

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

	-9	-5	-1	1	5	10	15	20	25	30	35	40																																										
Mammals	Human, chimpanzee				a	G	D	V	E	K	G	K	K	I	F	I	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Rhesus monkey				a	G	D	V	E	K	G	K	K	I	F	I	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Horse				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Donkey				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Cow, pig, sheep				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Dog				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Rabbit				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	California gray whale				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
Great gray kangaroo				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	I	N	G	I	F	G	R	K	T	G	Q	A							
Other vertebrates	Chicken, turkey				a	G	D	I	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Pigeon				a	G	D	I	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Pekin duck				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Snapping turtle				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	N	G	L	I	G	R	K	T	G	Q	A						
	Rattlesnake				a	G	D	V	E	K	G	K	K	I	F	T	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A						
	Bullfrog				a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	C	E	K	G	G	K	H	K	V	G	P	N	L	Y	G	L	I	G	R	K	T	G	Q	A						
	Tuna				a	G	D	V	A	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	N	G	G	K	H	K	V	G	P	N	L	W	L	F	G	R	K	T	G	Q	A							
	Dogfish				a	G	D	V	E	K	G	K	K	V	F	V	Q	K	C	A	Q	C	H	T	V	E	N	G	G	K	H	K	T	G	P	N	L	S	G	L	F	G	R	K	T	G	Q	A						
Insects	<i>Samia cynthia</i> (a moth)				h	G	V	P	A	G	N	A	E	N	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	F	Y	G	R	K	T	G	Q	A		
	Tobacco hornworm moth				h	G	V	P	A	G	N	A	D	N	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	F	F	G	R	K	T	G	Q	A		
	Screwworm fly				h	G	V	P	A	G	D	V	E	K	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A		
	<i>Drosophila</i> (fruit fly)				h	G	V	P	A	G	D	V	E	K	G	K	K	L	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	L	I	G	R	K	T	G	Q	A		
Fungi	Baker's yeast				h	T	E	F	K	A	G	S	A	K	K	G	A	T	L	F	K	T	R	C	L	Q	C	H	T	V	E	K	G	G	P	H	K	V	G	P	N	L	H	G	I	F	G	R	H	S	G	Q	A	
	<i>Candida krusei</i> (a yeast)				h	P	A	P	F	E	Q	G	S	A	K	K	G	A	T	L	F	K	T	R	C	A	Q	C	H	T	I	E	A	G	G	P	H	K	V	G	P	N	L	H	G	I	F	S	R	H	S	G	Q	A
	<i>Neurospora crassa</i> (a mold)				h	G	F	S	A	G	D	S	K	K	G	A	N	L	F	K	T	R	C	A	Q	C	H	T	L	E	E	G	G	G	N	K	I	G	P	A	L	H	G	L	F	G	R	K	T	G	S	V		
Higher plants	Wheat germ	a	A	S	F	S	E	A	P	P	G	N	P	D	A	G	A	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	A	G	A	G	H	K	Q	G	P	N	L	H	G	L	F	G	R	Q	S	G	T		
	Buckwheat seed	a	A	T	F	S	E	A	P	P	G	N	I	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T		
	Sunflower seed	a	A	S	F	A	E	A	P	P	G	D	P	T	T	G	A	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T		
	Mung bean	a	A	S	F	B	E	A	P	P	G	B	S	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T		
	Cauliflower	a	A	S	F	B	E	A	P	P	G	B	S	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T		
	Pumpkin	a	A	S	F	B	E	A	P	P	G	B	S	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T		
	Sesame seed	a	A	S	F	B	E	A	P	P	G	B	V	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T		
	Castor bean	a	A	S	F	B	E	A	P	P	G	B	V	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T		
	Cottonseed	a	A	S	F	Z	E	A	P	P	G	B	A	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T		
	<i>Abutilon</i> seed	a	A	S	F	Z	E	A	P	P	G	B	A	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T		

Number of different amino acids 1 3 5 5 5 1 3 3 4 1 4 3 2 1 3 1 1 1 1 4 2 4 1 2 3 2 1 4 1 1 2 1 5 1 3 3 2 1 3 2 1 3 3

^aThe 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 indicates 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.

