

1 Introdução

Hello (Kendrew et al., [1958](#))

2 Revisão de literatura

2.1 Métodos de atribuição de estruturas secundárias

2.1.1 DSSP

Em 1983, Kabsch e Sander publicaram o algoritmo de atribuição de estruturas secundárias de proteínas que viria a ser o mais utilizado até os dias atuais, o DSSP (*Dictionary of Protein Secondary Structure*).

No trabalho, os autores afirmam que a atribuição de estruturas secundárias a partir das coordenadas atômicas de estruturas proteicas é um problema de reconhecimento de padrões. Nesse contexto, eles optaram por identificar esses padrões através de ligações de hidrogênio entre átomos da cadeia principal ao invés de ângulos Φ e Ψ ou de posições relativas de C_α . A justificativa utilizada foi que a presença ou ausência de ligações de hidrogênio poderiam ser avaliadas por um simples critério energético, enquanto que outras características precisariam do ajuste de um número maior de parâmetros.

As ligações de hidrogênio foram definidas por eles utilizando um modelo eletrostático. Nesse modelo, uma ligação de hidrogênio HB ocorrerá se, e somente se, a energia E for menor que -0.5 kcal/mol. Para o cálculo são utilizadas as cargas parciais $+q_1, -q_1$ nos átomos C e O , e $-q_2, +q_2$ nos átomos N e H , onde $q_1 = 0.42e$ e $q_2 = 0.20e$.

$$E < -0.5 \text{ kcal/mol} \implies HB = \text{Verdade} \quad (2.1)$$

onde

$$E = q_1 q_2 (1/r(ON) + 1/r(CH) - 1/r(OH) - 1/r(CN)) * f \quad (2.2)$$

Na equação (2.2), $r(AB)$ é a distância interatômica entre A e B em ângstroms e o fator dimensional $f = 332$.

Os autores afirmam que, por este modelo, uma boa ligação de hidrogênio teria aproximadamente -3 kcal/mol. Assim, a escolha de um limiar em -0.5 kcal/mol torna o modelo mais tolerante à erros nas coordenadas atômicas e à ligações de hidrogênios bifurcadas (Kabsch and Sander, 1983).

Uma vez definido o modelo para identificar ligações de hidrogênio, essas são testadas e anotadas na cadeia polipeptídica em duas classes, ou padrões, elementares: (1) padrão *n-Turn* e (2) padrão *Bridge*.

O padrão *n-Turn*, onde $n \in \{3, 4, 5\}$, apresentam uma ligação de hidrogênio entre o CO do resíduo i e o NH do resíduo $i + n$.

$$n\text{-Turn} \Leftarrow Hbond(i, i + n), n \in \{3, 4, 5\} \quad (2.3)$$

O padrão *Bridge* pode ocorrer de duas formas, a paralela e a antiparalela.

$$Parallel\ Bridge \Leftarrow [Hbond(i - 1, j) \wedge Hbond(j, i + 1)] \vee [Hbond(j - 1, i) \wedge Hbond(i, j + 1)] \quad (2.4)$$

$$Antiparallel\ Bridge \Leftarrow [Hbond(i, j) \wedge Hbond(j, i)] \vee [Hbond(i - 1, j + 1) \wedge Hbond(j - 1, i + 1)] \quad (2.5)$$

Sendo que as sequências $i - 1, i, i + 1$ e $j - 1, j, j + 1$ não apresentam sobreposição (*overlap*) de resíduos entre si.

As ocorrências dos padrões elementares *n-Turn* e *Bridge* ao longo da cadeia polipeptídica é utilizada para a atribuição dos elementos de estrutura secundária. Como exemplos, a repetição de padrões *4-Turn* consecutivos indica a ocorrência de uma hélice α , enquanto que resíduos consecutivos que apresentem padrões *Bridge* formam uma fita de uma folha β .

No trabalho, os autores mencionam que esse algoritmo proposto produz hélices mais curtas, com um resíduo a menos em cada extremidade, em relação a anotação seguindo as regras da IUPAC. Outra característica do algoritmo é que hélices que apresentem algumas ligações de hidrogênio ausentes, são mantidas como uma hélice única ao invés de múltiplas hélices com *kinks*. O mesmo ocorre com resíduos que formam *bulges* em fitas, sendo os mesmo também anotados como parte da fita.

