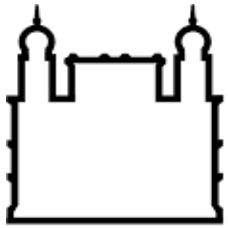


Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

PROCC
PROGRAMA DE COMPUTAÇÃO CIENTÍFICA

Biologia computacional usando o R

Leo Bastos (PROCC/Fiocruz)

Parceiros:

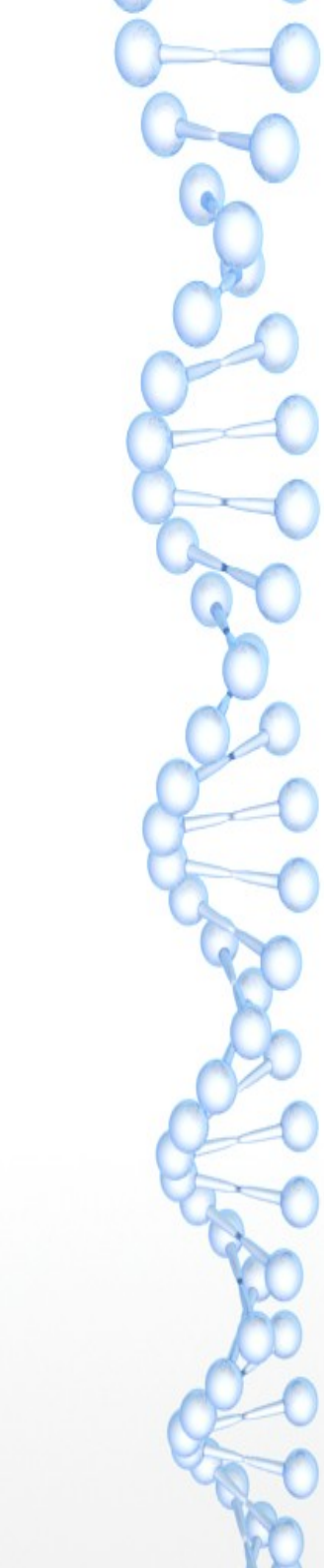
Maurício Costa (PROCC)

Gisele Rocha (doutoranda BCS)

Grupo de Biofísica Computacional e Modelagem Molecular

Caution!

- Essa apresentação ~~pode conter~~ contém erros básicos de biologia/biofísica!
- Meus colaboradores estão isentos de culpa.
- Peço desculpas adiantadamente.



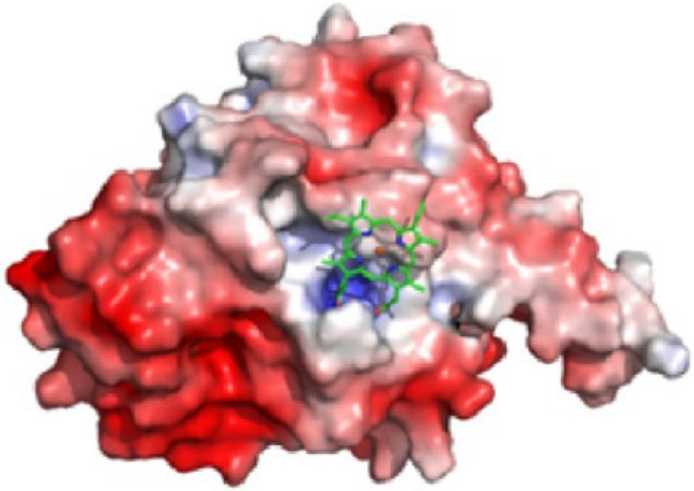
Biologia computacional

- Área interdisciplinar:
  - Biologia, física, química
  - Matemática aplicada, **estatística**
  - Computação
- Algumas sub-áreas:
  - Biologia de sistemas
  - Genônica computacional
  - Biologia evolucionária
  - **Modelagem molecular estrutural**
  - etc

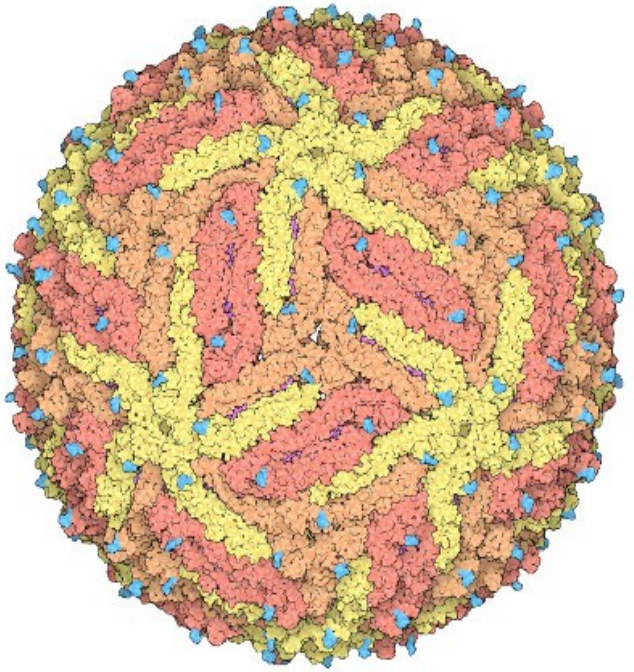

Experimental Biologist
Computational Biologist
NO EATING
NO DRINKING
NO BREATHING
NO OPEN TOE SHOES
NO SHORTS
NO MINI SKIRTS
ALWAYS LAB COATS ON
LIFE IS BEAUTIFUL
www.biocomicals.com

Modelagem molecular

É a ciência e arte que estuda a estrutura molecular através da computação usualmente partindo de dados experimentais.



Falcipaína-2
Cristalografia de raios X
Hogg et al. 2006



Capsídeo ZIKV
cryo EM
Sirohi, et al. 2016

Estudo da dinâmica molecular

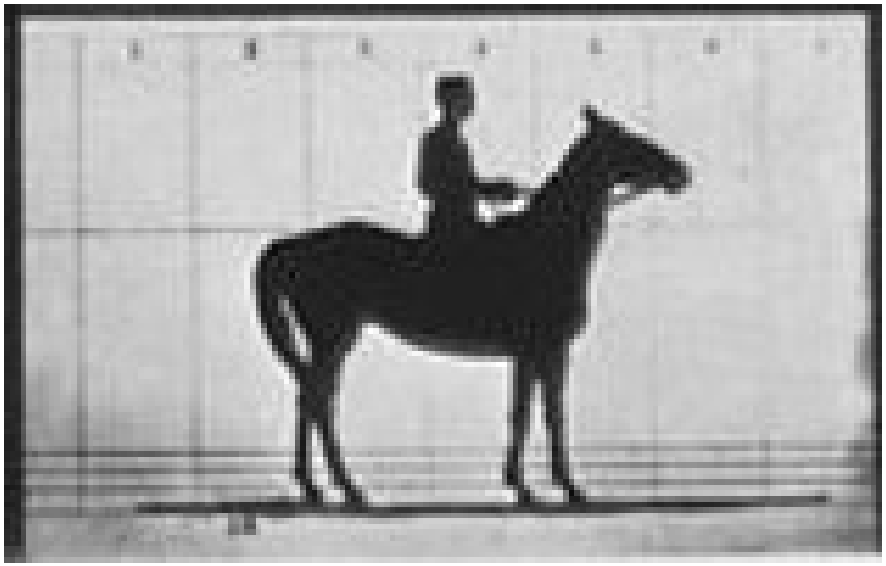
- O estudo do movimento molécular é feito a partir da estrutura 3D estática obtida experimentalmente.