45	50	55	60	65	70	75	80	85	90	95	100	104																																																
P	G	Y	S	T	A	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	V	G	I	K	K	K	E	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E	
P	G	Y	S	T	A	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	V	G	I	K	K	K	E	E	R	A	D	L	I	A	Y	L	K	K	A	A	N	E	
P	G	F	T	Y	T	D	A	N	K	N	K	G	I	T	W	K	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	T	E	R	E	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	K	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	T	E	R	E	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	T	G	E	R	A	D	L	I	A	Y	L	K	K	A	T	K	E
V	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	D	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
V	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	T	Y	T	D	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
E	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	S	E	R	V	D	L	I	A	Y	L	K	D	A	T	S	K
E	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	A	E	R	A	D	L	I	A	Y	L	K	Q	A	T	A	K
E	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	S	E	R	A	D	L	I	A	Y	L	K	D	A	T	A	K
E	G	F	S	Y	T	E	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	A	E	R	A	D	L	I	A	Y	L	K	D	A	T	S	K
V	G	Y	S	T	A	A	N	K	N	K	G	I	I	W	G	D	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	V	F	T	G	L	S	K	K	K	E	R	T	N	L	I	A	Y	L	K	E	A	A			
A	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	Q	D	L	I	A	Y	L	K	S	A	C	S	K
E	G	Y	S	Y	T	D	A	N	K	S	K	G	I	V	W	N	N	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	Q	D	L	V	A	Y	L	K	S	A	T	S	-
Q	G	F	S	Y	T	D	A	N	K	S	K	G	I	T	W	Q	E	T	L	R	I	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	L	K	K	K	S	E	R	Q	D	L	I	A	Y	L	K	K	T	A	A	S	
P	G	F	S	Y	S	N	A	N	K	A	K	G	I	T	W	G	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	V	F	A	G	L	K	K	A	N	E	R	A	D	L	I	A	Y	L	K	E	S	T	K	-
P	G	F	S	Y	S	N	A	N	K	A	K	G	I	T	W	Q	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	V	F	A	G	L	K	K	A	N	E	R	A	D	L	I	A	Y	L	K	Q	A	T	K	-
A	G	F	A	Y	T	N	A	N	K	A	K	G	I	T	W	Q	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	L	K	K	P	N	E	R	G	D	L	I	A	Y	L	K	S	A	T	K	-
A	G	F	A	Y	T	N	A	N	K	A	K	G	I	T	W	Q	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	L	K	K	P	N	E	R	G	D	L	I	A	Y	L	K	S	A	T	K	-
Q	G	Y	S	Y	T	D	A	N	I	K	N	V	L	W	D	E	N	N	M	S	E	Y	L	T	N	P	X	K	Y	I	P	G	T	K	M	A	F	G	G	L	K	K	E	K	D	R	N	D	L	I	T	Y	L	K	K	A	C	E	-	
Q	G	Y	S	Y	T	D	A	N	K	R	A	G	V	E	W	A	E	P	T	M	S	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	A	F	G	G	L	K	K	A	K	D	R	N	D	L	V	T	Y	M	L	E	A	S	K	-
D	G	Y	A	Y	T	D	A	N	K	Q	K	G	I	T	W	D	E	N	T	L	F	E	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	A	F	G	G	L	K	K	D	R	N	D	I	I	T	F	M	K	E	A	T	A	-		
A	G	Y	S	Y	S	A	A	N	K	N	K	A	V	E	W	E	E	N	T	L	Y	D	Y	L	L	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	K	A	T	S	S
A	G	Y	S	Y	S	A	A	N	K	N	K	A	V	T	W	G	E	D	T	L	Y	E	Y	L	L	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	E	R	A	D	L	I	A	Y	L	K	D	S	T	E	-
A	G	Y	S	Y	S	A	A	N	K	N	M	A	V	I	W	E	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	E	R	A	D	L	I	A	Y	L	K	T	S	T	A	-
A	G	Y	S	Y	S	T	A	N	K	N	M	A	V	I	W	E	E	K	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	S	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	K	A	V	E	W	E	E	K	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
P	G	Y	S	Y	S	A	A	N	K	N	R	A	V	I	W	E	E	K	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
P	G	Y	S	Y	S	A	A	N	K	N	M	A	V	I	W	G	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	E	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	M	A	V	Q	W	G	E	N	T	L	Y	A	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	M	A	V	Q	W	G	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	S	T	A	-
P	G	Y	S	Y	S	A	A	N	K	N	M	A	V	N	W	G	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	S	T	A	-

6 1 2 3 1 2 5 1 1 2 6 4 3 2 7 1 7 4 5 2 2 5 4 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 3 1 5 1 2 2 1 6 9 2 1 7 2 2 2 2 2 2 2 2 6 4 4 5 4

Hydrophilic, acidic: **D** Asp **E** Glu

Hydrophilic, basic: **H** His **K** Lys **R** Arg **X** TrimethylLys

Polar, uncharged: **B** Asn or Asp **G** Gly **N** Asn **Q** Gln
S Ser **T** Thr **W** Trp **Y** Tyr **Z** Gln or Glu

Hydrophobic: **A** Ala **C** Cys **F** Phe **I** Ile **L** Leu
M Met **P** Pro **V** Val

