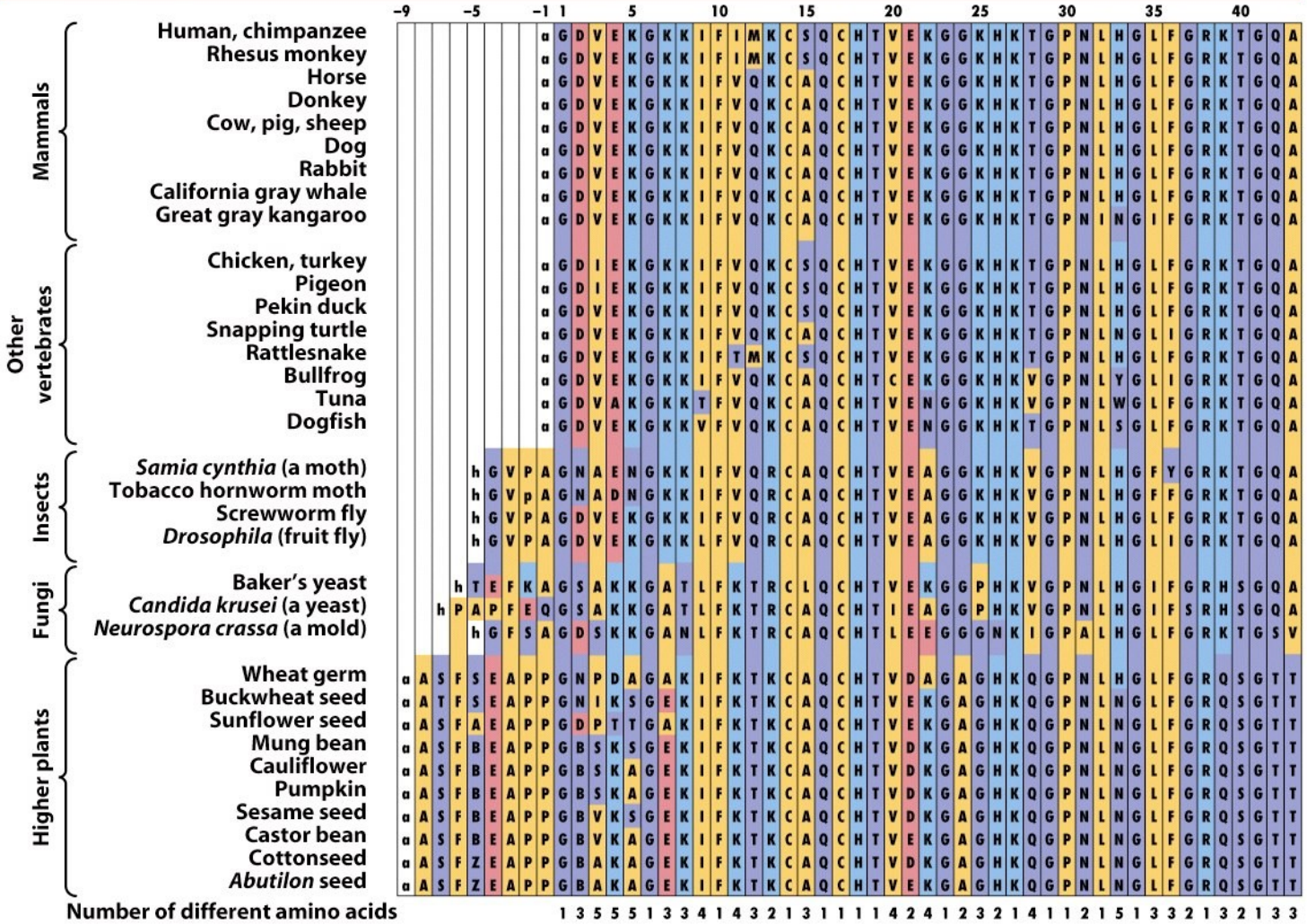
Predição da estrutura terciária/quaternária de proteínas com base na evolução de suas sequências de aminoácidos (estrutura primária)

Morcos et al., (2011). PNAS 108, E1293-E1301
Kamisetty et al., (2013). PNAS 110, 15674-15679
Jumper et al., (2021). Nature 596, 583-589

Alinhamento de estruturas primárias (sequências) de proteínas

TABLE 5-5 Amino Acid Sequences of Cytochrome *c* from 38 species[a]



[a]The amino acid side chains have been shaded according to their polarity characteristics so that an invariant or conservatively substituted residue is identified by a vertical band of a single color. The letter a at the beginning of the chain indicates that the N-terminal amino group is acetylated; an h indi-cates that the acetyl group is absent.

Source: After Dickerson, R.E., *Sci. Am.* 226(4); 58–72 (1972), with corrections from Dickerson, R.E., and Timkovich, R., *in* Boyer, P.D. (Ed.), *The Enzymes* (3rd ed.), Vol. 11, pp. 421–422, Academic Press (1975). Table copyrighted © by Irving Geis.

