

Introdução à Bioinformática

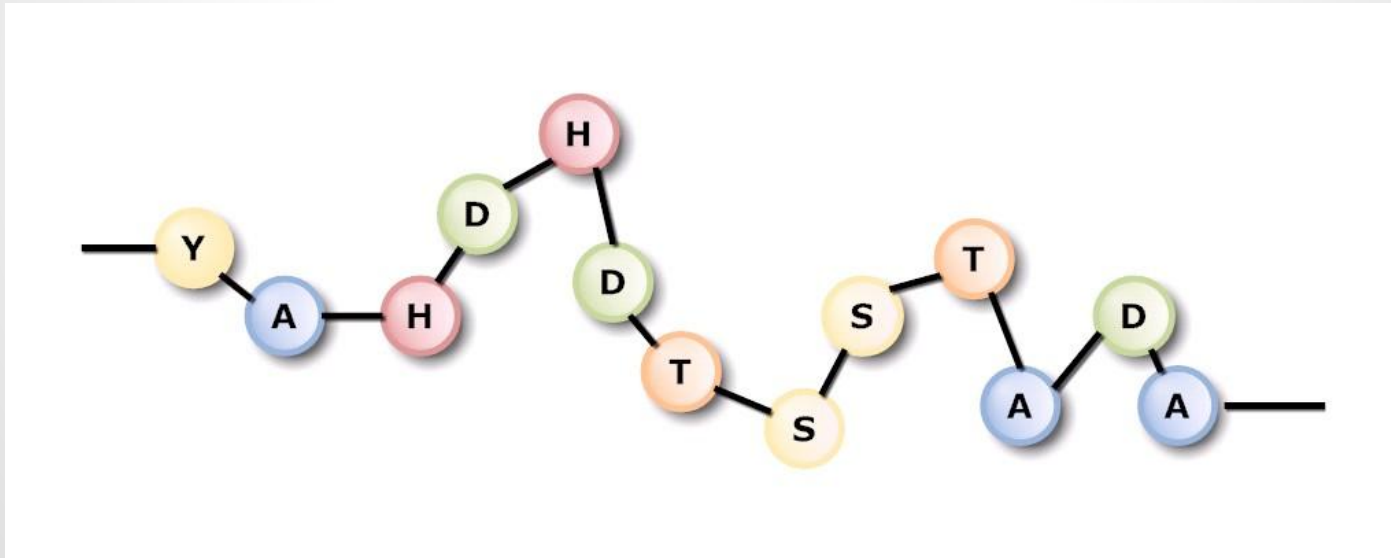
TP1 - Alinhamento

Hugo Richard Amaral

Luís Eduardo Oliveira Lizardo

Problema

Identificar se a função de uma proteína é alterada por mutações em sua sequência.



Objetivo

- Utilizar um algoritmo de alinhamento de sequências protéicas;
- Sugerir modificações que permitam resgatar a funcionalidade da proteína.