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								

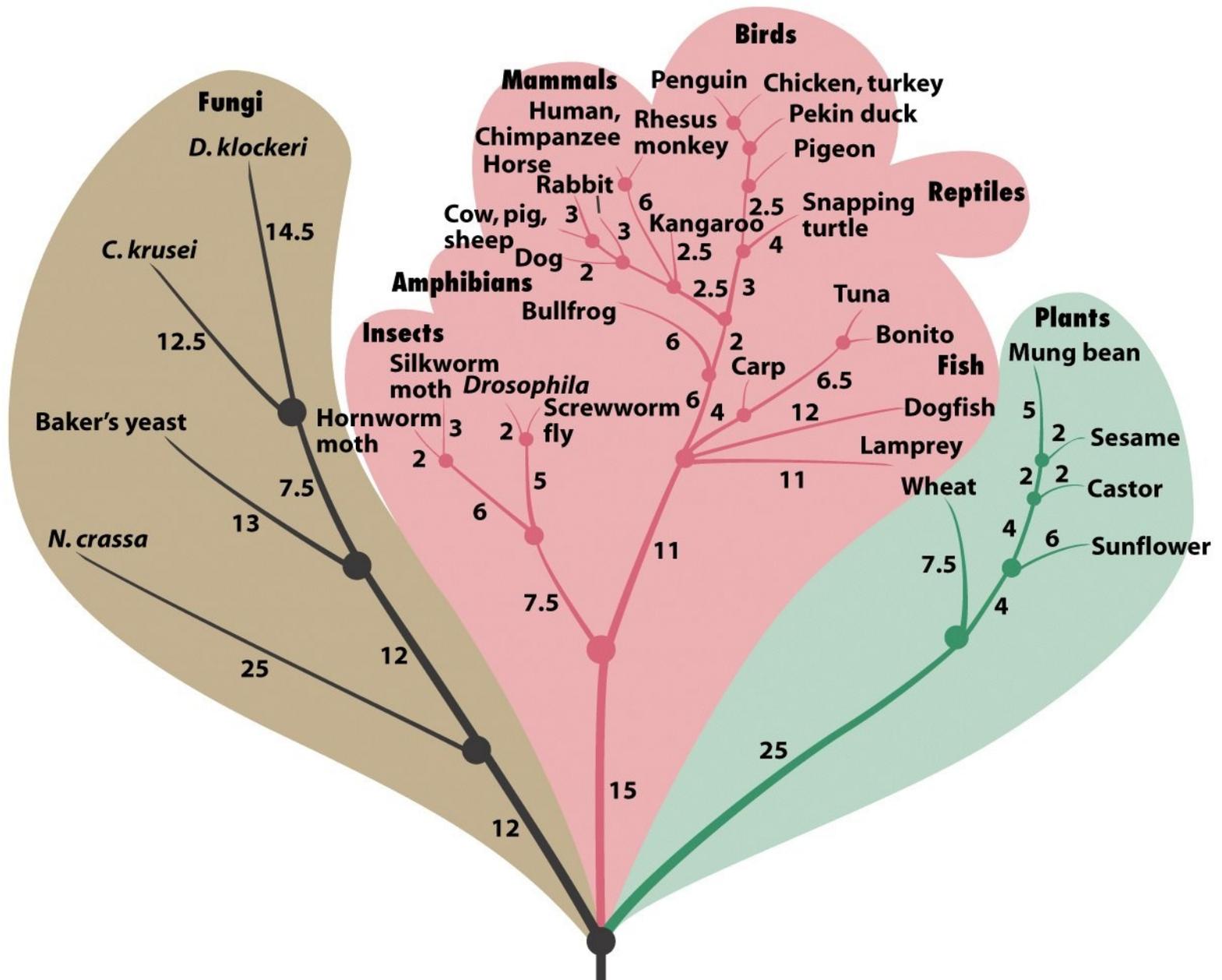


Figure 5-23 Fundamentals of Biochemistry, 2/e
 © 2006 John Wiley & Sons

Comparação das estruturas primárias de proteínas

-alinhamento

-cálculo de identidade e similaridade

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

Mb	G	L	S	D	G	E	W	Q	L	V	L	N	V	W	G	K	V	E	A	D	I	P	G	H	G	Q	E	V	L	I	R	L	F	K	G	H	P	E	T	L	40
Hb α	V	L	S	P	A	D	K	T	N	V	K	A	A	W	G	K	V	G	A	H	A	G	E	Y	G	A	E	A	L	E	R	M	F	L	S	F	P	T	T	K	40
Mb	E	K	F	D	K	F	K	H	L	K	S	E	D	E	M	K	A	S	E	D	L	K	K	H	G	A	T	V	L	T	A	L	G	G	I	L	K	K	K	G	80
Hb α	T	Y	F	P	H	F	-	-	-	-	-	D	L	S	H	G	S	A	Q	V	K	G	H	G	K	K	V	A	D	A	L	T	N	A	V	A	H	V	D	74	
Mb	H	H	E	A	E	I	K	P	L	A	Q	S	H	A	T	K	H	K	I	P	V	K	Y	L	E	F	I	S	E	C	I	I	Q	V	L	Q	S	K	H	P	120
Hb α	D	M	P	N	A	L	S	A	L	S	D	L	H	A	H	K	L	R	V	D	P	V	N	F	K	L	L	S	H	C	L	L	V	T	L	A	A	H	L	P	114
Mb	G	D	F	G	A	D	A	Q	G	A	M	N	K	A	L	E	L	F	R	K	D	M	A	S	N	Y	K	E	L	G	F	Q	G						153		
Hb α	A	E	F	T	P	A	V	H	A	S	L	D	K	F	L	A	S	V	S	T	V	L	T	S	K	Y	R												141		

% ID = 27.0

Como quantificar o melhor alinhamento (probabilidade de homologia)?

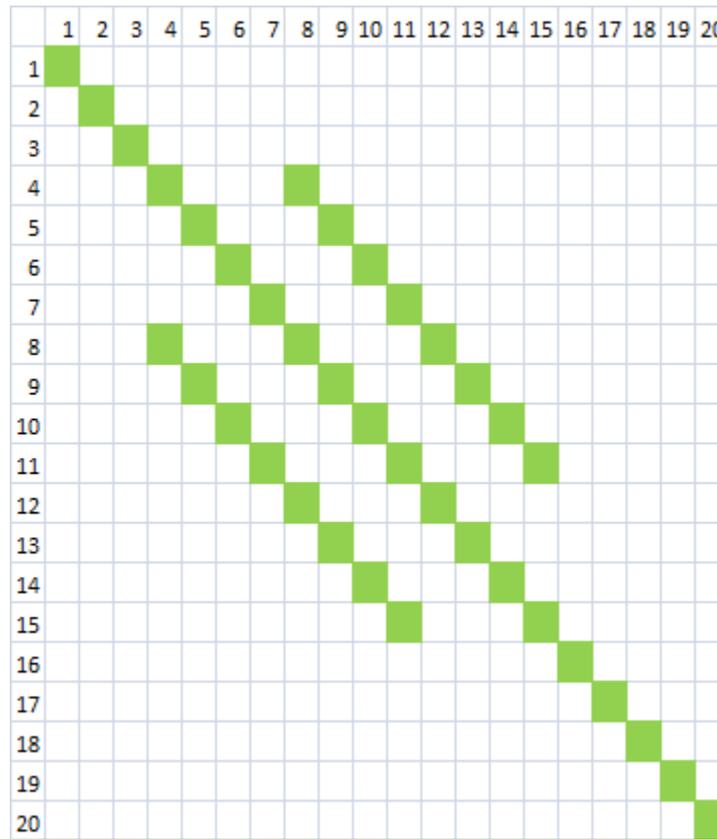
Solução “empírica”

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



Mapa de Contatos da Estrutura Terciária de Proteínas



Distância entre $C\alpha < 8 \text{ \AA}$

Distância entre quaisquer átomos $< 5 \text{ \AA}$

Mapa de Contatos da Estrutura Terciária de Proteínas



Mapa de Contatos da Estrutura Terciária de Proteínas

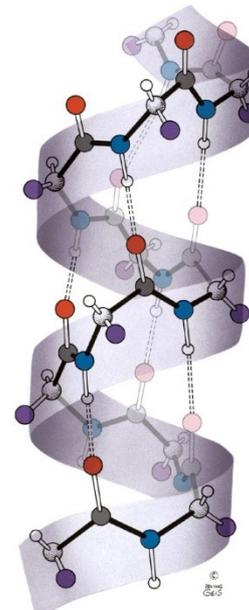
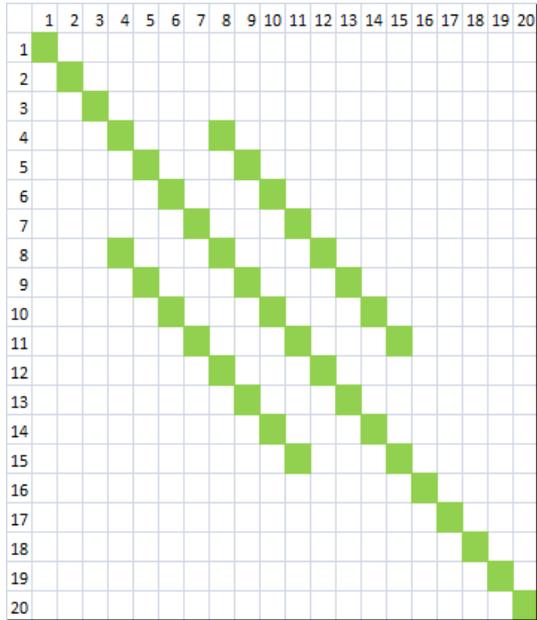