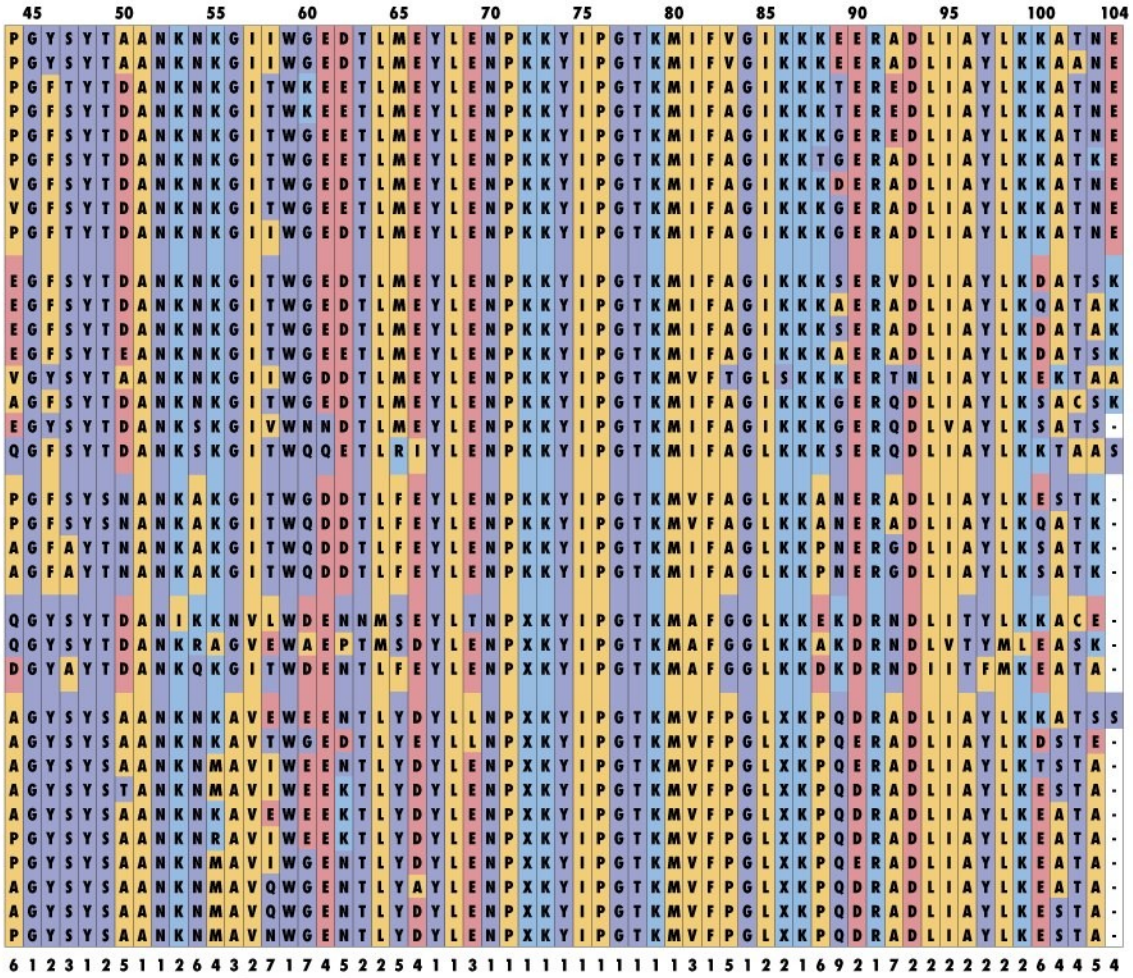


Table 5-5 part 2 Fundamentals of Biochemistry, 2/e

	Chimpanzee	Sheep	Rattlesnake	Carp	Snail	Moth	Yeast	Cauliflower
Human	0	10	14	18	29	31	44	44
Chimpanzee		10	14	18	29	31	44	44
Sheep			20	11	24	27	44	46
Rattlesnake				26	28	33	47	45
Carp					26	26	44	47
Garden snail						28	48	51
Tobacco hornworm moth							44	44
Baker's yeast (iso-1)								47
Cauliflower								

Lehninger Principles of Biochemistry
Nelson & Cox

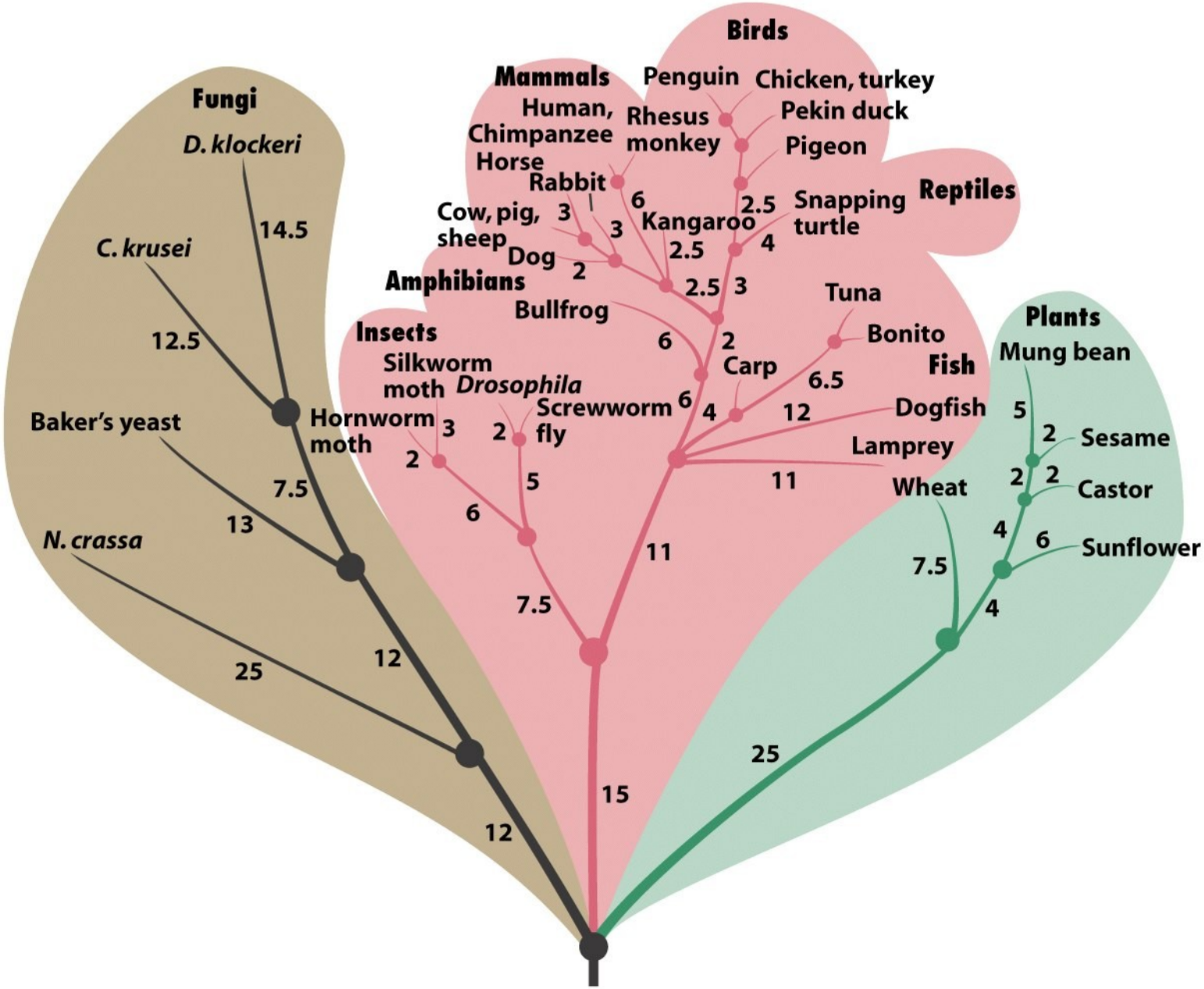


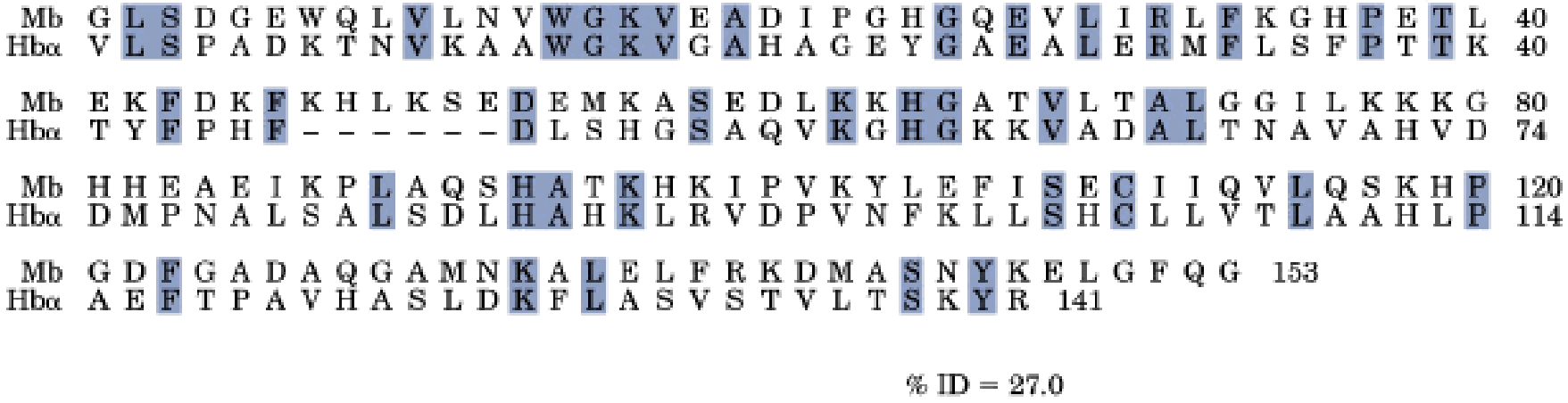
Figure 5-23 Fundamentals of Biochemistry, 2/e

Comparação das estruturas primárias de proteínas

-alinhamento

-cálculo de identidade e similaridade

- probabilidade de homologia (sequências proteicas compartilharem uma ancestral comum)



Lehninger Principles of Biochemistry
Nelson & Cox

Como quantificar o melhor alinhamento (probabilidade de homologia)?