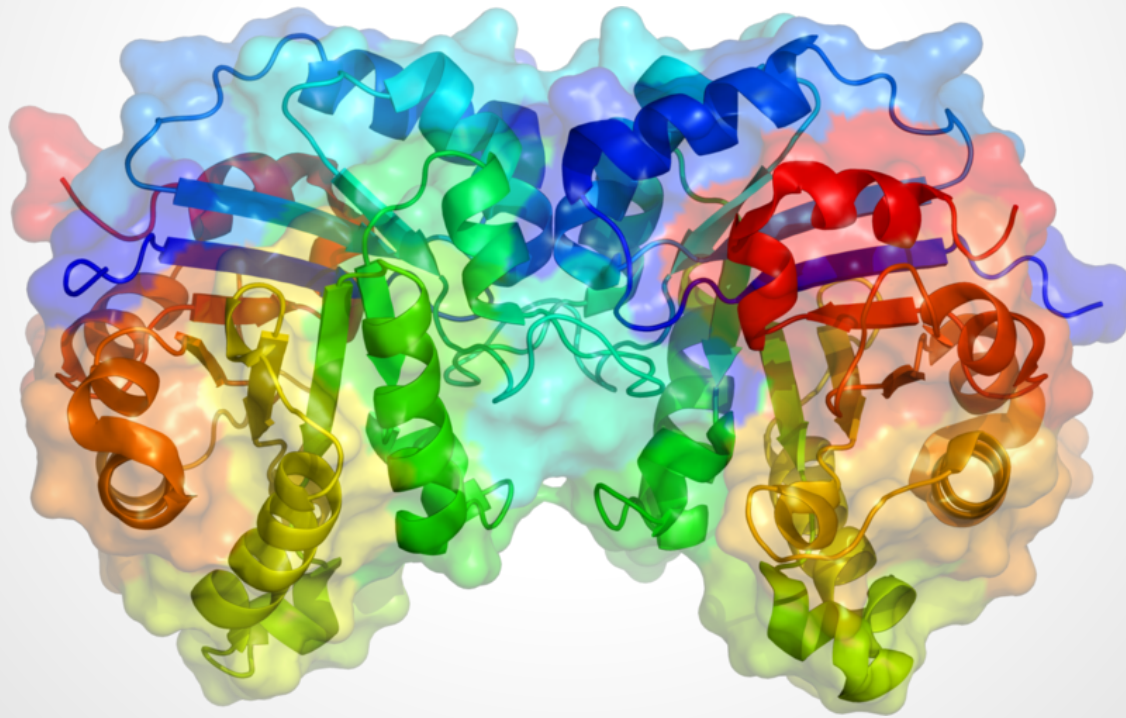
$$H = \begin{pmatrix} 0 & 2 & 2 & 1 & 0 \\ 0 & 2 & 4 & 3 & 2 \\ 0 & 1 & 3 & 5 & 5 & 4 & 3 & 2 & 1 & 0 \\ 0 & 0 & 2 & 5 & 8 & 7 & 6 & 5 & 4 & 3 \\ 2 & 1 & 1 & 4 & 7 & 6 & 5 & 4 & 3 & 2 \\ 1 & 4 & 3 & 3 & 6 & 5 & 4 & 3 & 2 & 1 \end{pmatrix}$$

Sequência 1: A T T C A - C C - G - T - A

Sequência 2: - T T C A A - - T G G T C -

Triose-fosfato isomerase - 2YPIA

248 aminoácidos



Fonte: Wikimedia Commons

Metodologia - Alinhamento

- Algoritmo de Needleman-Wunsch:
 - Alinhamento global de duas sequências;
 - Programação dinâmica;
 - Mede a similaridade entre as sequências com base em uma tabela de pontuação;
- Quanto maior o *score* de alinhamento, mais similares são as sequências.

Fonte: Needleman, Saul B.; and Wunsch, Christian D. (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins"

Metodologia - 1ª Etapa

- Identificar aminoácidos com características semelhantes e agrupá-los.
- **Objetivo:** a troca desses aminoácidos causam pouco impacto na função da proteína.



Fonte: Vermont ViewER MutatiON Tool

Metodologia - 2ª Etapa

- Alinhamento entre dTIM e 2YPIA
- **Objetivo:** identificar os pontos em que há alteração.

```
2ypia:  MARTFFVGGNFKLNGSKQSIKEIVERLNTASIPENVEVVICPPATYLDYSVSLVKKPQVTVGAQNAYLKASGAFTGENSV 80
        MART FVGGN K NG K  KE VE L  A P  VEVV  PPA YLD      K    V AQN Y  A GAFTGE S
dTIM:   MARTPFVGGNWKMNNGTKAEAKELVEALK-AKLPDDVEVVVAPPVYLDTAREALKGSKIKVAAQNCYKEAKGAFTGEISP 80

2ypia:  DQIKDVGAKWVILGHSERRSYFHEDDKFIADKTKFALGQGVGVILCIGETLEEKKAGKTLDVVERQLNAVLEEVKD-WTN 160
        KD GA  VILGHSERR YF E D   A K   AL  G  VI CIGETLEE  AGKT  VV RQ  A L    D W N
dTIM:   EMLKDLGADYVILGHSERRHYFGETDELVAKKVAHALEHGLKVIACIGETLEEREAGKTEEVVFRQTKALLAGLGDEWKN 160

2ypia:  VVVAYEPVWAIGTGLAATPEDAQDIHASIRKFLASKLGDKAASELRILYGGSSANGSNAVTFKDKADVDGFLVGGASLKPE 240
        VV AYEPVWAIGTG  ATPE AQ  HA IRK LA      A  RILYGGSS  NA      D DGFLVGGASLKPE
dTIM:   VVIAYEPVWAIGTGKTATPEQAQEVHAFIRKWLAENVSAEVAESVRILYGGSVKPANAKELAAQPDIDGFLVGGASLKPE 240

2ypia:  FVDIINSRN 249
        F DIINSRN
dTIM:   FLDIINSRN 249
```