Figure 6-7 Fundamentals of Biochemistry, 2/e

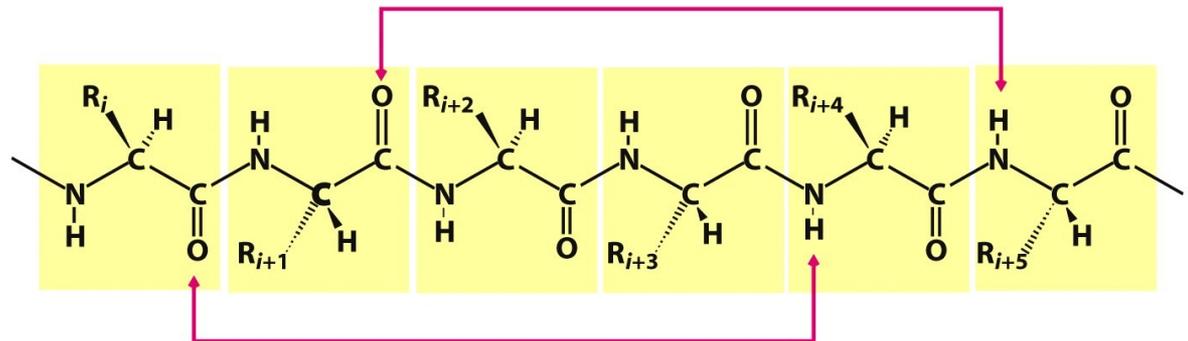
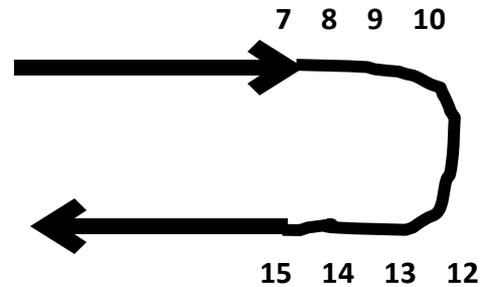
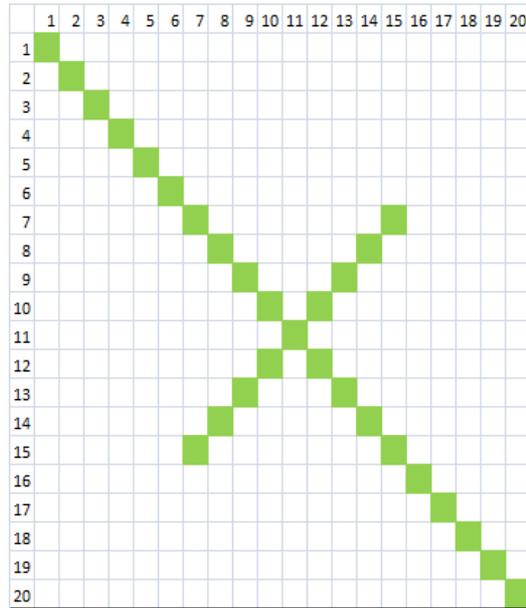
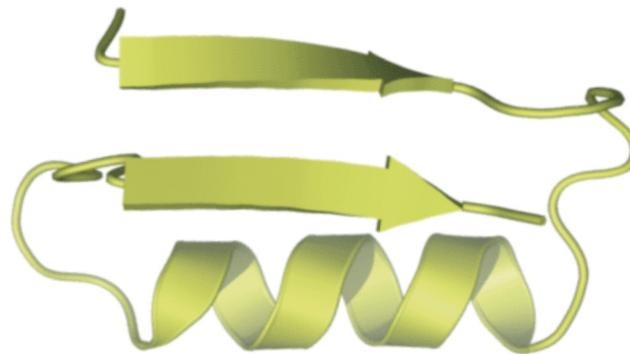
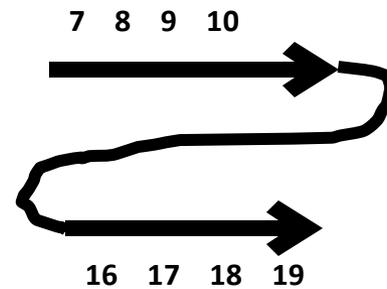
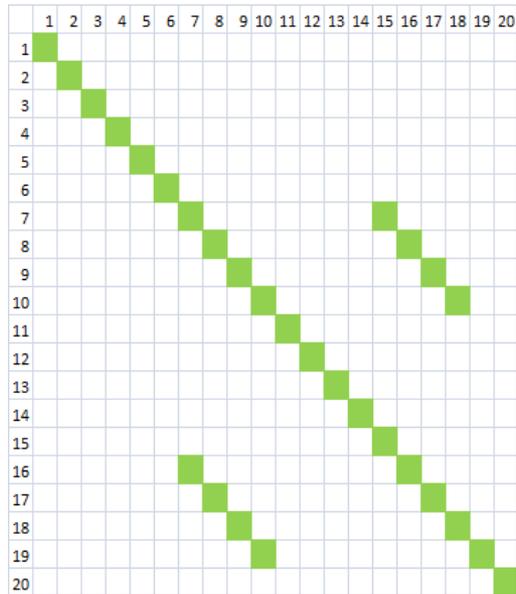


