa	<b>-35,814</b>
ED/Rand/1	-12,976
ED/Current-to-Rand/	-35,391
ED/Rand/2	-5,278
ED/Non-Linear	-33,242

A versão adaptativa obteve o melhor (menor) valor de energia potencial, quando comparada com as versões não adaptativas. Assim, a versão adaptativa para as estratégias de mutação foi a escolhida para o comparativo com a literatura.

A Tabela V apresenta os resultados obtidos pelo algoritmo proposto em comparação com abordagens de otimização mono-objetivo e multiobjetivo da literatura. A tabela mostra os valores de energia, os valores de RMSD considerando todos os átomos da proteína predita e IPLW (RMSD<sub>t</sub>) e os valores de RMSD considerando somente os carbonos alfa (RMSD<sub>C<sub>α</sub></sub>) da proteína predita e IPLW.

Tabela V. RESULTADOS OBTIDOS PELO ALGORITMO PROPOSTO E OUTRAS ABORDAGENS PARA *Met-Enkephalin*. CÉLULAS EM DESTAQUE REPRESENTAM OS MELHORES VALORES.

Algoritmo	Otimização	Energia (kcal mol <sup>-1</sup> )	RMSD <sub>C<sub>α</sub></sub> (Å)	RMSD <sub>t</sub> (Å)
Adaptativo proposto	Mono-objetivo	<b>-35,81</b>	1,913	3,424
PSAGC [21]	Mono-objetivo	-11,10	-	-
NOMAD-PSP [22]	Mono-objetivo	-30,14	<b>1,55</b>	-
SaDE/RGA [24]	Mono-objetivo	-12,43	-	-
ADEMO/D [25]	Multiobjetivo	-24,68	1,681	<b>2,866</b>
IPAES [12], [23]	Multiobjetivo	-20,56	1,740	3,605

Para a proteína *Met-enkephalin* o algoritmo proposto provou ser um bom otimizador considerando os valores de energia potencial, associado com bons valores de RMSD em comparação com a literatura.

## V. CONCLUSÕES

Neste trabalho foi apresentado um algoritmo evolutivo adaptativo baseado no método evolucionário da evolução diferencial aplicado ao problema da predição da estrutura de proteínas. A abordagem adaptativa considerada foi a *Probability Matching* a fim de controlar a seleção da melhor estratégia de mutação, dentre um conjunto de quatro estratégias disponíveis.

A principal contribuição deste trabalho foi o desenvolvimento de um otimizador mono-objetivo a fim de avaliar do uso do método de evolução diferencial adaptativa com *Probability Matching* para o problema da predição da estrutura de proteínas, considerando a abordagem *ab initio* e a modelagem computacional livre de látece. Os resultados obtidos pelo algoritmo proposto foram competitivos em comparação com a literatura, principalmente em termos de minimização nos valores de energia potencial.

Como trabalhos futuros pretende-se investigar outros métodos adaptativos, assim como realizar testes com outras proteínas.

## AGRADECIMENTOS

Os autores agradecem à Fundação Araucária pela bolsa de iniciação científica concedida e ao CNPq (Projeto 456179/2014-3) pelo apoio financeiro.