Metodologia - 3ª Etapa

- Alinhar a proteína selvagem com as proteínas da família;
- **Objetivo:** identificar nas proteínas da família as mesmas alterações verificadas na dTIM.

```
2YPIA:  MARTFFVGGNFKLNGSKQSIKEIVERLNTASIPENVEVVICPPATYLDYSVSLVKKPQVTVGAQNAYLKASGAFTGENSV 80
        R FFVGGN K NG K S E   LN A       EVV   P   YLD               V AQN Y   GAFTGE S
8TIMA:  APRKFFVGGNWKMNQDKKSLGELIHTLNGAKLSADTEVVCGAPSIYLDFAHQKL-DAKIGVAAQNCYKVPKGAFTGEISP 80

2YPIA:  DQIKDVGAKWVILGHSERRSYFHEDDKFIADKTKFALGQGVGVILCIGETLEEKKAGKTLDVVERQLNAVLEEVKDWTNV 160
        IKD GA WVILGHSERR  F E D   I K   AL  G GVI CIGE L E  AG T  VV  Q  A   VKDW  V
8TIMA:  AMIKDIGAAWVILGHSERRHVFGEDELIGQKVAHALAEGLGVIACIGEKLDEREAGITEKVVFEQTKAIADNVKDWSKV 160

2YPIA:  VVAYEPVWAIGTGLAATPEDAQDIHASIRKFLASKLGDKAASELRILYGGSANGSNAVTFKDKADVDGFLVGGASLKPEF 240
        V AYEPVWAIGTG  ATP  AQ  H   R L       D A   RI YGGS  G N           DVDGFLVGGASLKPEF
8TIMA:  VLAYEPVWAIGTGKTATPQQAQEVHEKLRGWLKTHVSDAVAQSTRIIYGGSVTGGNCKELASQHDVDGFLVGGASLKPEF 240

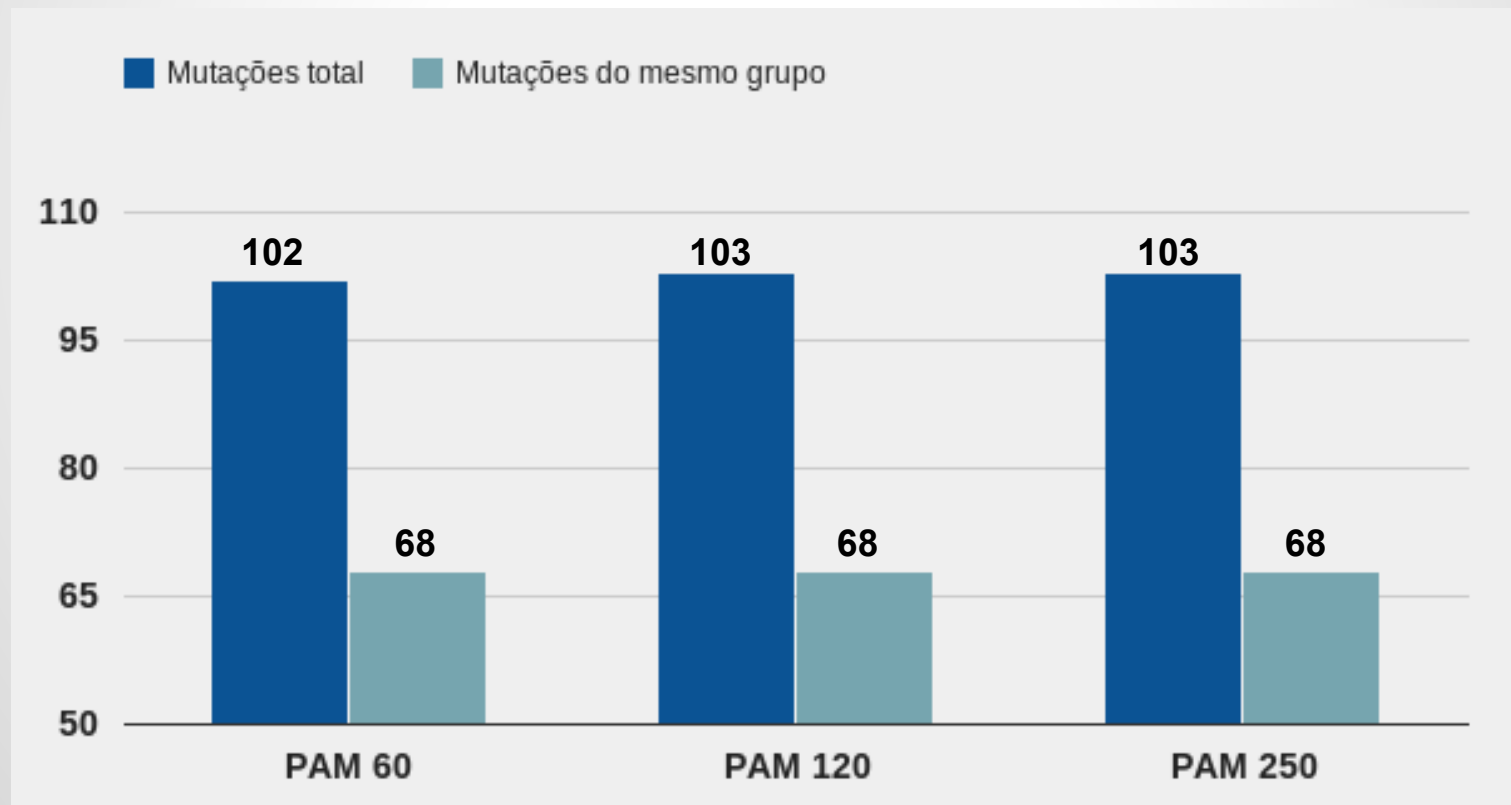
2YPIA:  VDIINSRN 248
        VDIIN
8TIMA:  VDIINAKH 248
```

