

## **QBQ0204 Bioquímica: Estrutura de Biomoléculas e Metabolismo**

### **Guia de estudos**

#### ***Aula 3: titulação de aminoácidos e estrutura de proteínas***

Seguindo o mesmo esquema de marcação de aulas anteriores há diversos tipos de leituras nos textos a seguir (D.L. Nelson e M.M. Cox - Princípios de Bioquímica de Lehninger, 6ª ed. e Marzzocco e B. B. Torres - Bioquímica Básica, 4ª ed.). Novamente, não se assustem com o número de páginas. Estamos mantendo a leitura básica em torno de 20 páginas. Mas, como trata-se de um assunto extenso, os livros de bioquímica avançada possuem bastante informação. É importante que vocês tenham esse material para eventuais consultas. No meio das leituras avançadas, na página 16, há o tema “3.3 trabalhando com proteínas”. Observem que esta é uma leitura complementar somente da aula 4 (marcado no texto), mas para não ficar cortando o capítulo no meio ele já se encontra neste guia.

## 2

## Aminoácidos e Proteínas

As proteínas, além de constituírem o componente celular mais abundante, são as biomoléculas mais diversificadas quanto a forma e função. São o componente principal, ou único, de estruturas tão diferentes quanto a clara dos ovos, os cascos e chifres, a pele e o cabelo, o bico e as penas das aves. As funções que desempenham são estruturais e dinâmicas. Fazem parte de todas as membranas e organelas celulares, do citoesqueleto e da matriz extracelular. Participam de quase todos os processos biológicos, já que incluem as enzimas, catalisadores das milhares de reações químicas que ocorrem nos organismos. Outra função dinâmica das proteínas é o transporte de moléculas (oxigênio, lipídios etc.) e íons pelo plasma e a transferência destes compostos através das membranas. Os mecanismos de defesa do organismo incluem diversas proteínas, como as imunoglobulinas e o interferon, que atuam no combate a infecções bacterianas e virais. Muitas proteínas participam do controle global do metabolismo, devido à sua ação hormonal, como é o caso da insulina e do glucagon. São também responsáveis por mecanismos contráteis, sendo de particular importância a actina e a miosina, que atuam na contração muscular. Até mesmo a atividade dos genes é controlada por proteínas: proteínas reguladoras ligam-se ao DNA em sítios específicos, localizados próximo às extremidades dos genes, sinalizando o início e o término corretos da transcrição. Estas proteínas são capazes de reconhecer, no genoma de mamíferos, o sítio regulador de um determinado gene, dentre dezenas de milhares de genes diferentes.

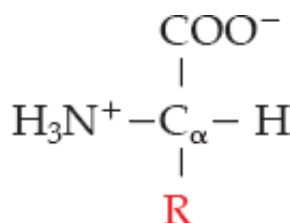
## 2.1 Aminoácidos componentes de proteínas

### As proteínas são polímeros de aminoácidos

Apesar de apresentarem estruturas e funções tão variadas, as proteínas são sintetizadas a partir de apenas 20 aminoácidos diferentes. Ainda que o número dos monômeros precursores pareça pequeno, as possibilidades de existirem proteínas distintas são espantosamente grandes. Considerando-se a formação de proteínas hipotéticas contendo somente 20 aminoácidos, um de cada tipo, poderiam ser obtidas  $2,4 \times 10^{18}$  moléculas diferentes! Como as proteínas são compostas por centenas de aminoácidos, cada um deles podendo estar representado mais de uma vez, a possibilidade de construção de moléculas diferentes é praticamente infinita.

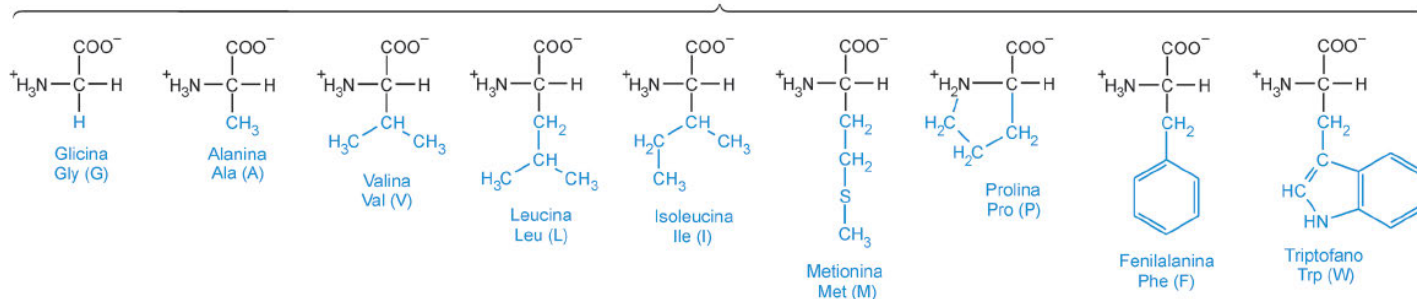
### Os aminoácidos diferem entre si pela cadeia lateral

Aminoácidos são compostos que apresentam, na sua molécula, um *grupo amino* ( $-NH_2$ ) e um *grupo carboxila* ( $-COOH$ ). Entre os aminoácidos que compõem as proteínas, a única exceção é a prolina, que contém um grupo imino ( $-NH-$ ) no lugar do grupo amino, sendo a rigor um *iminoácido*. Em pH fisiológico, esses grupos estão na forma ionizada:  $-NH_3^+$ ,  $-COO^-$  e  $-NH_2^+$ . Os aminoácidos têm uma fórmula básica comum, com os grupos amino e carboxila ligados ao carbono  $\alpha$ , ao qual também se liga um átomo de hidrogênio e um grupo variável chamado *cadeia lateral* ou *grupo R*:



As propriedades das cadeias laterais dos aminoácidos — principalmente a afinidade pela água — são importantes para a conformação das proteínas e, portanto, para sua função. De acordo com a polaridade do grupo R, os aminoácidos são classificados em duas grandes categorias: aminoácidos apolares (grupo R hidrofóbico) e aminoácidos polares (grupo R hidrofílico) (Figura 2.1).