Solução "empírica"

Comparar proteínas com alta identidade (> 85%) e calcular a frequência de trocas de $aac_i \rightarrow aac_j$

Indica a probabilidade da troca ser aceita

C Cys	12																			
S Ser	0	2																		
T Thr	−2	1	3																	
P Pro	−3	1	0	6																
A Ala	−2	1	1	1	2															
G Gly	−3	1	0	−1	1	5														
N Asn	−4	1	0	−1	0	0	2													
D Asp	−5	0	0	−1	0	1	2	4												
E Glu	−5	0	0	−1	0	0	1	3	4											
Q Gln	−5	−1	−1	0	0	−1	1	2	2	4										
H His	−3	−1	−1	0	−1	−2	2	1	1	3	6									
R Arg	−4	0	−1	0	−2	−3	0	−1	−1	1	2	6								
K Lys	−5	0	0	−1	−1	−2	1	0	0	1	0	3	5							
M Met	−5	−2	−1	−2	−1	−3	−2	−3	−2	−1	−2	0	0	6						
I Ile	−2	−1	0	−2	−1	−3	−2	−2	−2	−2	−2	−2	−2	2	5					
L Leu	−6	−3	−2	−3	−2	−4	−3	−4	−3	−2	−2	−3	−3	4	2	6				
V Val	−2	−1	0	−1	0	−1	−2	−2	−2	−2	−2	−2	−2	2	4	2	4			
F Phe	−4	−3	−3	−5	−4	−5	−4	−6	−5	−5	−2	−4	−5	0	1	2	−1	9		
Y Tyr	0	−3	−3	−5	−3	−5	−2	−4	−4	−4	0	−4	−4	−2	−1	−1	−2	7	10	
W Trp	−8	−2	−5	−6	−6	−7	−4	−7	−7	−5	−3	2	−3	−4	−5	−2	−6	0	0	17
	C Cys	S Ser	T Thr	P Pro	A Ala	G Gly	N Asn	D Asp	E Glu	Q Gln	H His	R Arg	K Lys	M Met	I Ile	L Leu	V Val	F Phe	Y Tyr	W Trp

The PAM-250 Log Odds Substitution Matrix

Probabilidade da troca j para i = $10^{X/10}$. Frequência do aac_i

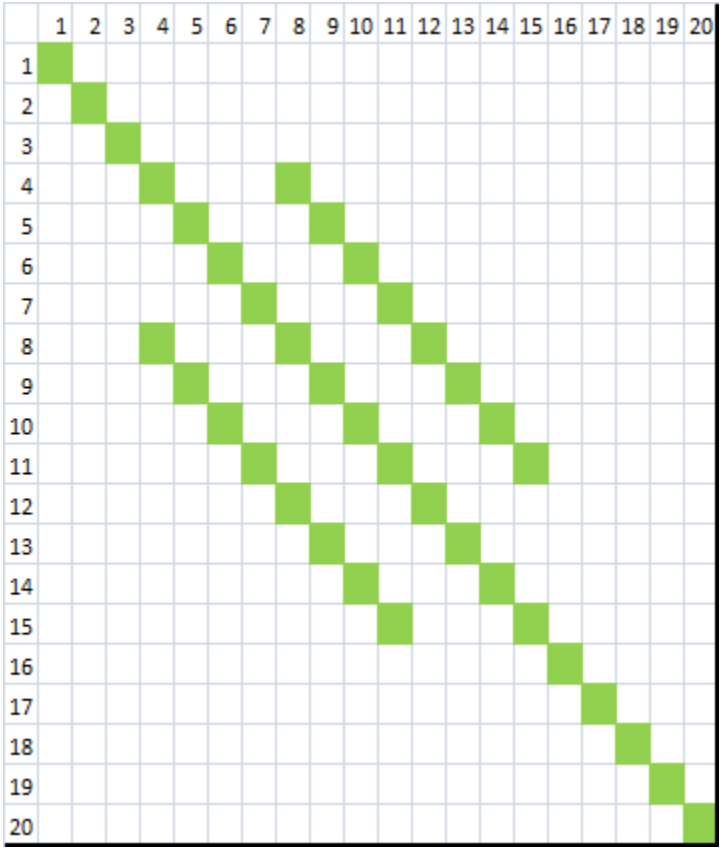
Matrizes de Probabilidade de Substituição de Aminoácidos

PAM

BLOSUM

Programas CLUSTAL e BLAST

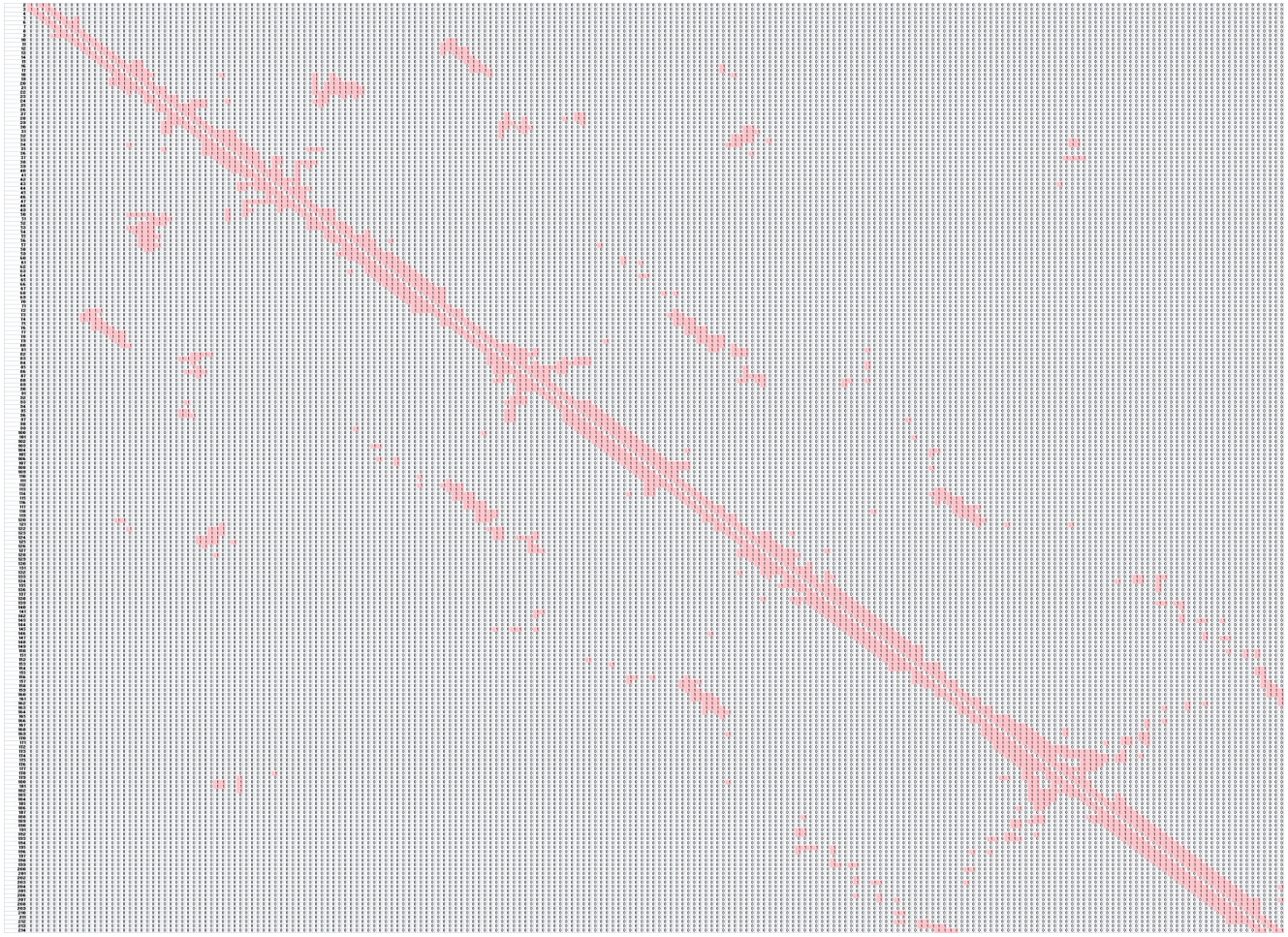
Mapa de Contatos da Estrutura Terciária de Proteínas