## REFERÊNCIAS

- [1]: J. M. Bujnicki, Ed., *Prediction of Protein Structures, Functions, and Interactions*. John Wiley, 2009.
- [2]: F. E. Cohen and J. W. Kelly, "Therapeutic approaches to protein-misfolding diseases," *Nature*, vol. 426, pp. 905–909, December 2003.
- [3]: A. Tramontano, *Protein Structure Prediction: Concepts and Applications*. John Wiley and Sons, 2006.
- [4]: D. J. Osguthorpe, "Ab initio protein folding," *Current Opinion in Structural Biology*, vol. 10, no. 2, pp. 146 – 152, 2000. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0959440X00000671>
- [5]: A. D. Mackerel, C. L. Brooks, L. Nilsson, B. Roux, Y. Won, and M. Karplus, *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, ser. The Encyclopedia of Computational Chemistry. John Wiley & Sons: Chichester, 1998, vol. 1, pp. 271–277.
- [6]: R. Storn and K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997.
- [7]: D. E. Goldberg, "Probability matching, the magnitude of reinforcement, and classifier system bidding," *Mach. Learn.*, vol. 5, pp. 407–425, October 1990.
- [8]: A. Lehninger, D. Nelson, and M. Cox, *Lehninger princípios de bioquímica*. ARTMED, 2011. [Online]. Available: <https://books.google.com.br/books?id=Ziu4uAAACAAJ>
- [9]: C. Brändén and J. Tooze, *Introduction to Protein Structure*, ser. Introduction to Protein Structure Series. Garland Pub., 1999. [Online]. Available: <https://books.google.com.br/books?id=zsWcpqrgG74C>
- [10]: W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *Journal of The American Chemical Society*, vol. 117, pp. 5179–5197, 1995.
- [11]: M. Mijajlovic, M. J. Biggs, and D. P. Djurdjevic, "On potential energy models for ea-based ab initio protein structure prediction," *Evol. Comput.*, vol. 18, pp. 255–275, June 2010.
- [12]: V. Cutello, G. Narzisi, and G. Nicosia, "A multi-objective evolutionary approach to the protein structure prediction problem," *J R Soc Interface*, vol. 3, no. 6, pp. 139–51, 2006.
- [13]: D. Becerra, A. Sandoval, D. Restrepo-Montoya, and L. F. Niño, "A parallel multi-objective ab initio approach for protein structure prediction," in *BIBM*, T. Park, S. K.-W. Tsui, L. Chen, M. K. Ng, L. Wong, and X. Hu, Eds. IEEE Computer Society, 2010, pp. 137–141.
- [14]: A. M. Anile, V. Cutello, G. Narzisi, G. Nicosia, and S. Spinella, "Determination of protein structure and dynamics combining immune algorithms and pattern search methods," *Natural Computing*, vol. 6, no. 1, pp. 55–72, 2007.
- [15]: D. Dasgupta, D. Becerra, A. Banceanu, F. Nino, and J. Simien, "A parallel framework for multi-objective evolutionary optimization," in *IEEE Congress on Evolutionary Computation (CEC), 2010*, July 2010, pp. 1–8.
- [16]: A. P. Engelbrecht, *Computational Intelligence: An Introduction*, 2nd ed. Wiley Publishing, 2007.
- [17]: C.-L. Lu, S.-Y. Chiu, C.-H. Hsu, and S.-J. Yen, "Enhanced differential evolution based on adaptive mutation and wrapper local search strategies for global optimization problems," *Journal of Applied Research and Technology*, vol. 12, no. 6, pp. 1131 – 1143, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1665642314716724>
- [18]: S. M. Venske, R. A. Gonçalves, E. M. Benelli, and M. R. Delgado, "A multiobjective algorithm for protein structure prediction using adaptive differential evolution," in *Brazilian Conference on Intelligent Systems, BRACIS 2013, Fortaleza, CE, Brazil, 19-24 October, 2013*, 2013, pp. 263–268. [Online]. Available: <http://dx.doi.org/10.1109/BRACIS.2013.51>

- [19] K. Sindhya, S. Ruuska, T. Haanpaa, and K. Miettinen, "A new hybrid mutation operator for multiobjective optimization with differential evolution," *Soft Comput.*, vol. 15, no. 10, pp. 2041–2055, 2011.
- [20] D. Thierens, "An adaptive pursuit strategy for allocating operator probabilities," in *Conference on Genetic and Evolutionary Computation*, 2005, pp. 1539–1546.
- [21] Y.-L. Liu and L. Tao, "An improved parallel simulated annealing algorithm used for protein structure prediction," in *Machine Learning and Cybernetics, 2006 International Conference on*, 2006, pp. 2335–2338.
- [22] G. Nicosia and G. Stracquadiano, "Generalized pattern search and mesh adaptive direct search algorithms for protein structure prediction," in *Proceedings of the 7th International Conference on Algorithms in Bioinformatics*, ser. WABI'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 183–193.
- [23] V. Cutello, G. Narzisi, and G. Nicosia, "Computational studies of peptide and protein structure prediction problems via multiobjective evolutionary algorithms," in *Multiobjective Problem Solving from Nature*, ser. Natural Computing Series, J. Knowles, D. Corne, K. Deb, and D. Chair, Eds. Springer Berlin Heidelberg, 2008, pp. 93–114.
- [24] S. Sudha, S. Baskar, and S. Krishnaswamy, "Protein tertiary structure prediction using evolutionary algorithms," *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)*, vol. 3, no. 3, pp. 338–348–595, 2013.
- [25] S. M. Venske, R. A. Gonçalves, E. B. Martin, and M. R. Delgado, "A multiobjective algorithm for protein structure prediction using adaptive differential evolution," in *Brazilian Conference on Intelligent Systems*, 2013, pp. 263–268.