2.1.2 Stride

stride

2.1.3 KAKSI

KAKSI é um método de atribuição de estruturas secundárias proposto por Martin e colaboradores 2005. Esse método foi desenvolvido utilizando padrões de distâncias entre C_α e de ângulos Φ e Ψ .

Resumidamente, a heurística de atribuição de estrutura secundárias busca primeiramente por hélices, sendo que um resíduo é classificado como hélice se atender aos critérios de distâncias entre C_α ou aos critérios de ângulos Φ e Ψ . Em seguida, é feito a classificação dos resíduos em fitas. Somente os resíduos não-hélice podem ser classificados em fitas, e para tal, eles precisam atender aos critérios de distâncias entre C_α e aos critérios de ângulos Φ e Ψ .

1. Critérios para classificação de hélices:

Distância entre C_α

Todas as distâncias entre C_α em uma janela de seis resíduos $[i, i+5]$ precisam

estar dentro do intervalo $[M_\alpha - \varepsilon_H \times SD_\alpha, M_\alpha + \varepsilon_H \times SD_\alpha]$, onde M_α e SD_α são respectivamente a média e o desvio padrão observado em hélices α .

Ângulos Φ e Ψ

Todos os pares de ângulos Φ e Ψ em uma janela de quatro resíduos precisam satisfazer as condições: $\Phi < 0^\circ$ e $-90^\circ < \Psi < 60^\circ$. Além disso, ao menos um par de ângulos precisa estar em uma região densamente povoada, com densidade $> \sigma_H$.

2. Critérios para classificação de folhas:

Distância entre C_α

Todas as distâncias entre C_α em duas janelas de três resíduos precisam estar no intervalo $[M_\beta - \varepsilon_b \times SD_\beta, M_\beta + \varepsilon_b \times SD_\beta]$, onde M_β e SD_β são respectivamente a média e o desvio padrão observado em folhas β .

Ângulos Φ e Ψ

Cada par de ângulos Φ e Ψ presente na zona povoada de resíduos em folhas β incrementa um contador em 1. Quando um resíduo central da janela apresenta $-120^\circ < \Psi < 50^\circ$, o contador é reiniciado em zero. Esse critério é satisfeito se o contador $\geq \sigma_b$.

Além destes critérios, há critérios para detecção de *kink* em hélices e um critério de correção de segmentos, que altera um resíduo para o estado de *coil* quando ocorre continuidade de segmentos hélice-fita, ou fita-hélice, tornando as hélices 1 resíduo menor.

Os vários parâmetros necessários ao método foram ajustados empiricamente utilizando um conjunto de 2880 domínios estruturais, com identidade sequencial inferior à 40%, resolvidos por cristalografia e com resolução superior a 2.25Å.

2.1.4 PROSS

pross

2.2 Métodos de predição de estruturas secundárias

2.2.1 Primeira geração (1957-1978)

primeira geração - chou e fasman - gor

2.2.2 Segunda geração (1983-1992)

segunda geração - GORIII

2.2.3 Terceira geração

terceira geração - aprendizado de máquina (NN) - PHDsec

2.2.4 Quarta geração

quarta geração - aprendizado de máquina (NN) + dados evolutivos (PSSM) - PSIPred

3 Resultados

3.1 Análise de métodos de atribuição de estrutura secundária

Diferentemente de vários métodos publicados de predição de estrutura secundária que utilizaram os resultados de atribuição provenientes de apenas um método, comumente o DSSP, nós optamos por utilizar os resultados de um consenso entre quatro métodos. Consequentemente, é importante analisar como esses métodos diferem entre si.

As diferenças entre as metodologias de atribuição foram previamente discutidas na seção 2.1. Nesta seção, analisaremos as diferenças nas estruturas secundárias atribuídas aos resíduos.