RCSB PDB
PROTEIN DATA BANK

Contém estrutura de 118587 macromoléculas!

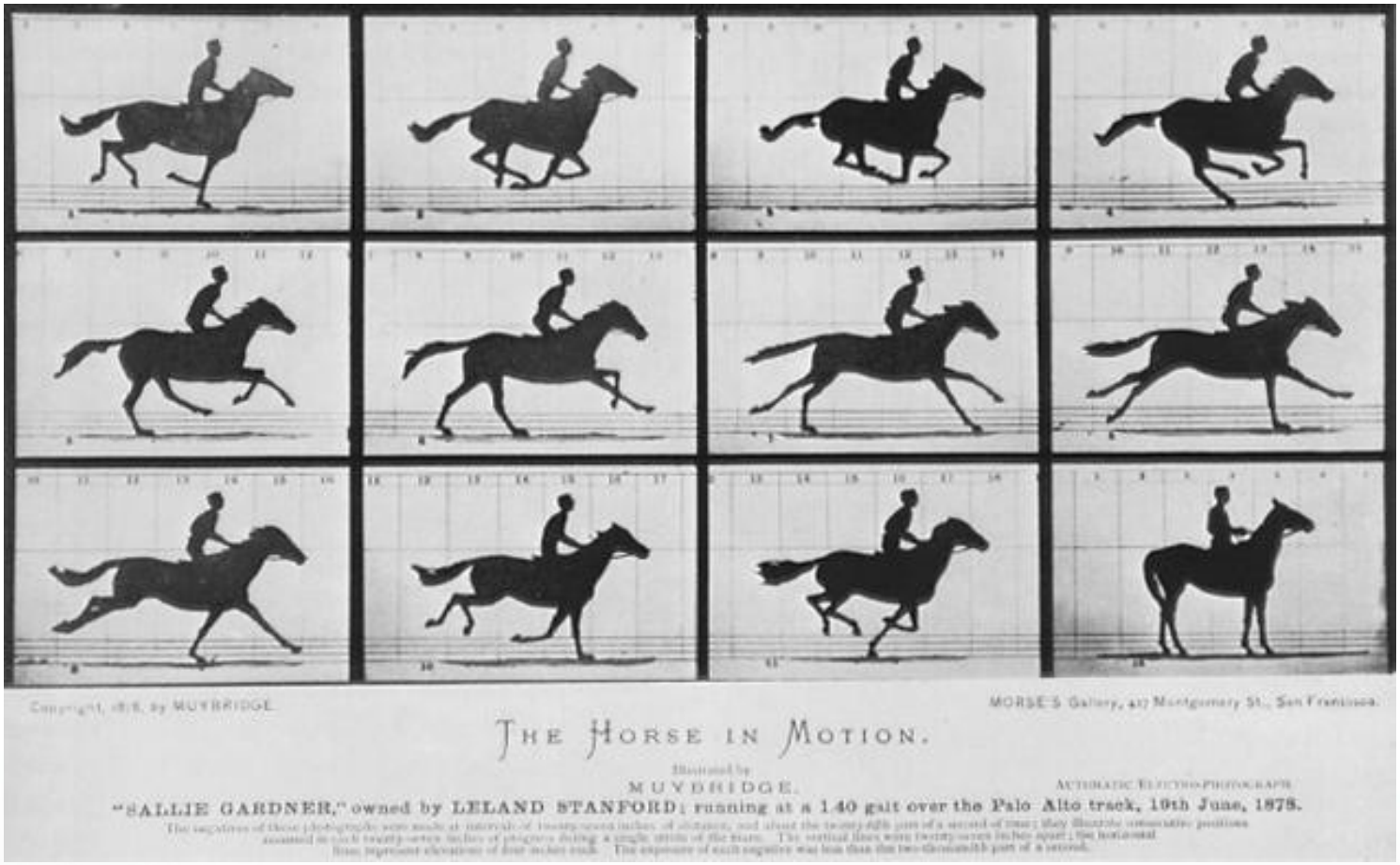
- Computacionalmente os movimentos atômicos de biomoléculas podem ser descritos via:
  - Dinâmica molecular
  - Análise de modos normais

Dinâmica molecular



Estrutura da proteína (experimental)

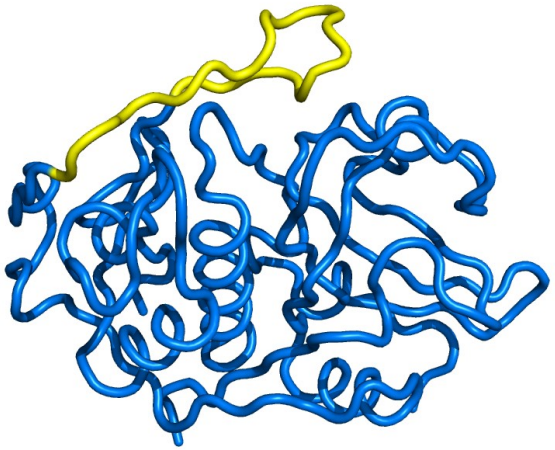
Dinâmica molecular



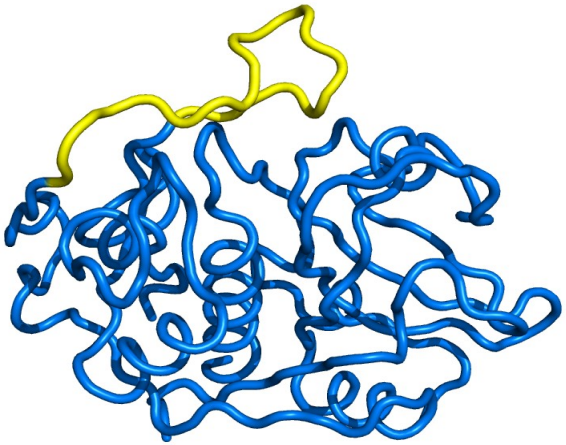
Movimentos da proteína (computacional)