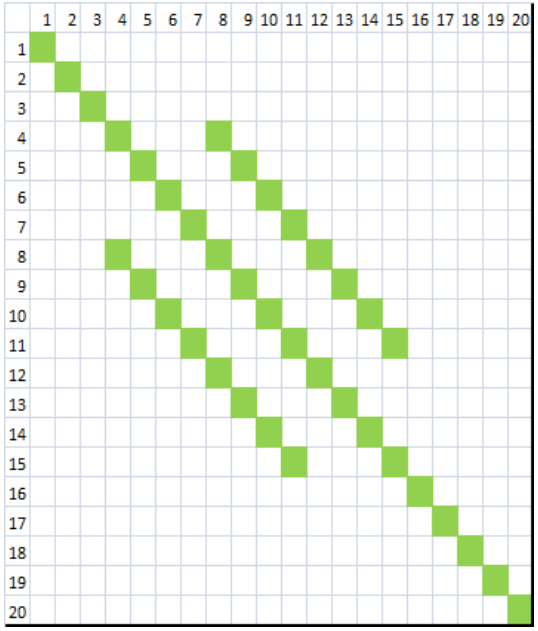
Distância entre Cα < 8 Å

Distância entre quaisquer átomos < 5 Å

Mapa de Contatos da Estrutura Terciária de Proteínas



Mapa de Contatos da Estrutura Terciária de Proteínas



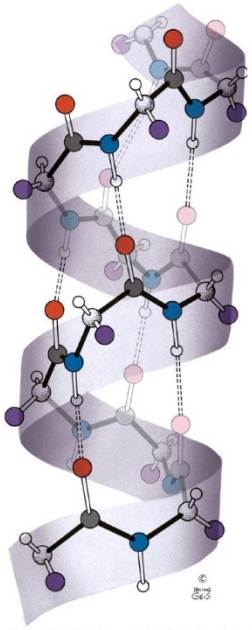


Figure 6-7 Fundamentals of Biochemistry, 2/e

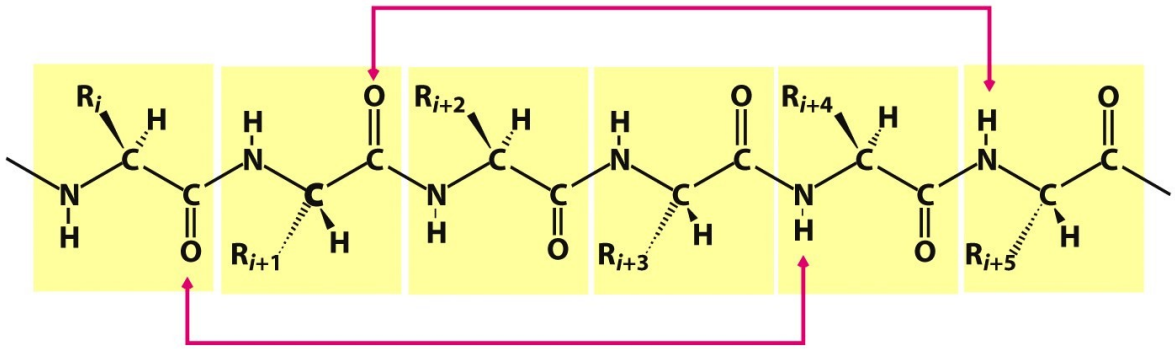
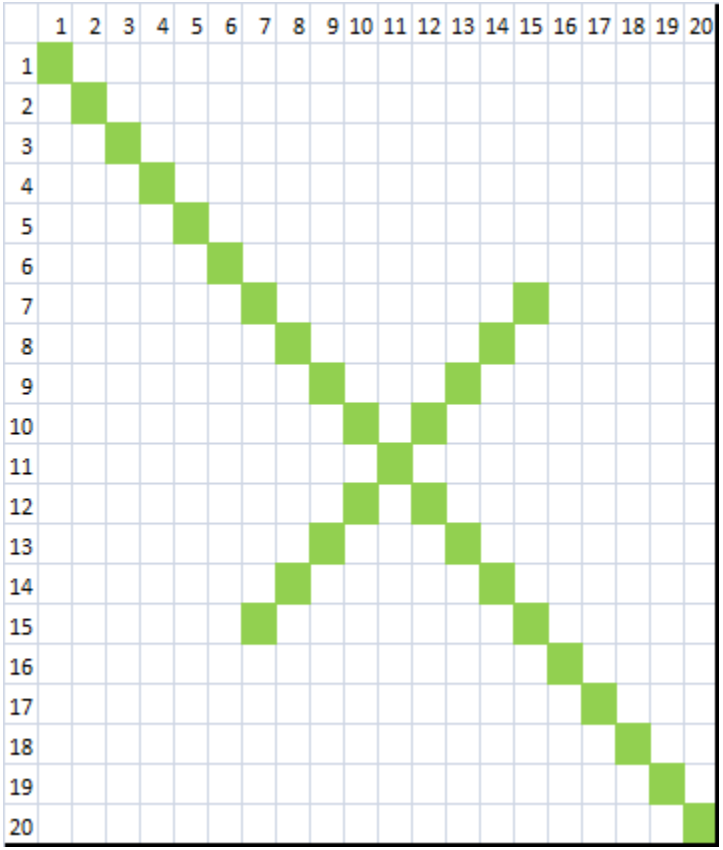
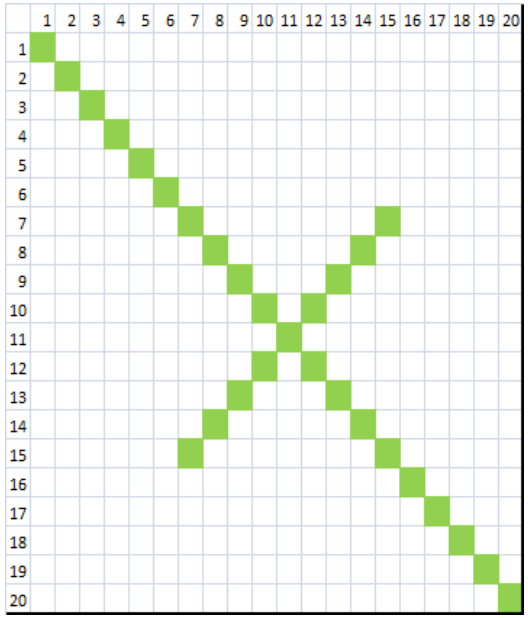