As análises a seguir foram feitas entre os resíduos de 6749 proteínas do banco de dados Top8000-HOM50 que tiveram estrutura secundária atribuída por todos os métodos. Isso exclui 5,19% dos resíduos que não tiveram a estrutura secundária atribuída por não terem sido visualizados na estrutura experimental e 0,05% dos resíduos que não tiveram sua estrutura atribuída por um ou mais programas devido à possíveis falhas dos programas.

Notamos que, entre os quatro métodos de atribuição analisados, há a ocorrência de uma variação considerável entre as estruturas secundárias atribuídas para cada resíduo. A figura 3.1 demonstra a similaridade observada entre os métodos. Como esperado, a maior similaridade é observada entre os métodos DSSP e STRIDE, uma vez que o último é inspirado no primeiro. Entretanto, a similaridade entre os demais métodos encontra-se na faixa entre 83% e 85%. A porcentagem de resíduos que apresentaram consenso entre os quatro métodos foi de apenas 74,43%.

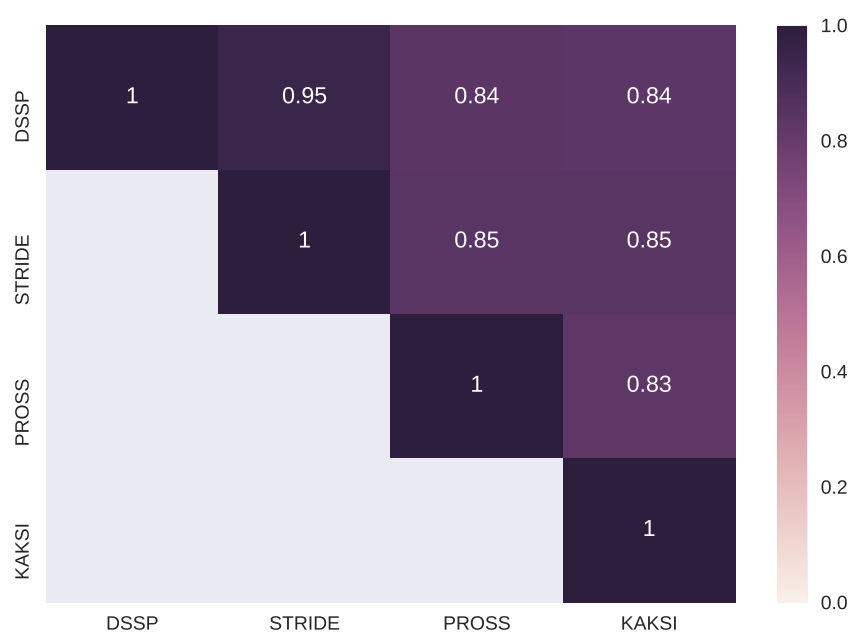


Figure 3.1: Similaridade entre as estruturas secundárias atribuídas para cada resíduo entre os quatro métodos de atribuição de estruturas secundárias: DSSP, STRIDE, KAKSI e PROSS.