Dinâmica molecular

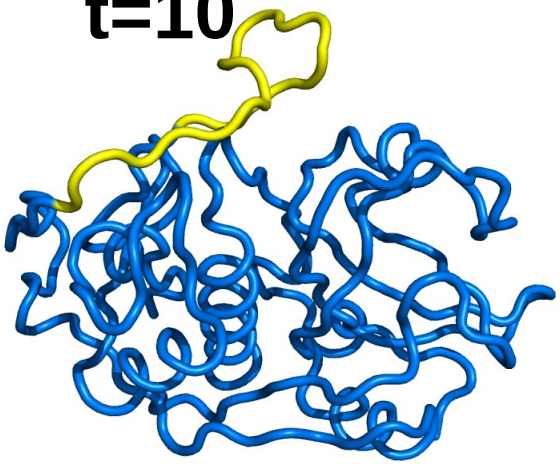
t=0



t=1

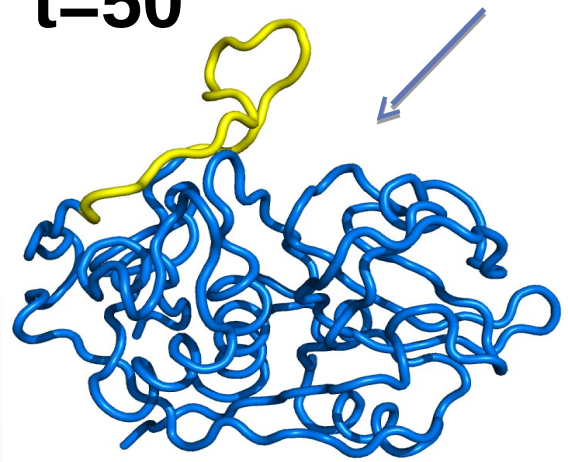


t=10



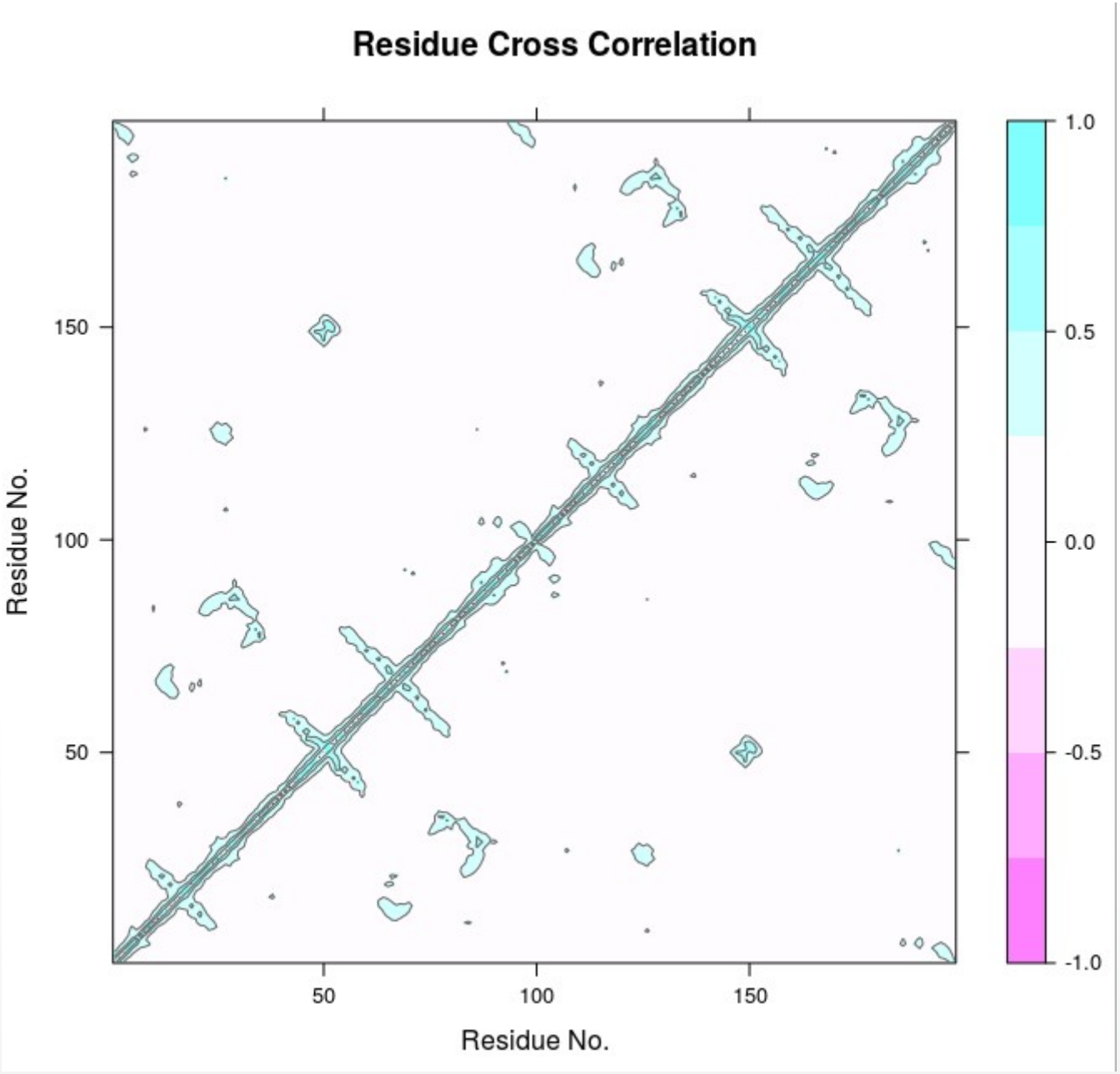
t=50

Sítio ativo exposto

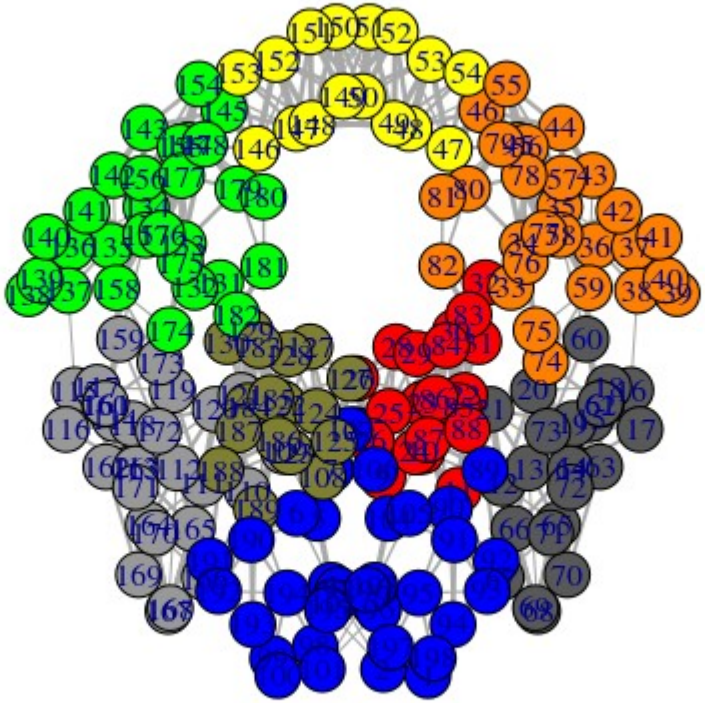


Tempo em ns (10^-9 segundos)