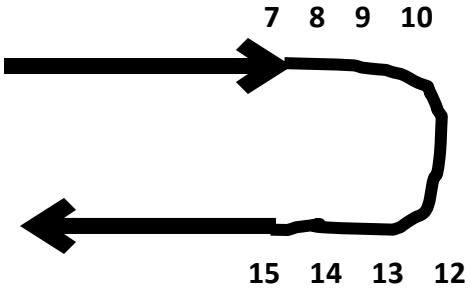
Figure 2.25
***Biochemistry,* Seventh Edition**

Mapa de Contatos da Estrutura Terciária de Proteínas

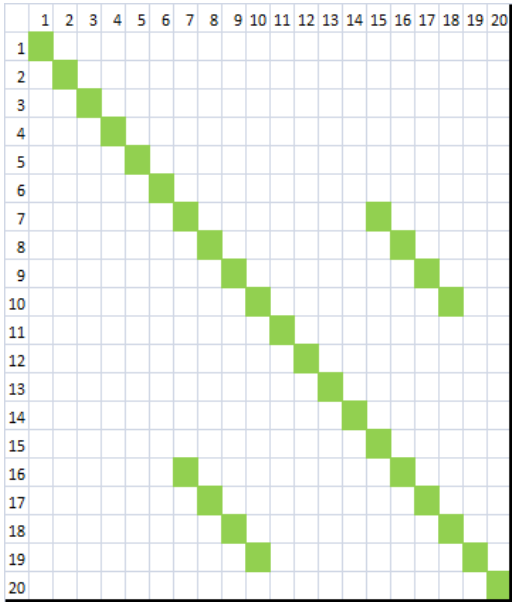


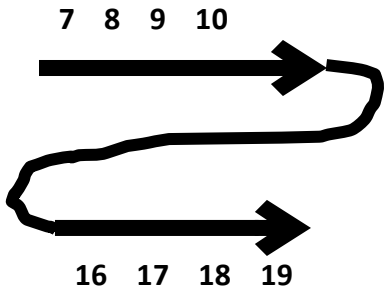
Mapa de Contatos da Estrutura Terciária de Proteínas

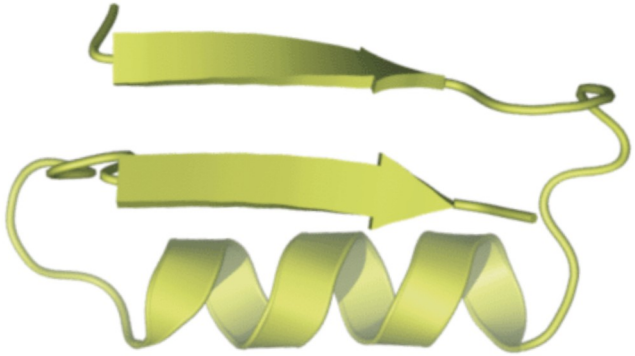


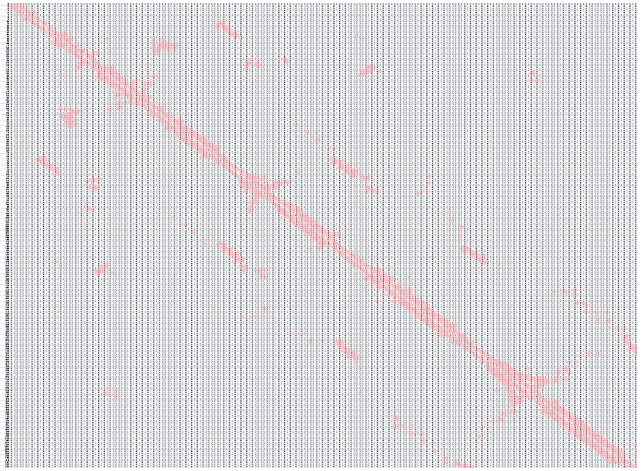


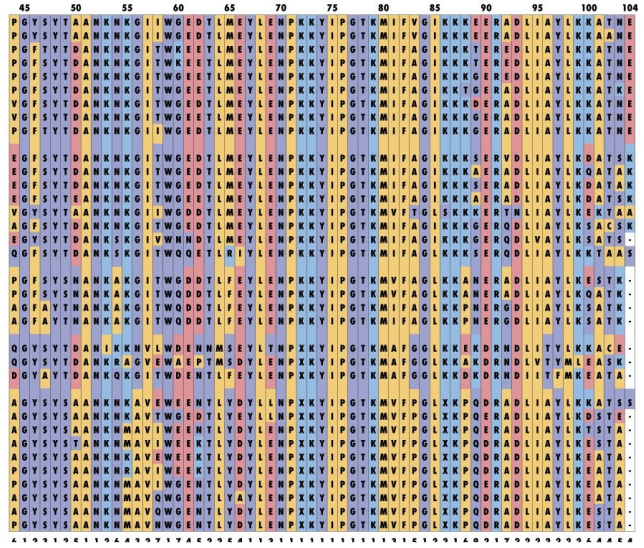
Mapa de Contatos da Estrutura Terciária de Proteínas