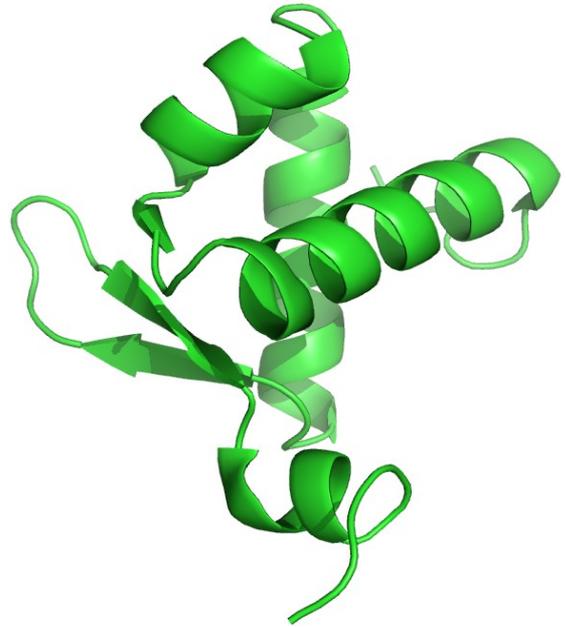
Figure 2.25
 Biochemistry, Seventh Edition
 © 2012 W. H. Freeman and Company

Mapa de Contatos da Estrutura Terciária de Proteínas



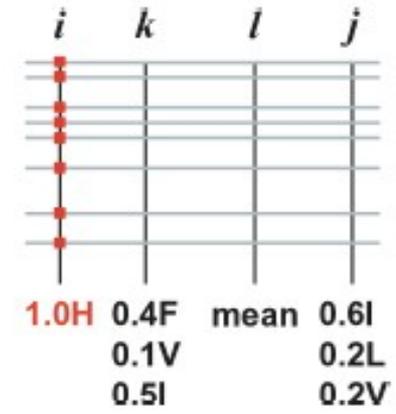
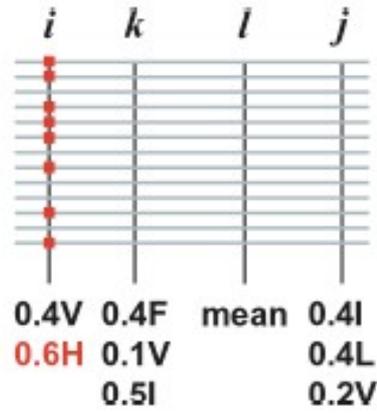
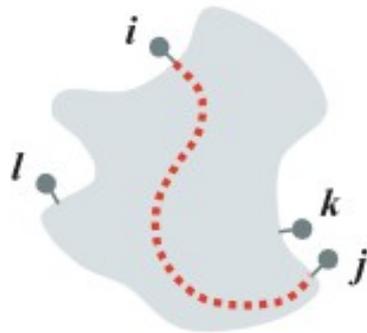
Mapa de Contatos da Estrutura Terciária de Proteínas





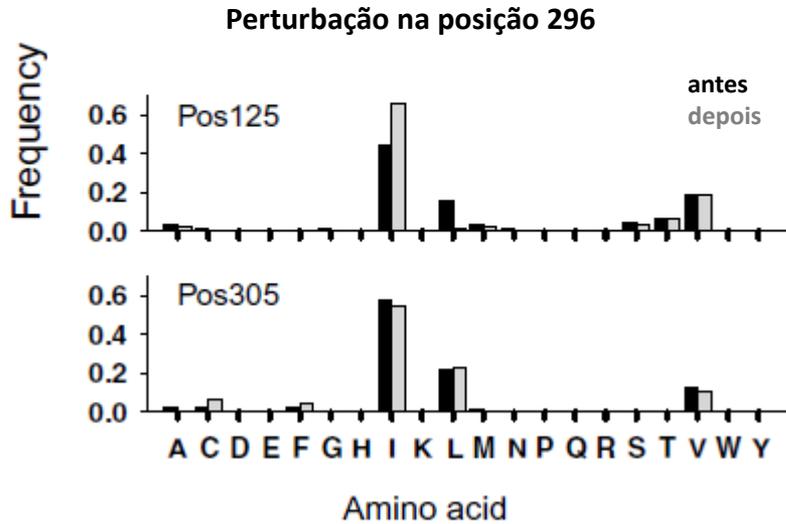


SCA – Statistical Coupling Analysis



SCA – Statistical Coupling Analysis

Análise aplicada a uma GPCR

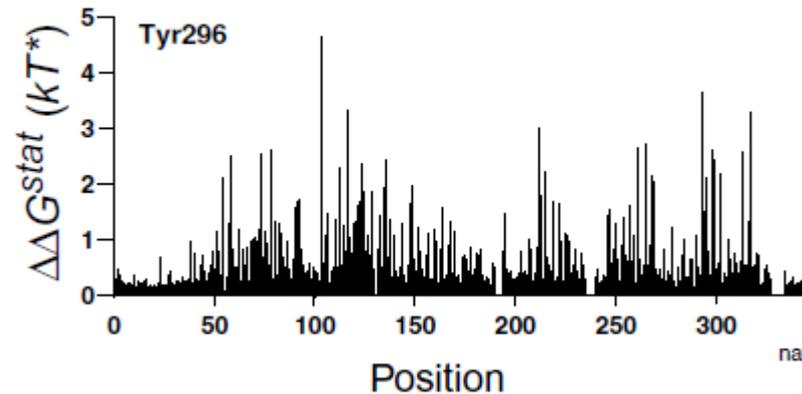


$$\Delta G_i = \sqrt{\sum_x (\ln P_i^x)^2}$$

antes

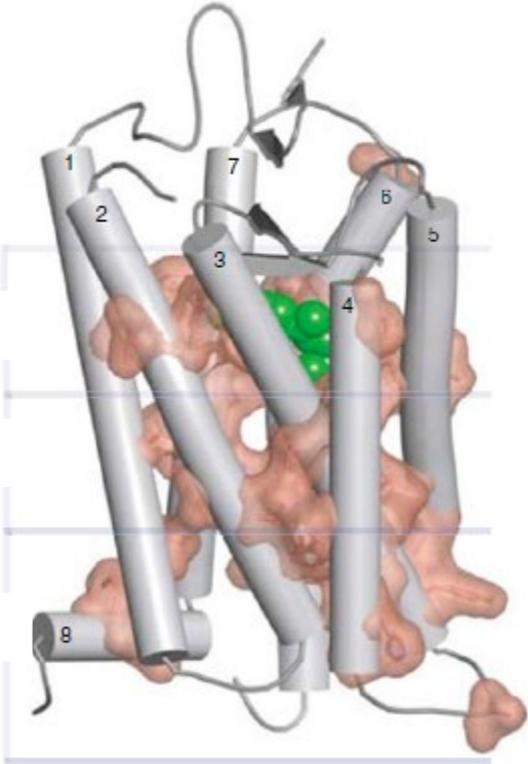
$$\Delta\Delta G_{i,j} = \sqrt{\sum_x (\ln P_{i|\delta j}^x - \ln P_i^x)^2}$$