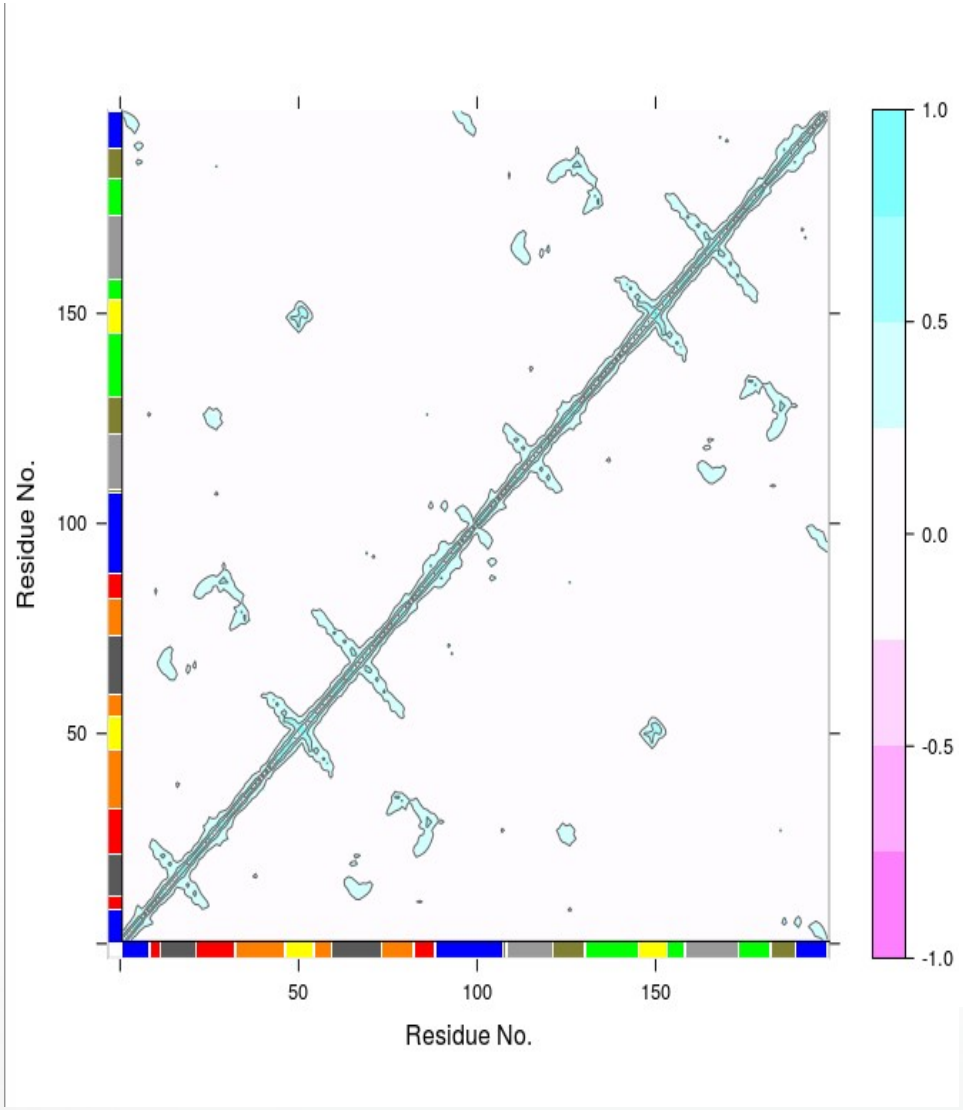
Correlação dinâmica

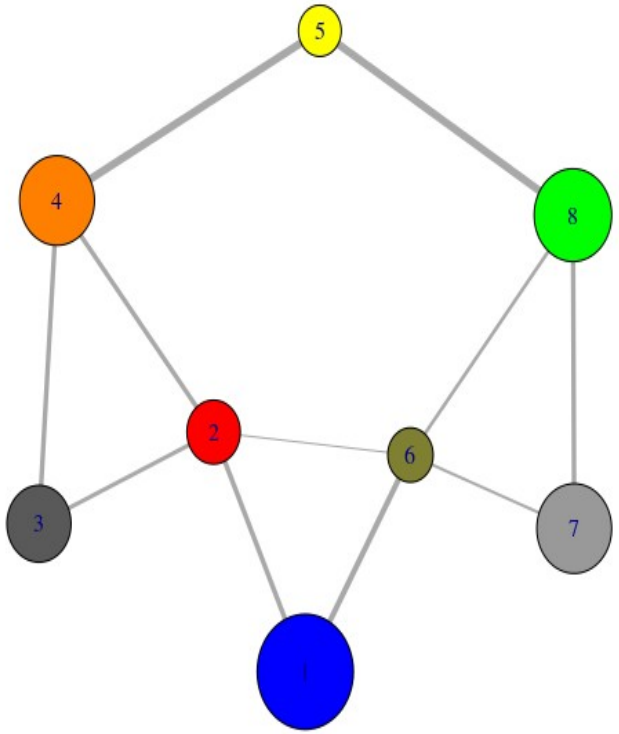


Rede de correlação dinâmica



Clusters


Residue No.
150
100
50
50
100
150
Residue No.
1.0
0.5
0.0
-0.5
-1.0


5
4
8
2
6
3
7
1

Como fazer isso no R?

- Bio3D é um pacote do R que contém funções para análise de estruturas, sequências e trajetórias de proteínas.
- http://thegrantlab.org/bio3d/

Fluxograma Bio3D

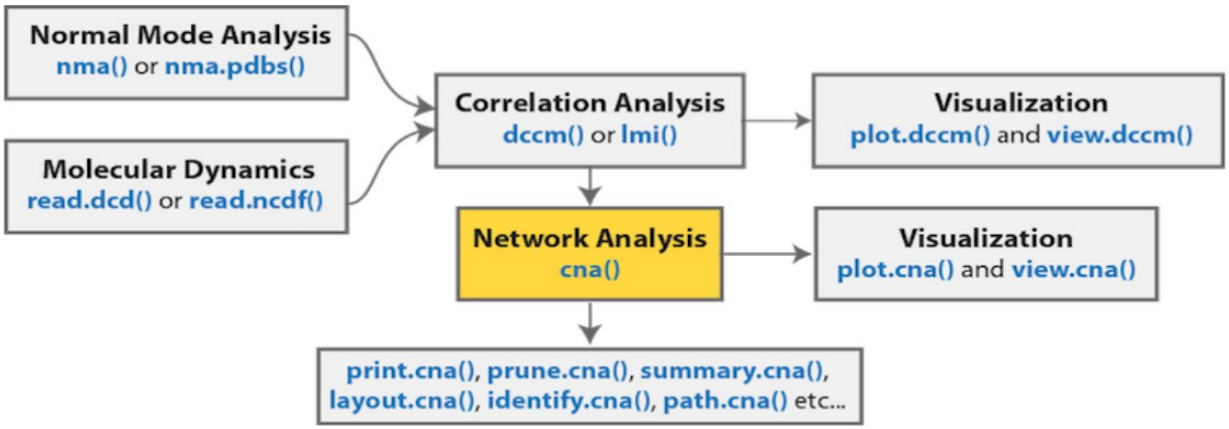
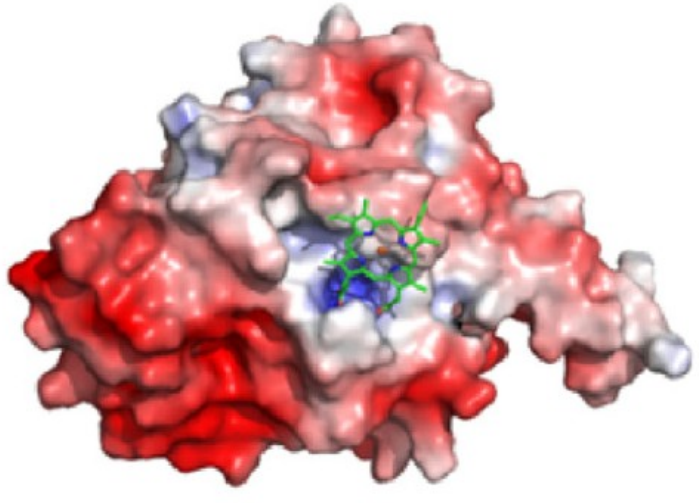