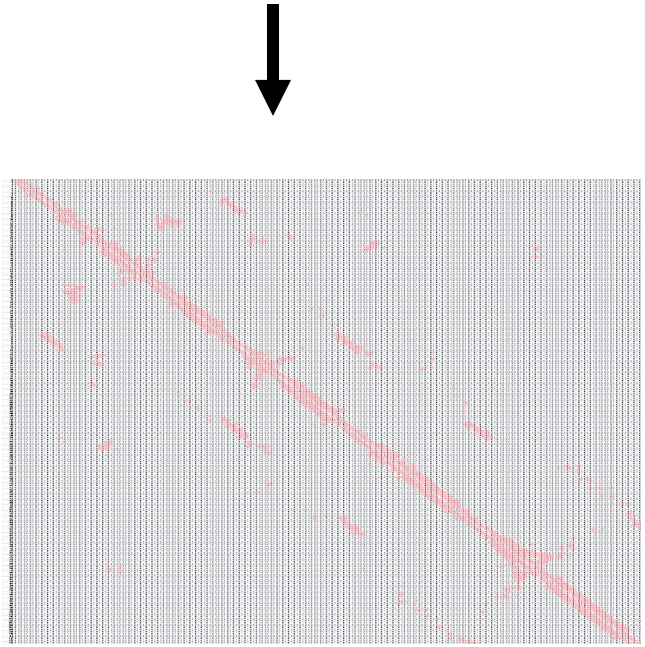




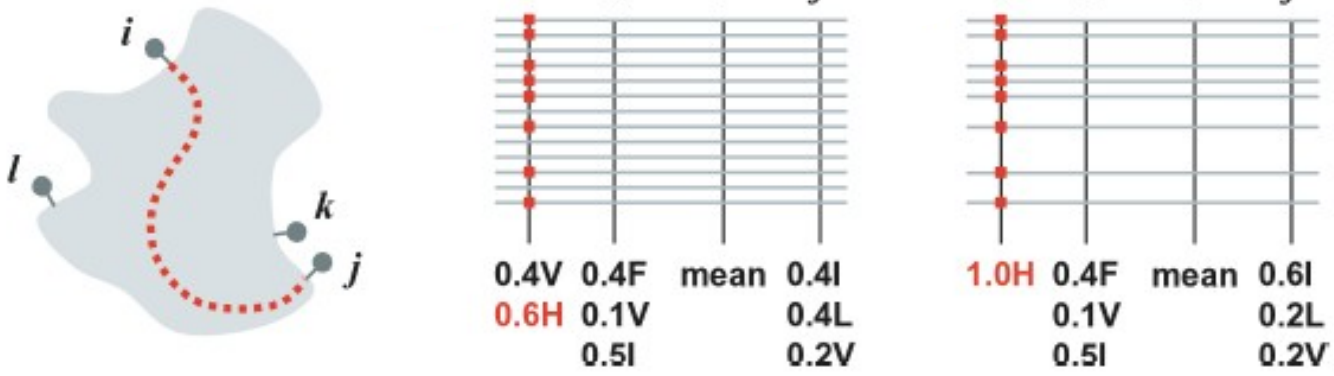






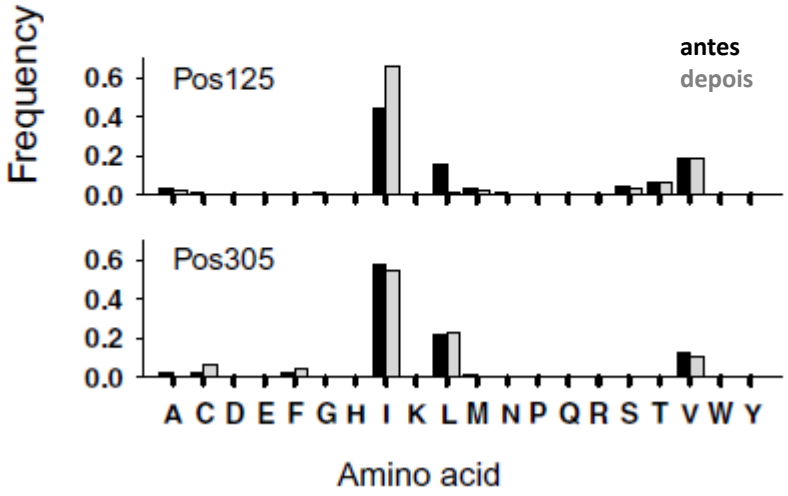


SCA – Statistical Coupling Analysis



SCA – Statistical Coupling Analysis
Análise aplicada a uma GPCR

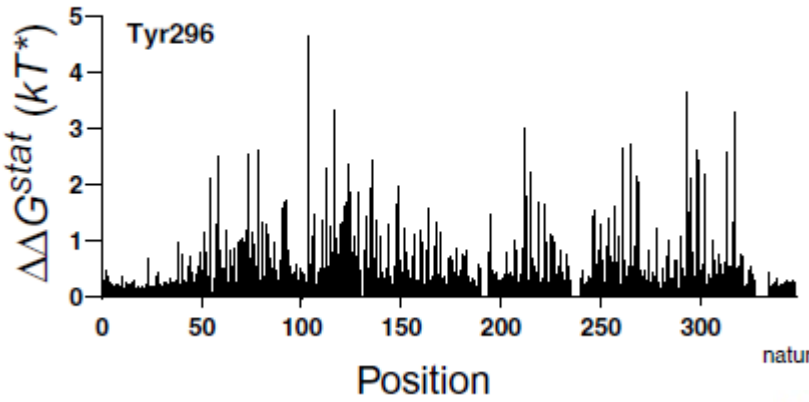
Perturbação na posição 296



antes
depois

$$\Delta G_i = \sqrt{\sum_x \left(\ln P_i^x\right)^2}$$ **antes**

$$\Delta\Delta G_{i,j} = \sqrt{\sum_x \left(\ln P_{i|\delta j}^x - \ln P_i^x\right)^2}$$ depois