depois

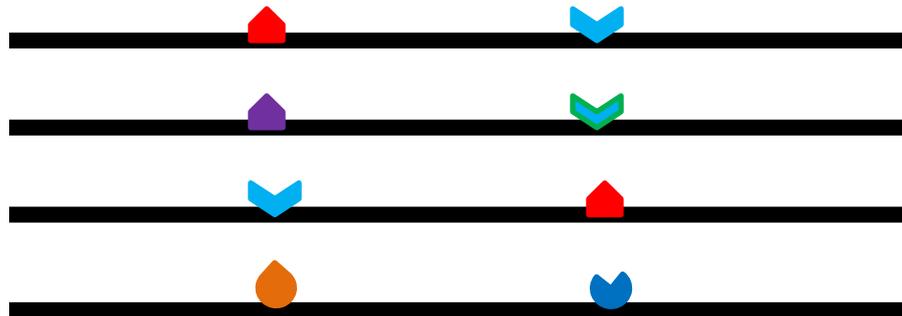
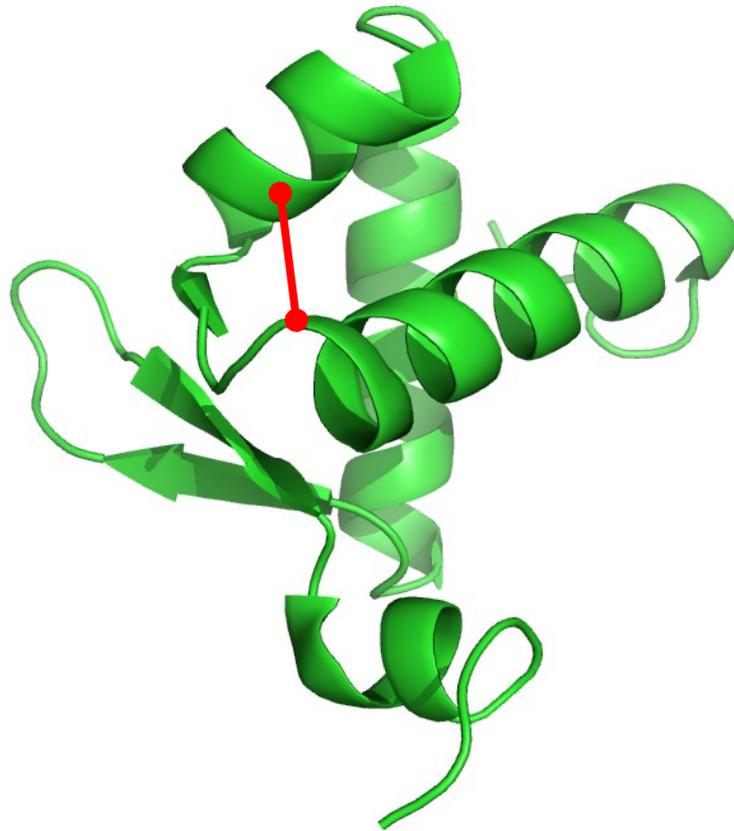