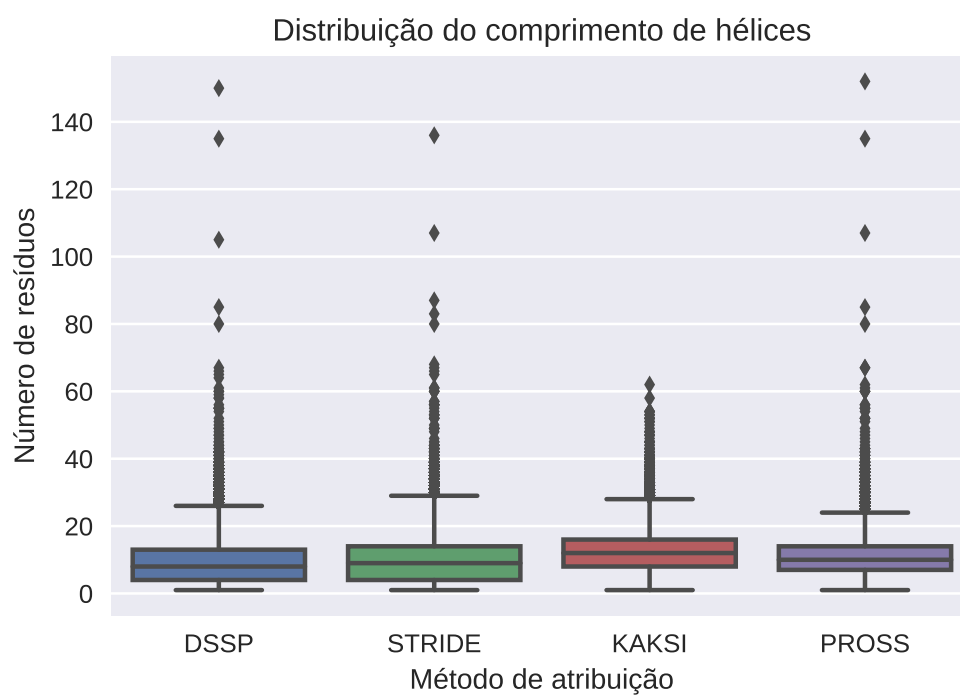


Figure 3.2: Distribuição do comprimento de hélices por método de atribuição de estrutura secundária.

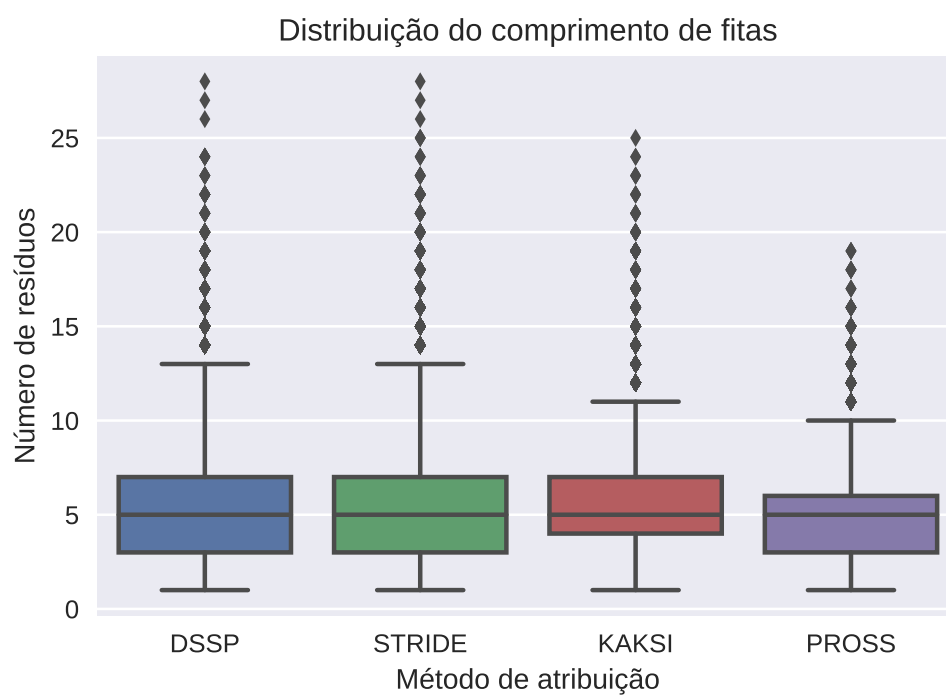


Figure 3.3: Distribuição do comprimento de fitas por método de atribuição de estrutura secundária.

Bibliography

- Kabsch, Wolfgang and Christian Sander (1983). “Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features”. In: *Biopolymers* 22.12, pp. 2577–2637. DOI: [10.1002/bip.360221211](https://doi.org/10.1002/bip.360221211).
- Kendrew, John C et al. (1958). “A three-dimensional model of the myoglobin molecule obtained by x-ray analysis”. In: *Nature* 181.4610, pp. 662–666. DOI: [10.1038/181662a0](https://doi.org/10.1038/181662a0).
- Martin, Juliette et al. (2005). In: *BMC Structural Biology* 5.1, p. 17. DOI: [10.1186/1472-6807-5-17](https://doi.org/10.1186/1472-6807-5-17).