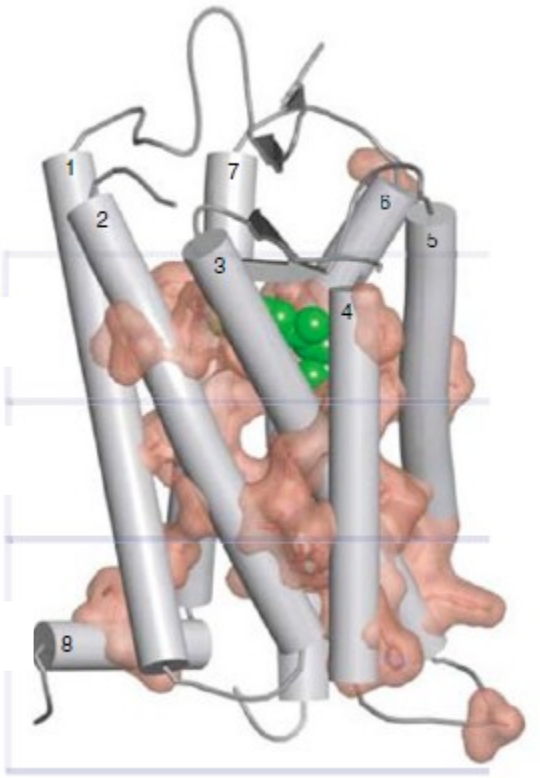


nature structural biology • volume 10 number 1 • january 2003

Vol. 20 no. 10 2004, pages 1565–1572
doi:10.1093/bioinformatics/bth128

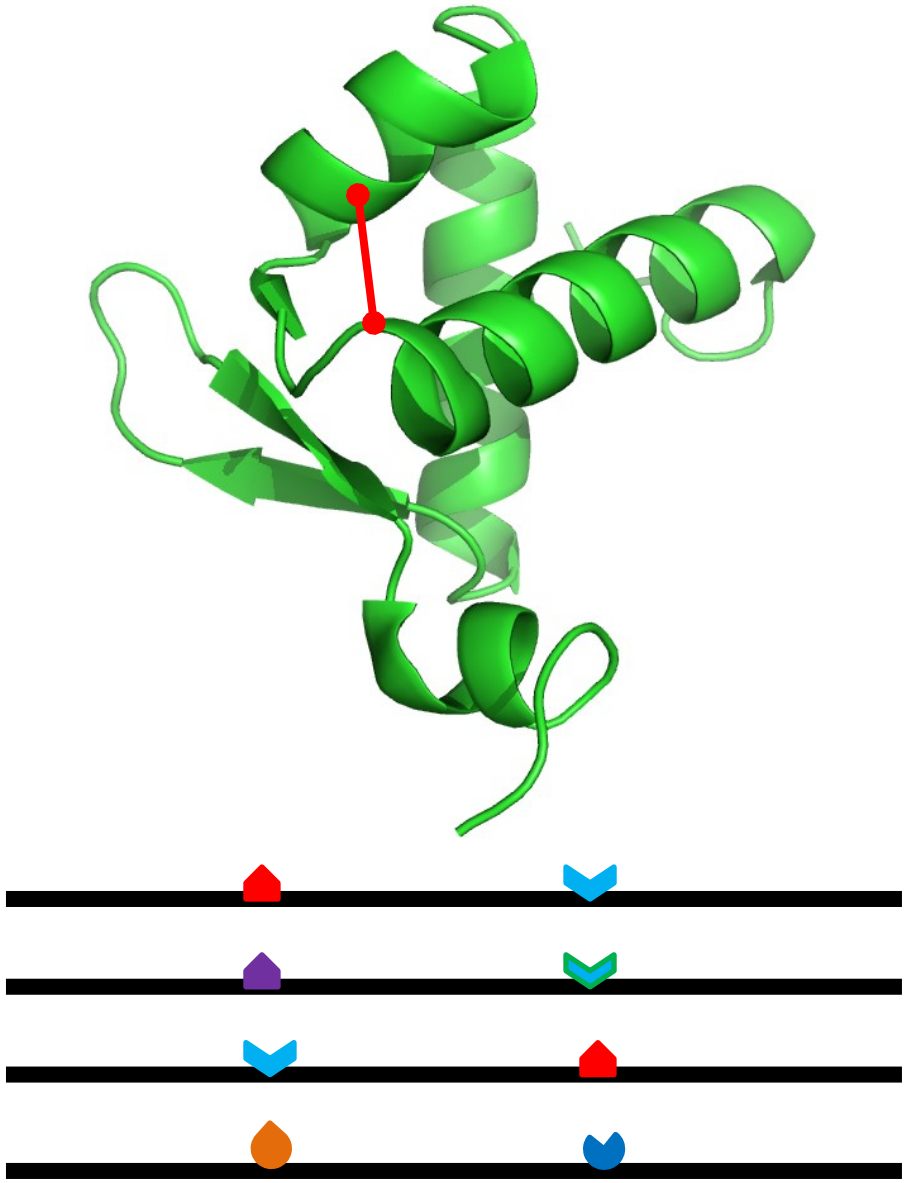
SCA – Statistical Coupling Analysis

Rotas de efeitos alostéricos

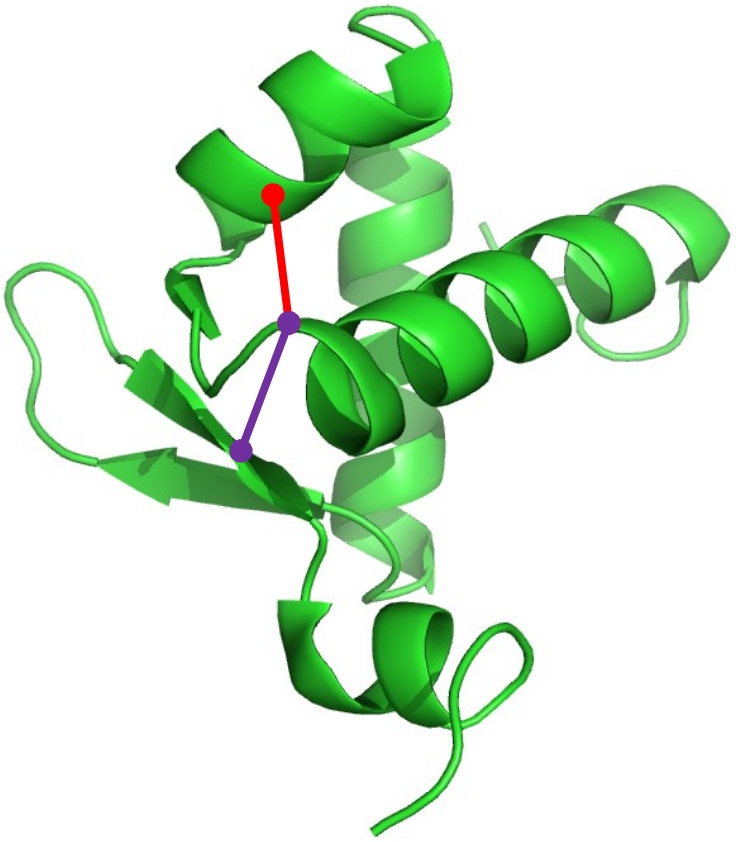


nature structural biology • volume 10 number 1 • january 2003

Covariação e Contatos: Predição de Estrutura



SCA – Statistical Coupling Analysis



SCA x DCA – Direct Coupling Analysis

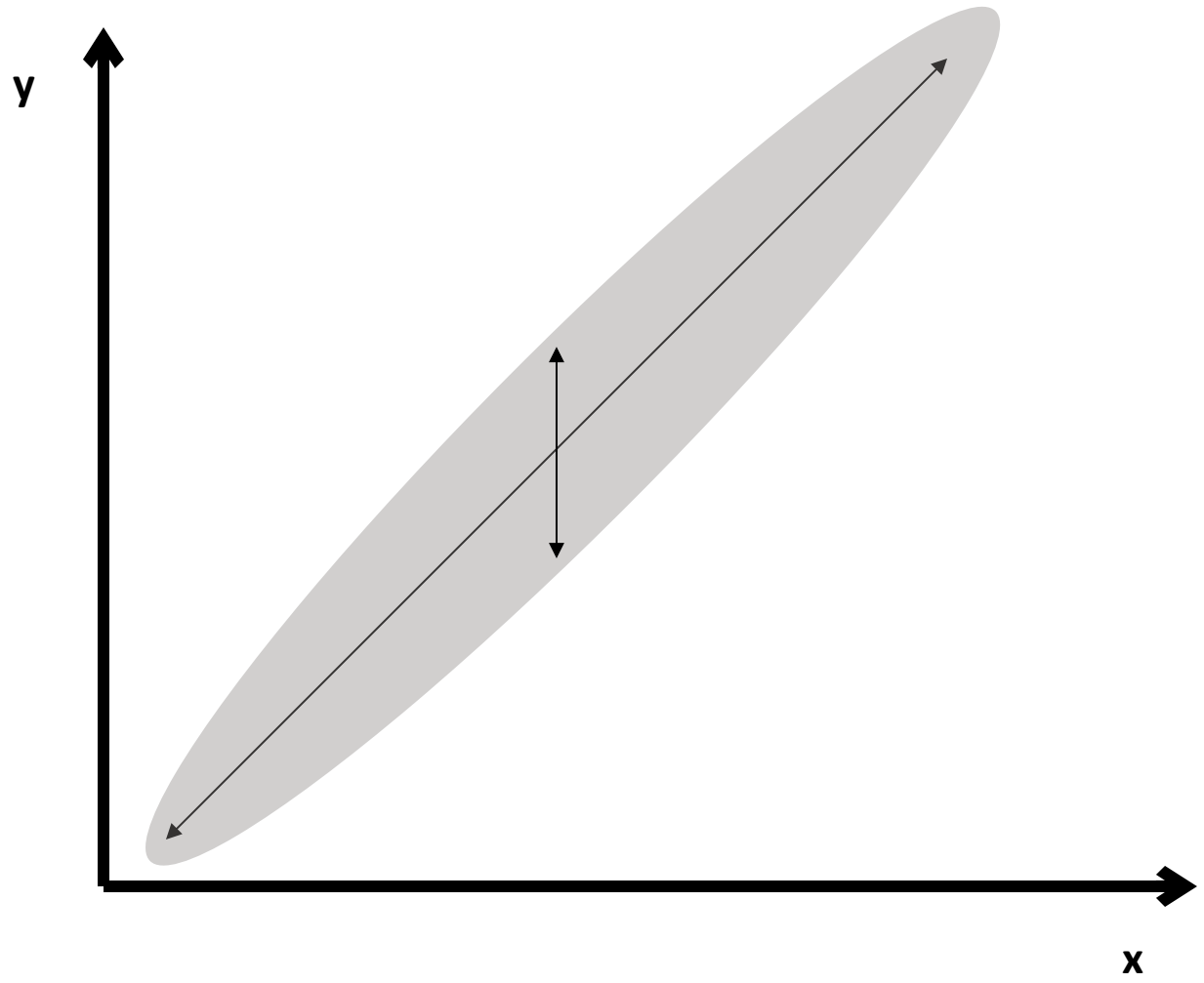
Utilização de alinhamentos de sequências com identidade menor que 80%

Comparação da f(A) na posição i e f(B) na posição j com a f(A,B) nas posições i,j