SCA – Statistical Coupling Analysis

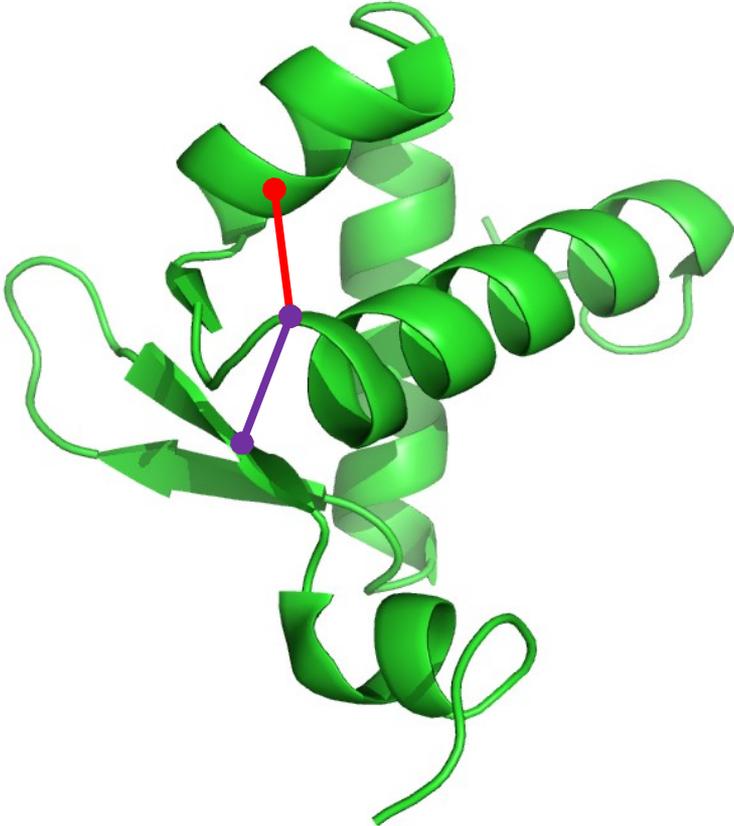
Rotas de efeitos alostéricos



Covariação e Contatos: Predição de Estrutura



SCA – Statistical Coupling Analysis



SCA x DCA – Direct Coupling Analysis

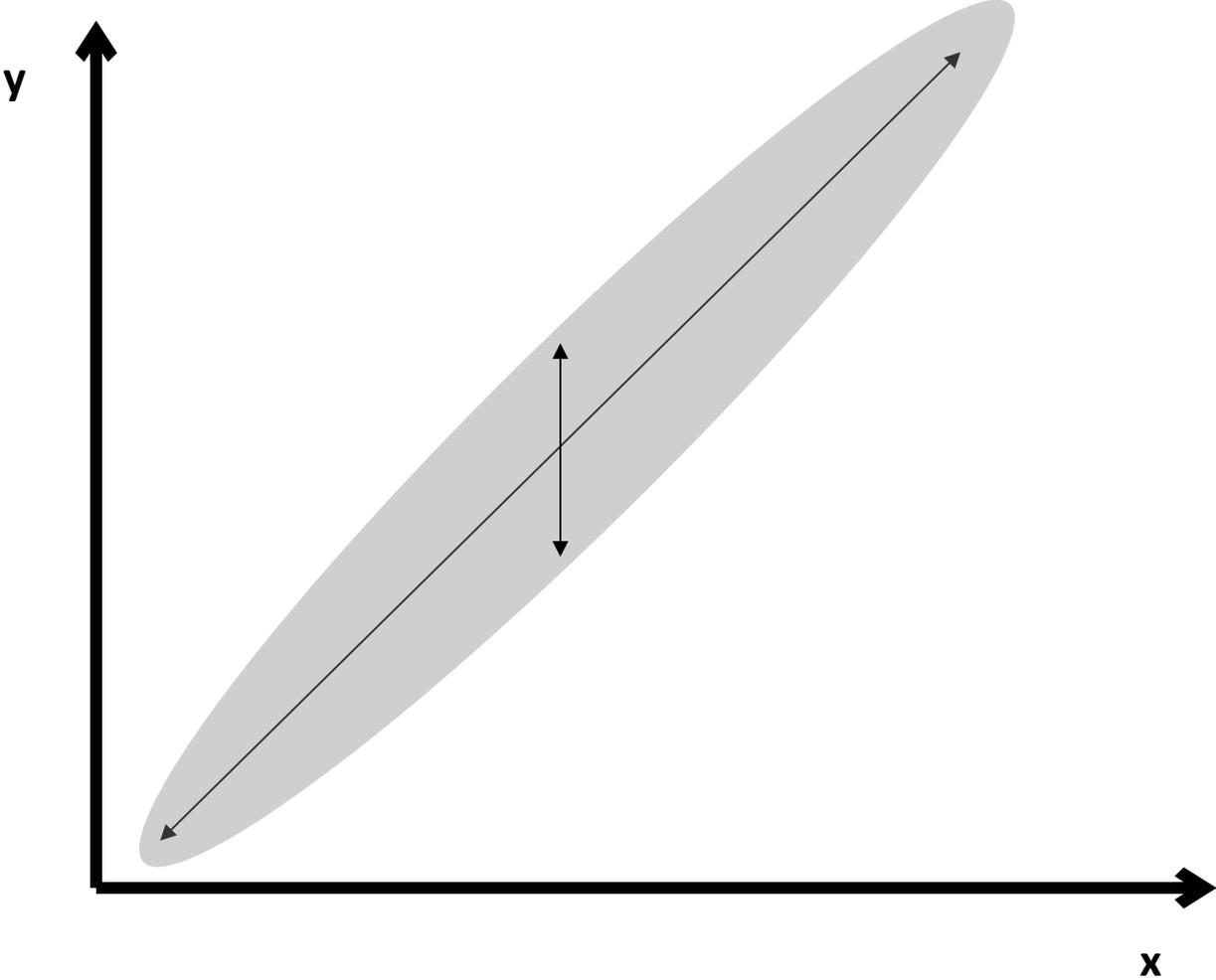
Utilização de alinhamentos de sequências com identidade menor que 80%

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

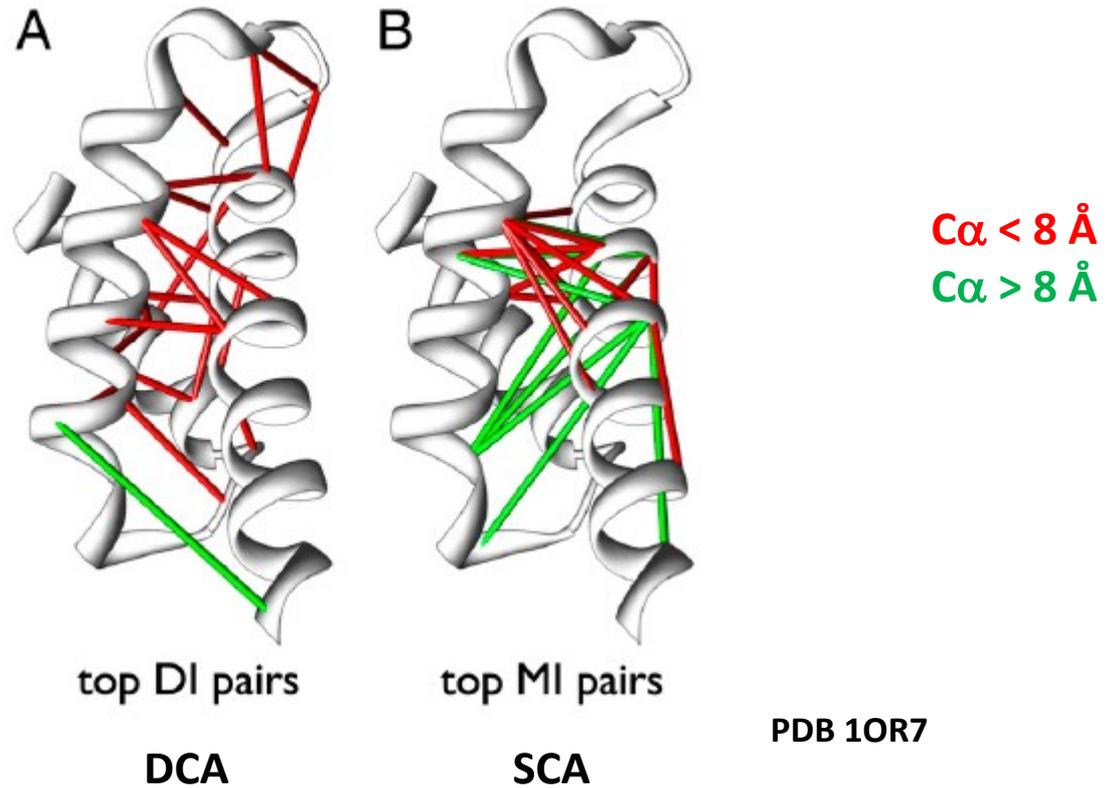
Construção de matrizes de covariação entre posições da sequência