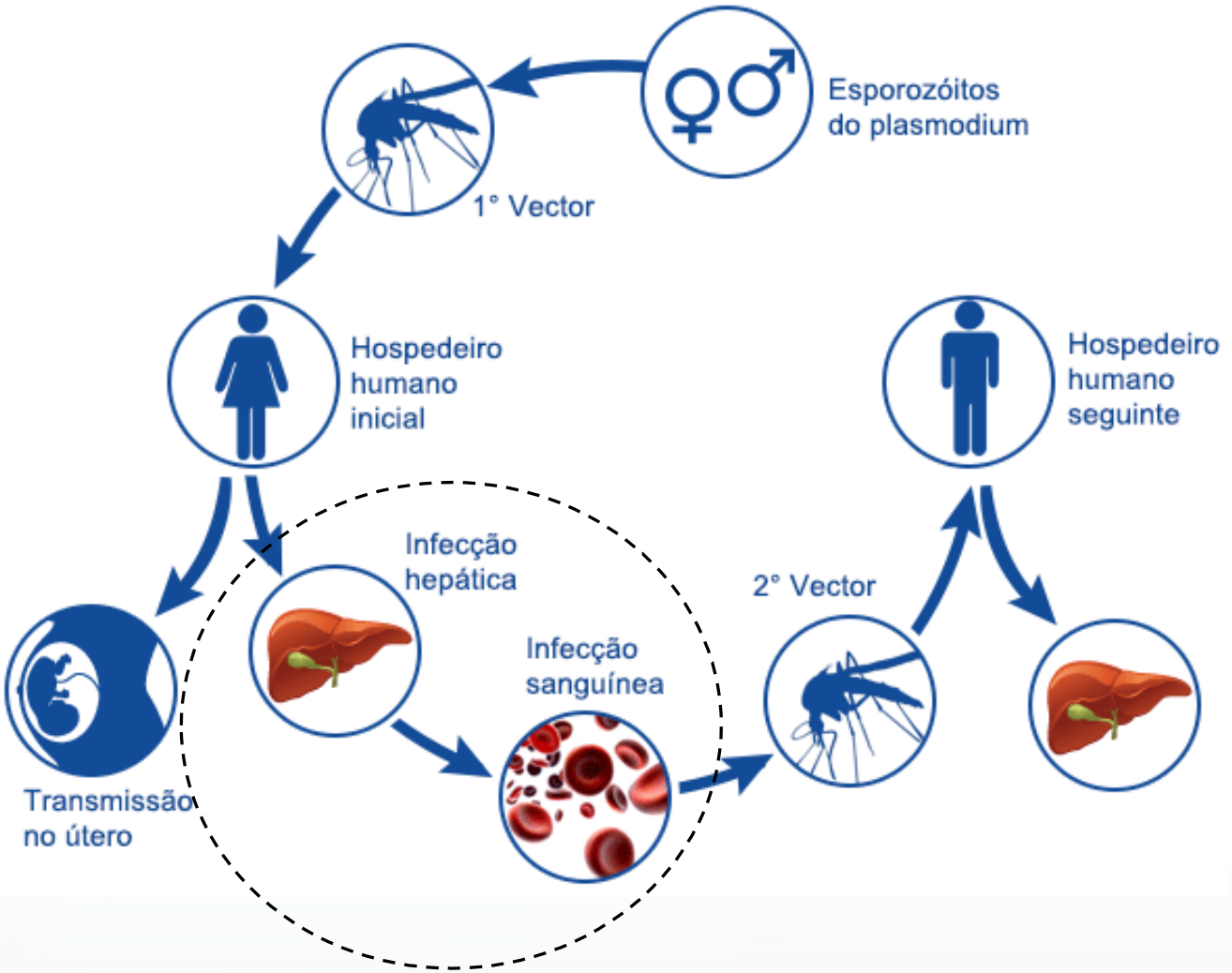
Figure 1: **Overview of a typical Bio3D network analysis protocol with key functions detailed in blue**. Normal mode and/or molecular dynamics input first undergoes **correlation analysis** (with the functions **dccm()** or **lmi()**). The output from these functions is typically a matrix of residue-by-residue cross-correlations. **Correlation network analysis** can then be performed (with the **cna()** function) to generate a correlation network with residues as nodes that are linked by weighted edges that are proportional to their degree of correlated motion. The **cna()** function also performs a **community clustering analysis** to identify *communities* of highly correlated residues. The full residue network and coarse-grained community network can be visualized and analyzed further. This methodology can provide valuable insight not readily available from the measures of correlation alone.

Exemplo: Malária

- Parasita: *Plamodium falciparum*
- Falcipaína-2:
  - Degradação da hemoglobina do hospedeiro
  - Promissor alvo molecular