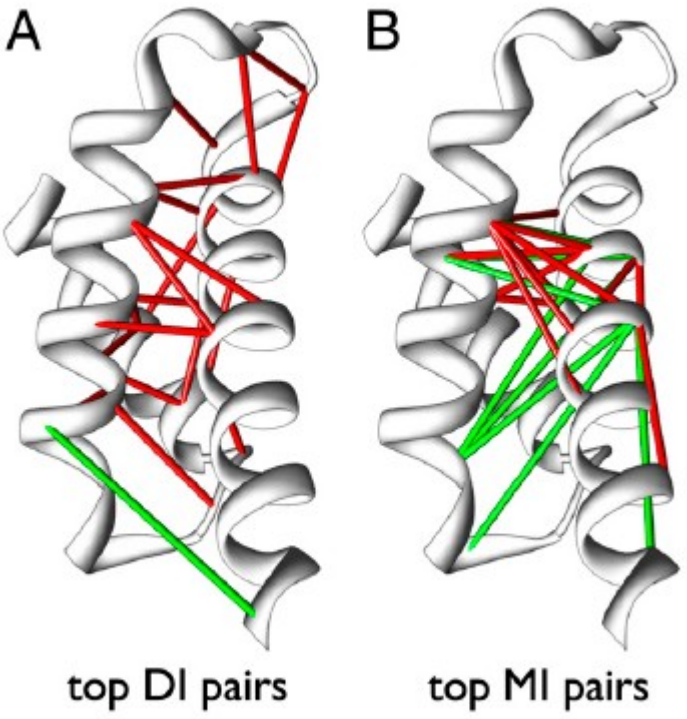
Construção de matrizes de covariação entre posições da sequência

Remoção das covariâncias secundárias (“contatos indiretos”)

SCA x DCA – Direct Coupling Analysis



DCA – Direct Coupling Analysis



DCA **SCA**

Cα < 8 Å
Cα > 8 Å

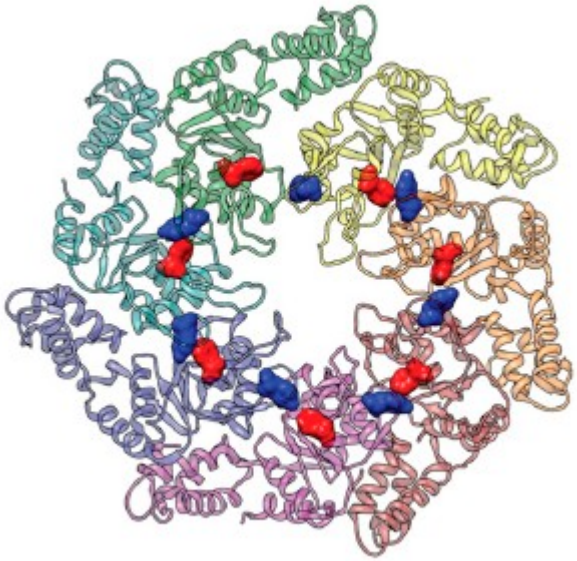
PDB 1OR7

PNAS | December 6, 2011 | vol. 108 | no. 49 | E1293–E1301

DCA – Direct Coupling Analysis

Contatos entre domínios na estrutura quaternária

Sítios de ligação de cofatores ou substratos



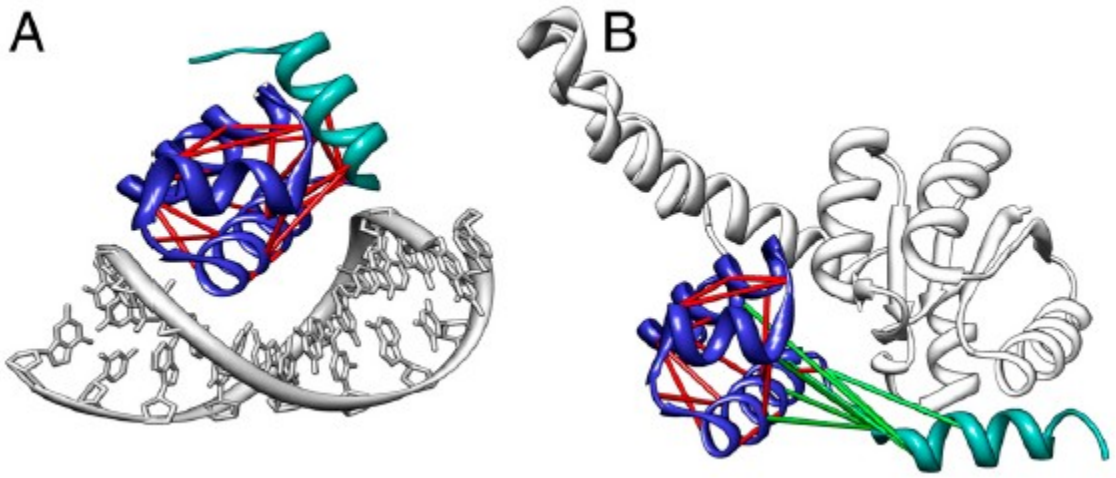
ATPase heptâmero

PDB 1NY6

PNAS | December 6, 2011 | vol. 108 | no. 49 | E1293–E1301

DCA – Direct Coupling Analysis

Conformações alternativas

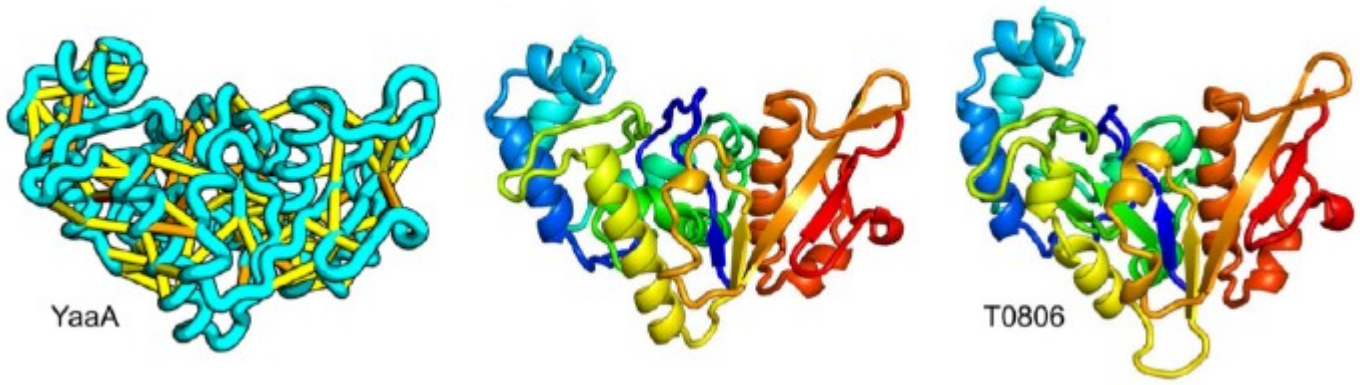


PDB 3C3W

GREMLIN + ROSETTA

Mudança no método de cálculo da covariância evolutiva
Significativo aumento no número de sequência disponíveis
Modelagem dos ângulos de torção da cadeia principal guiada pelo mapa de contatos previsto por covariância evolutiva + modelagem dos ângulos de torção das cadeias laterais

CASP 11
Critical Assessment of techniques for protein Structure Prediction (2014)



Pares de resíduo coevoluídos **Estrutura prevista** **Estrutura cristalográfica**

rmsd 2,9 Å

Ovchinnikov *et al*. eLife 2015;4:e09248. DOI: 10.7554/eLife.09248

AlphaFold

CASP 14 (2020)

rmsd 0,96 Å (cadeia principal)
rmsd 1,5 Å (todos os átomos)