Remoção das covariâncias secundárias (“contatos indiretos”)

SCA x DCA – Direct Coupling Analysis



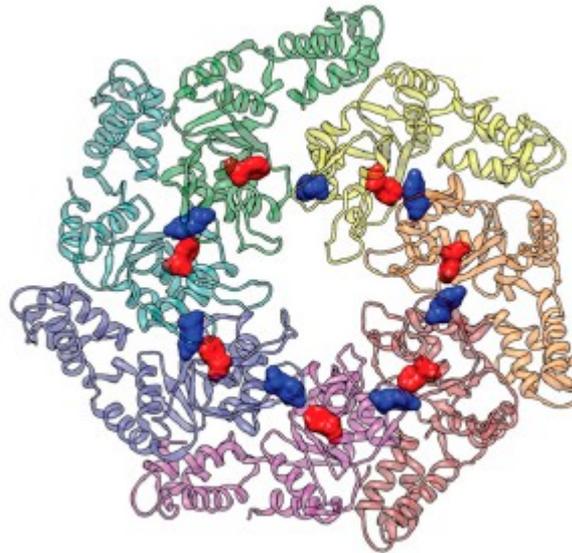
DCA – Direct Coupling Analysis



DCA – Direct Coupling Analysis

Contatos entre domínios na estrutura quaternária

Sítios de ligação de cofatores ou substratos

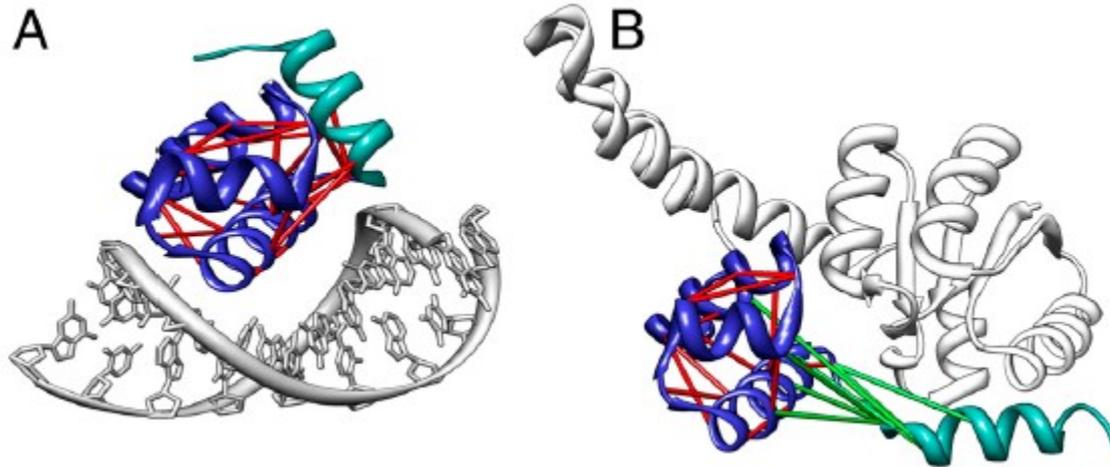


ATPase
heptâmero

PDB 1NY6

DCA – Direct Coupling Analysis

Conformações alternativas



PDB 3C3W

GREMLIN + ROSETTA

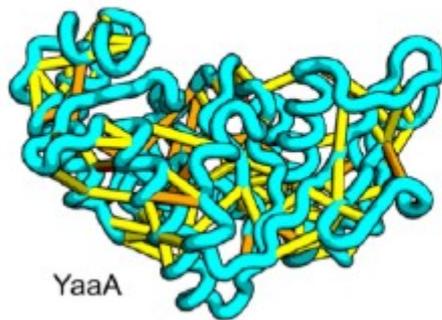
Mudança no método de cálculo da covariância evolutiva

Significativo aumento no número de sequência disponíveis

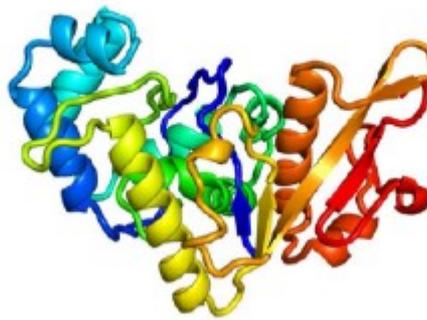
Modelagem dos ângulos de torção da cadeia principal guiada pelo mapa de contatos previsto por covariância evolutiva + modelagem dos ângulos de torção das cadeias laterais

CASP 11

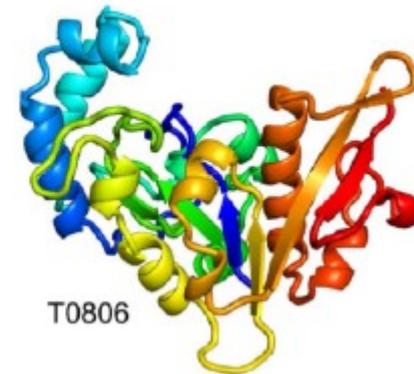
Critical Assessment of techniques for protein Structure Prediction (2014)



Pares de resíduo coevolúidos



Estrutura prevista



Estrutura cristalográfica

rmsd 2,9 Å

AlphaFold

CASP 14 (2020)

rmsd 0,96 Å (cadeia principal)

rmsd 1,5 Å (todos os átomos)