Metodologia - 4ª Etapa

- Comparar as mutações ocorridas na proteína selvagem com as mutações nas proteínas da família.
- **Objetivo:** contabilizar quais alterações que ocorreram tanto na selvagem como na família.

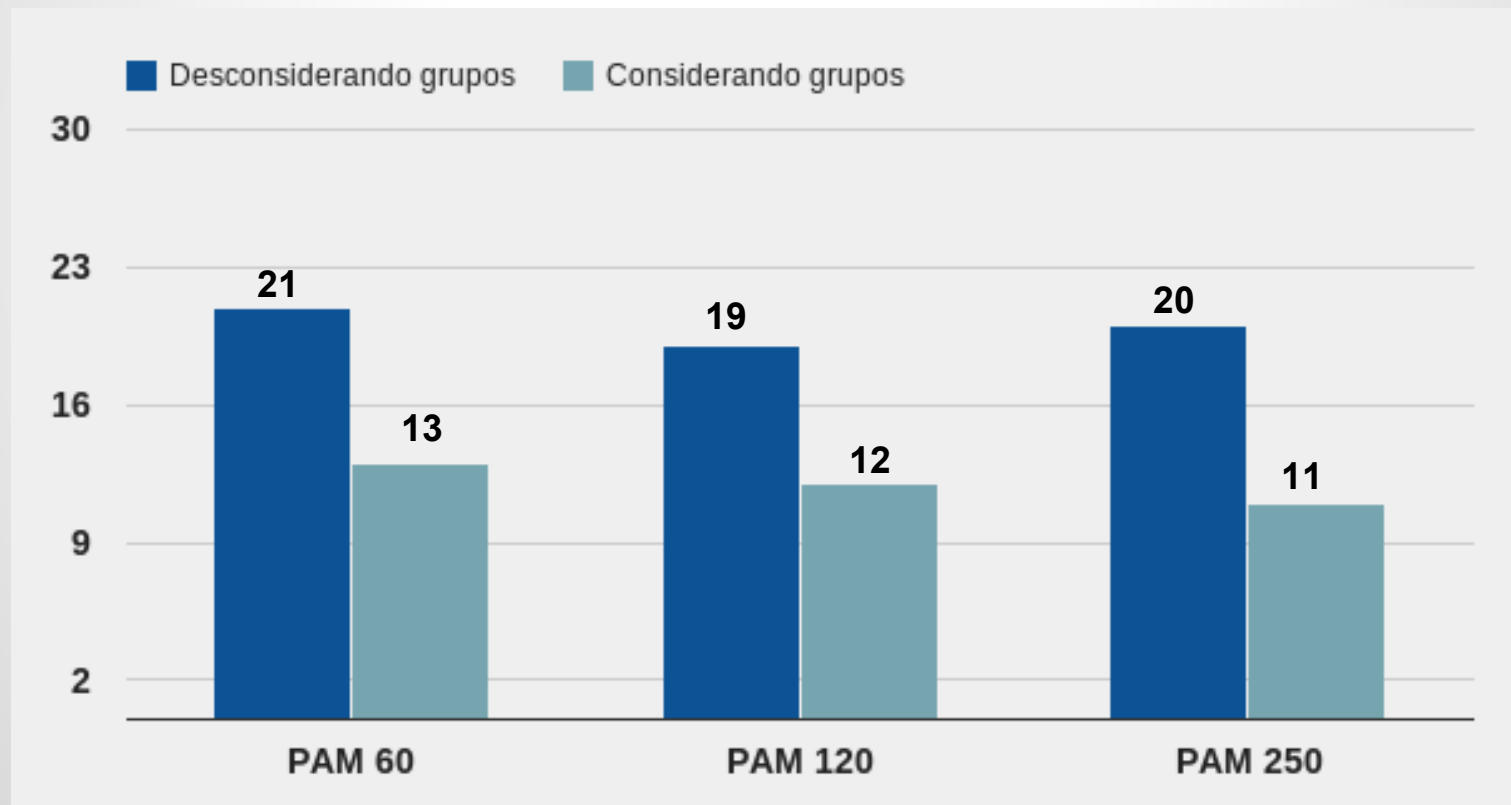
Resultados

Quantidade de mutações entre dTIM e 2YPIA:



Resultados

Quantidade de mutações entre dTIM e 2YPIA que alteram a função:



Resultados

- Para restaurar a função da proteína mutada, consideramos realizar as mutações na dTIM para que ela se torne a proteína selvagem ou uma das proteínas família, sendo a escolhida aquela cuja pontuação de alinhamento for a maior;
- Para tanto, alinhamos a proteína mutada com todas as proteínas da família a fim de encontrar a mais semelhante.

Resultados

- Proteína com maior semelhança: 1SW3A
- Pontuação: 941
- Número de mutações necessárias: $61/102 = 59,8\%$

```
dTIM:  MA-RTPFVGGNWKMNGTAEAKELVEALK-AKLPDDVEVVVAPPAVYLDTAREALKGSKIKVAAQNCYKEAKGAFTGEIS 80
      MA R  FVGGNWKMNG K   EL   L  AKL  D EVV   P  YLD AR  L   KI VAAQNCYK  KGFTGEIS