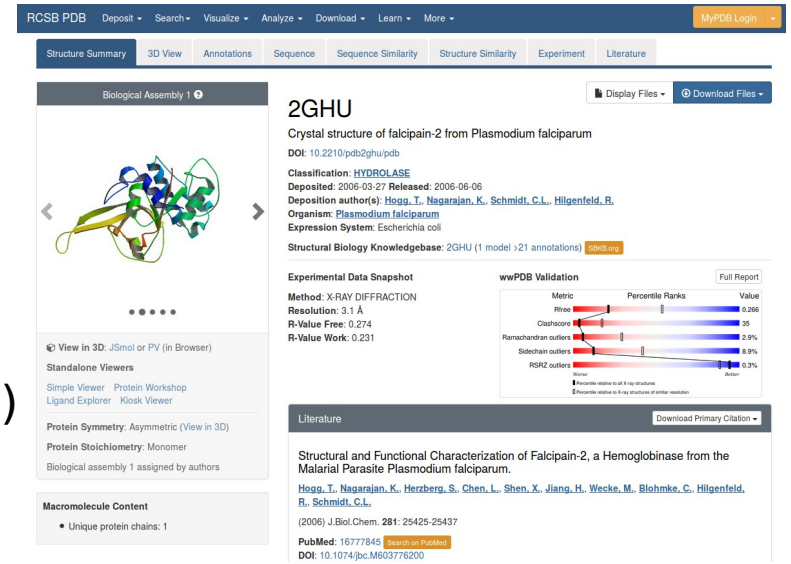


Ciclo de Transmissão da Malária

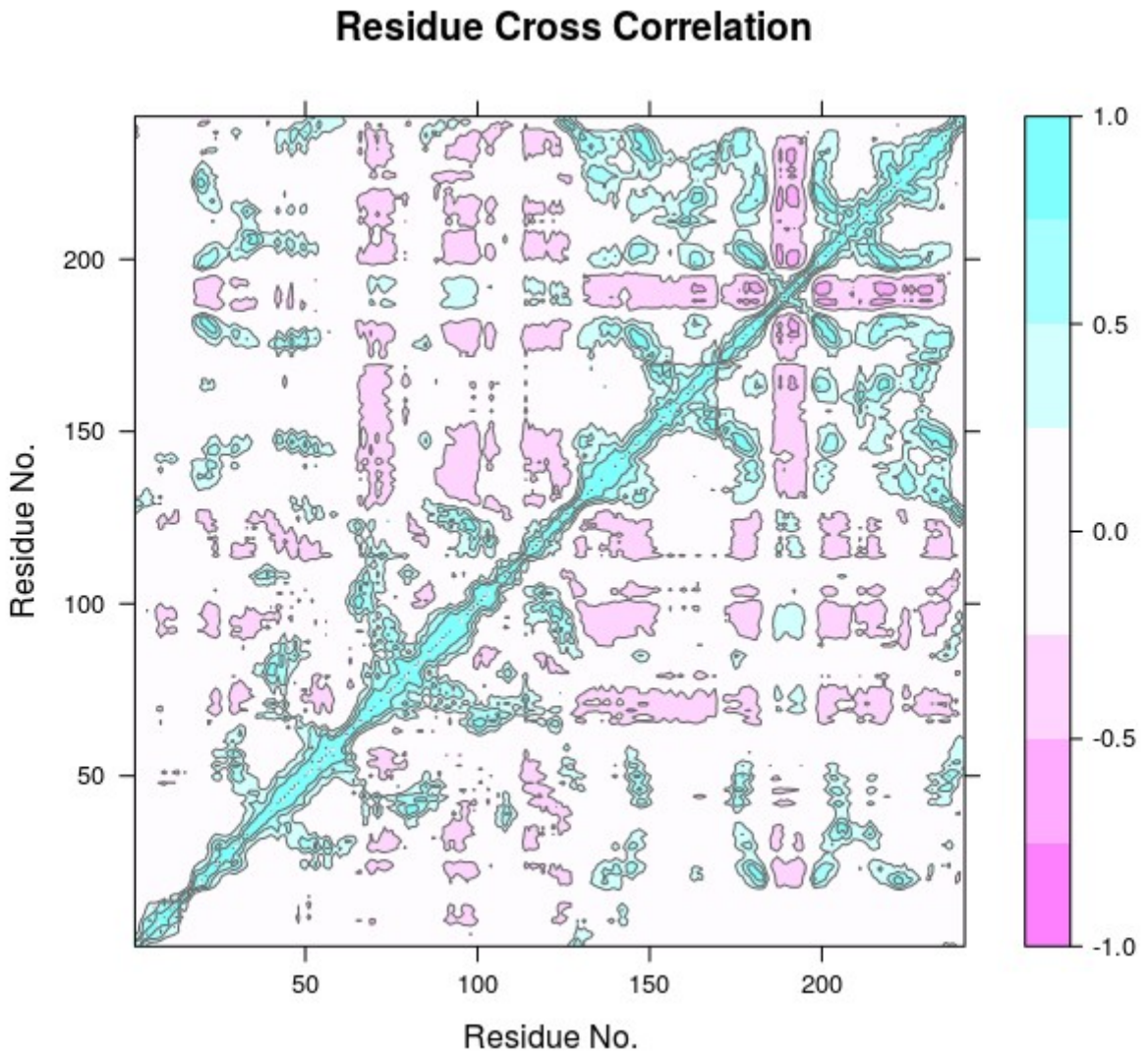


Exemplo: Códigos

```
# Lendo do pdb

require(bio3d)
falc2 <- read.pdb("2ghu")


# Dinâmica molecular via NMA

ind.A <- atom.select(falc2, chain = "A")
falc2.A <- trim.pdb(falc2, ind.A)
modes.falc2.A <- nma(falc2.A)


# Gerando a matriz de correlação

cor.mat <- dccm.nma(modes.falc2.A, nmodes = 50)
plot.dccm(cor.mat)
```

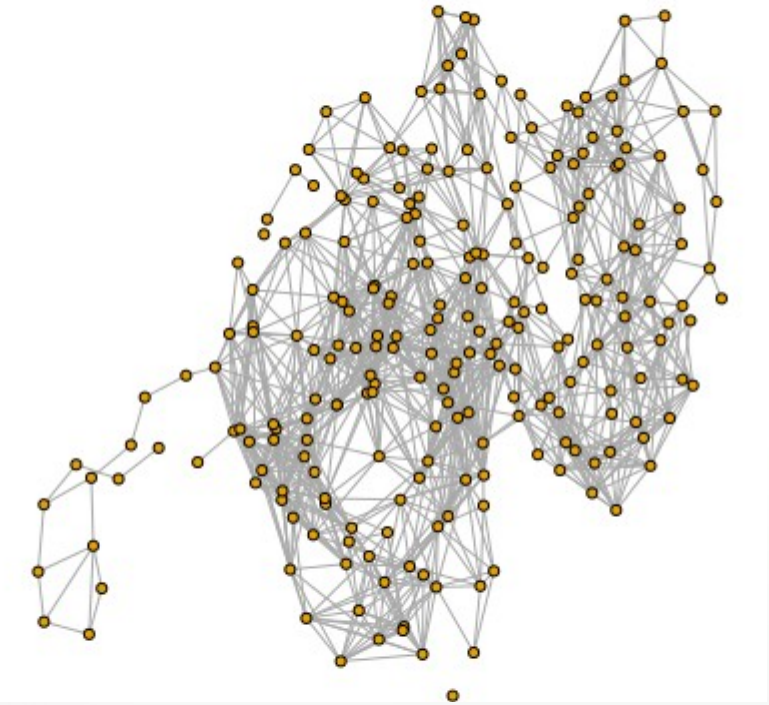


Matriz de correlação dinâmica

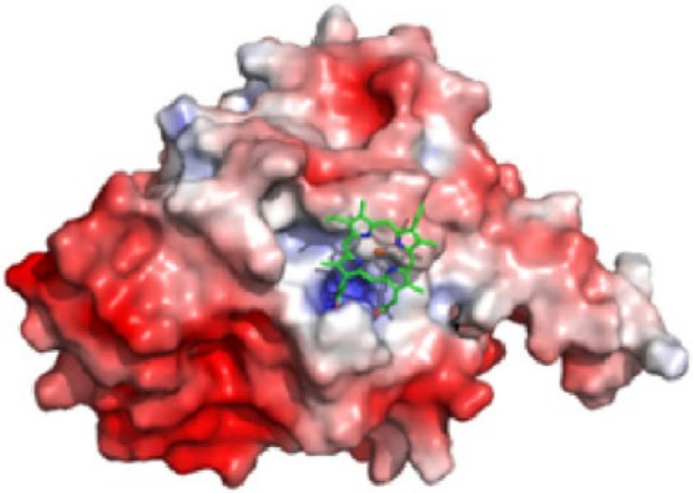


Rede de correlação dinâmica

```
# Criando a rede usando como ponto de corte: |cor_ij| > 0.65
net.falc2 <- cna(cor.mat, cutoff.cij = 0.65 )

# Plot da rede usando a estrutura cristalografica como referencia
plot(net.falc2, falc2.A, full = T, col=1, vertex.label=NA,
        vertex.size=3, weights=1)
```



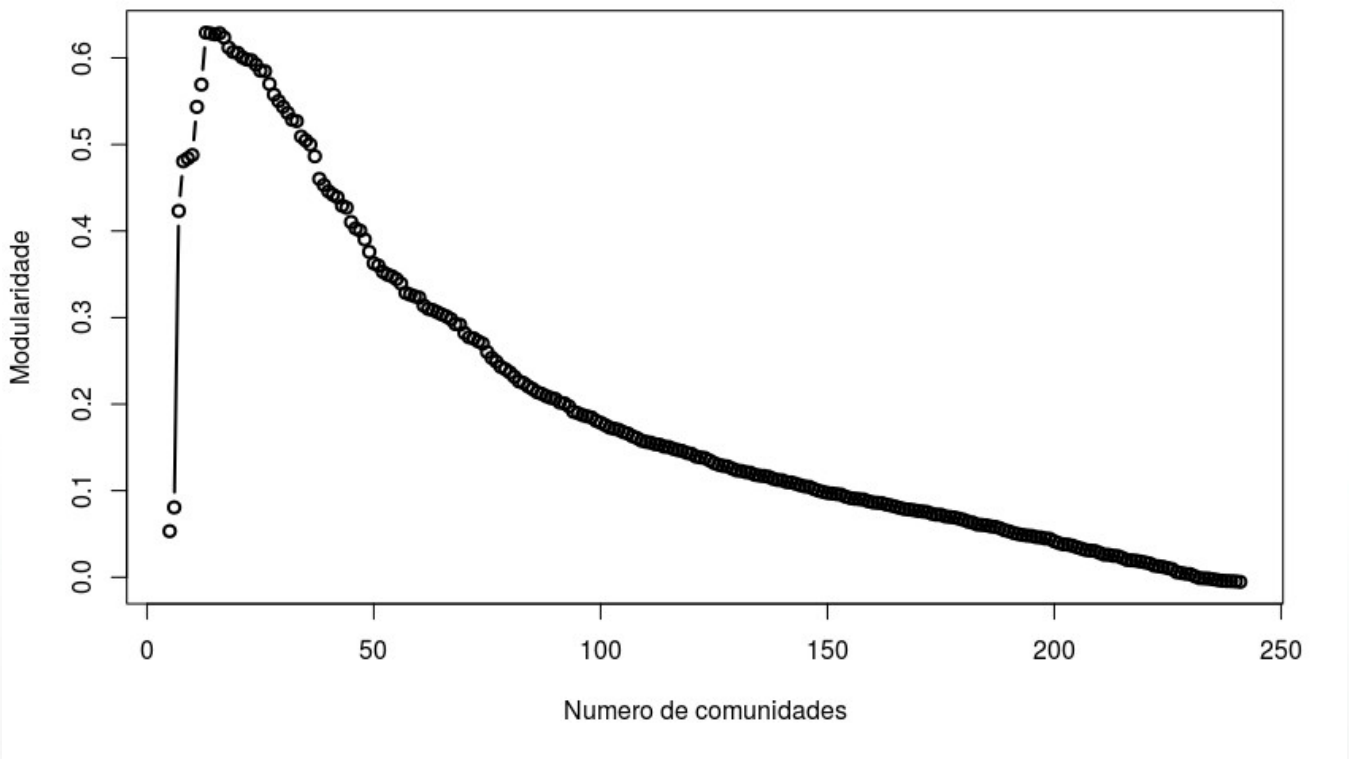
Rede de correlação dinâmica



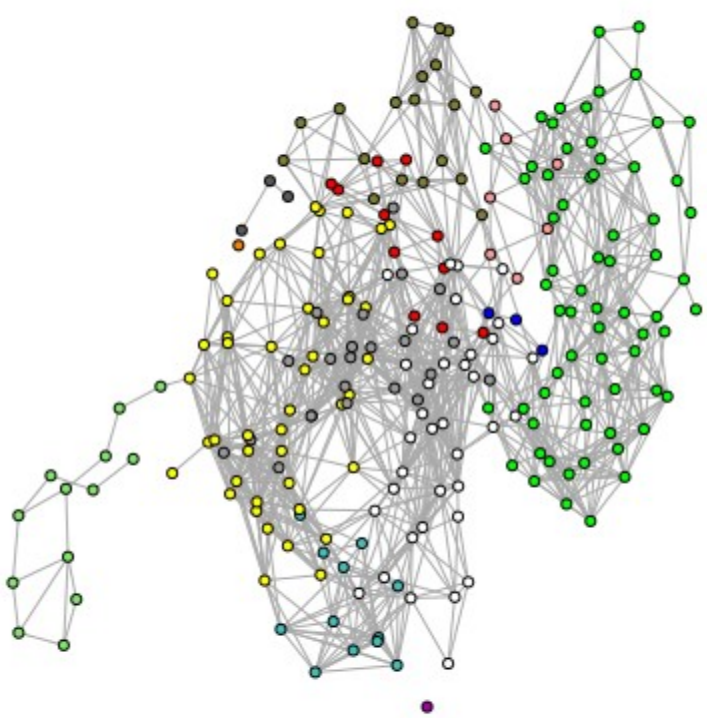
Estrutura 3D

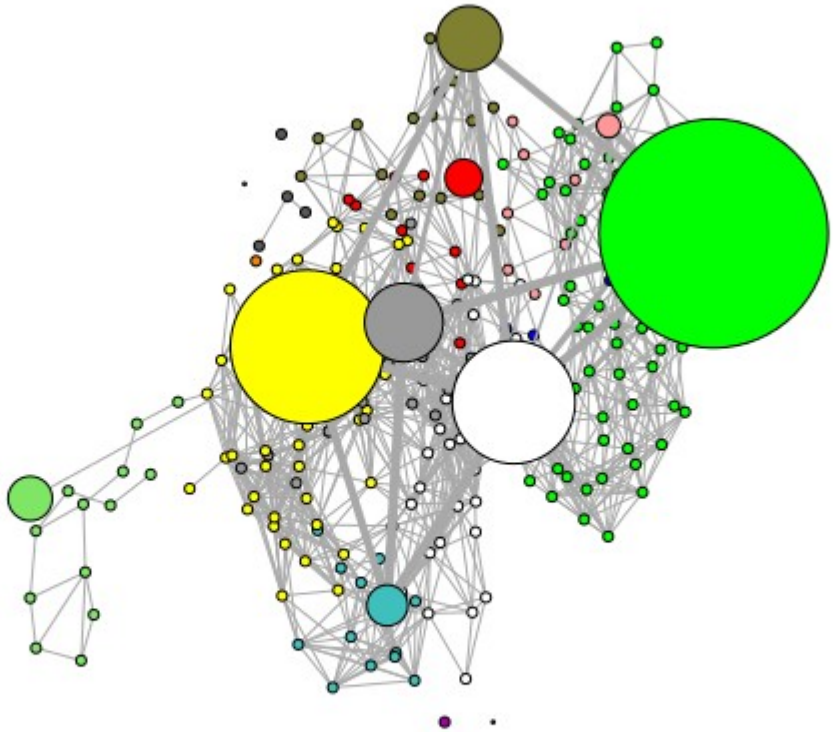
Clusters

```
# Modularidade
tree.falc2 <- community.tree(net.falc2, rescale=TRUE)

# Plot da modularidade
plot( tree.falc2$num.of.comms, tree.falc2$modularity )
```