1SW3A:  MAPRKFFVGGNWKMNWDKKSLGELIHTLNGAKLSADTEVVCGAPSIYLDFAEQKL-DAKIGVAAQNCYKVPKGFTGEIS 80

dTIM:  PEMPLKDLGADYVILGHSERRHYFGETDELVAKKVAHALEHGLKVIACIGETLEEREAGKTEEVVFRQTKALLAGLGDEWK 160
      P M KD GA  VILGHSERRH FGE DEL   KVAHAL  GL VIACIGE L EREAG TE VVF QTKA  A      W
1SW3A:  PAMIKDIGAAWVILGHSERRHVFGEDELIGQKVAHALAEGLGVIACIGEKLDEREAGITEKVVFEQTKA-IADNVKDWS 160

dTIM:  NVVIAIEPVWAIGTGKTATPEQAQEVHAFIRKWLAENVSAEVAESVRILYGGSVKPANAKELAAQPDIDGFLVGGASLKP 240
      VV AYEPVWAIGTGK ATP QAQEVH   R WL   VS  VA S RI YGGSV   N KELA Q  D DGFLVGGASLKP
1SW3A:  KVVLAYEPVWAIGTGKVATPQQAQEVHEKLRGWLKSHVSDAVAQSTRIIYGGSVTGGNCKELASQHDVDGFLVGGASLKP 240

dTIM:  EFLDIINSRN 250
      EF DIIN
1SW3A:  EFVDIINAKH 250
```

Conclusão

- O algoritmo de alinhamento de Needleman-Wunsch é simples e fácil de implementar.
- Porém, para tentar resolver o problema ainda é preciso uma boa tabela de pontuação, gap inicial, aminoácidos semelhantes, sítio ativo, entre outros;
- Mesmo assim, pode servir como um passo inicial para resolução do problema.



Perguntas



Obrigado