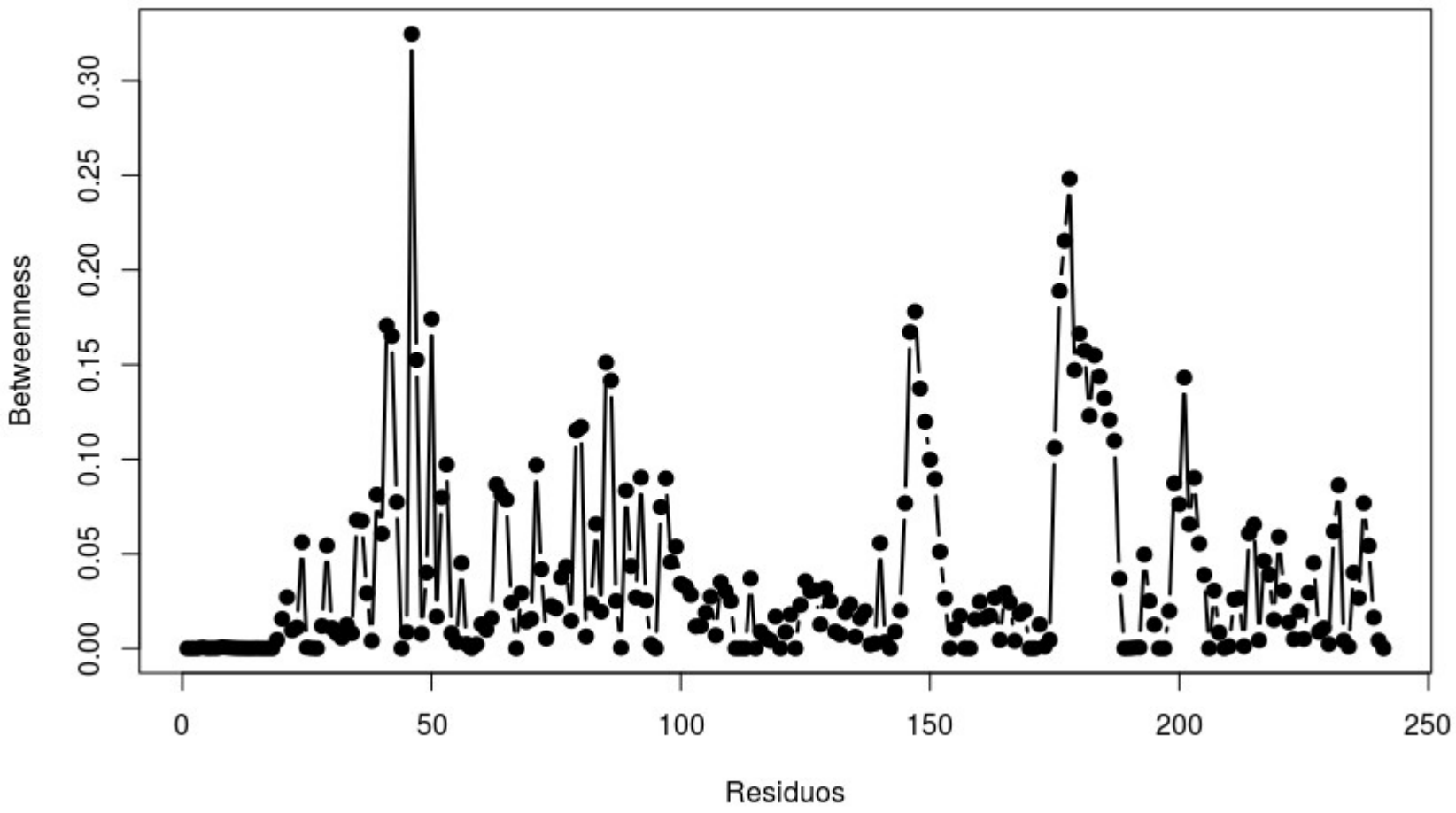
Clusters





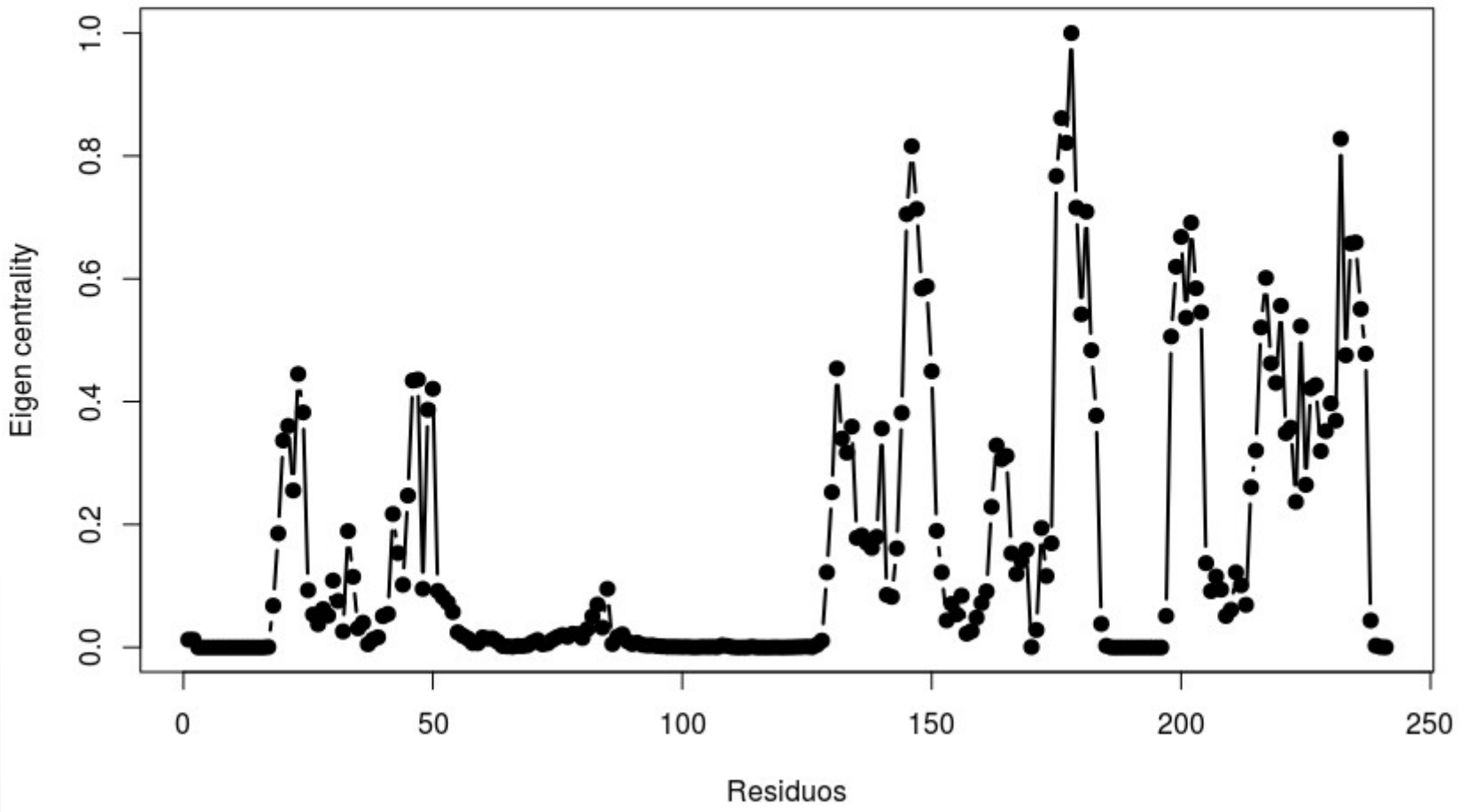
Medidas de centralidade da rede

```
##### betweenness ######
require(igraph)
bet.falc2.n <- normalize.vector(betweenness(net.falc2$network))
plot(bet.falc2.n)
```



Medidas de centralidade da rede

```
##### eigenvector ######
eigen.falc2 <-evcent(net.falc2$network)
plot(eigen.falc2$vector
```



Direções

- O R é uma ferramenta muito poderosa e bastante útil na biologia computacional.
- A análise estatística de redes têm grande potencial na modelagem molecular.
- Interessados: Pós graduação em Biologia Computacional e Sistemas da Fiocruz

Creditos: @alexandresalem