Apolares (Hidrofóbicos)



Polares (Hidrofílicos)

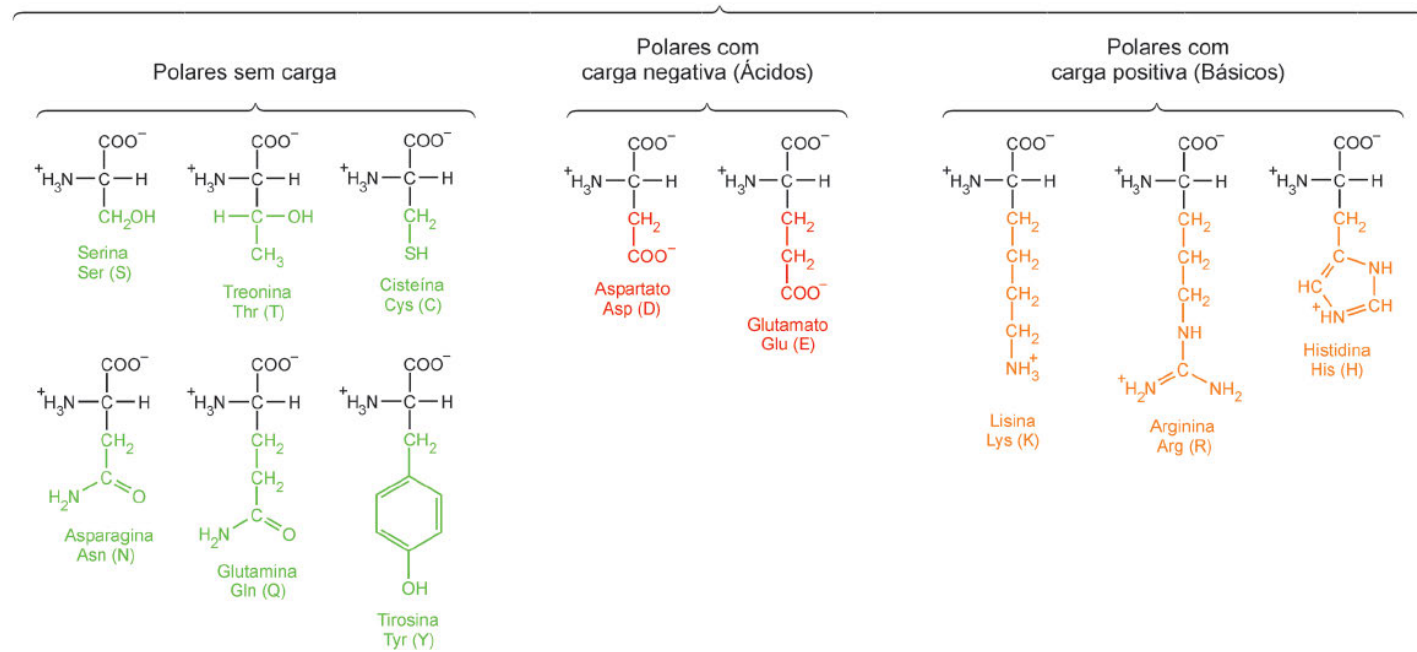


Figura 2.1 Estrutura e classificação dos aminoácidos.

Os aminoácidos *apolares* têm grupos R com caráter de hidrocarboneto, que não interagem com a água; por isso, frequentemente localizam-se no interior da molécula proteica. Pertencem a este grupo: *glicina, alanina, valina, leucina, isoleucina, metionina, prolina, fenilalanina e triptofano*.

Os aminoácidos *polares* têm, nas cadeias laterais, grupos com carga elétrica líquida ou grupos com cargas residuais, que os capacitam a interagir com a água. São geralmente encontrados na superfície da molécula proteica. Estes aminoácidos são subdivididos em três categorias, segundo a carga apresentada pelo grupo R em pH 7: *aminoácidos básicos*, se a carga for positiva; *aminoácidos ácidos*, se a carga for negativa; e *aminoácidos polares sem carga*, se a cadeia lateral não apresentar carga líquida.

Os aminoácidos básicos são *lisina, arginina e histidina*. O valor de  $pK_a$  do grupo ionizável presente na cadeia lateral da lisina (amino, com  $pK_a = 10,54$ ) e da arginina (guanidino, com  $pK_a = 12,48$ ) mostra que, em pH neutro, esses grupos estão protonados e com carga positiva (Tabela 2.1). A cadeia lateral da histidina (grupo imidazólico, com  $pK_a = 6,04$ ) está muito menos ionizada em pH 7; como o valor de seu  $pK_a$  está uma unidade abaixo deste pH, apenas 10% das moléculas de histidina estarão com o grupo R protonado (ver a equação de Henderson-Hasselbalch, Seção 1.3) e com carga positiva.

Os aminoácidos ácidos são os dicarboxílicos: *aspartato e glutamato*. Os valores de  $pK_a$  das carboxilas de suas cadeias laterais são 3,90 e 4,07, respectivamente, e em pH neutro, estão desprotonadas e com carga negativa.

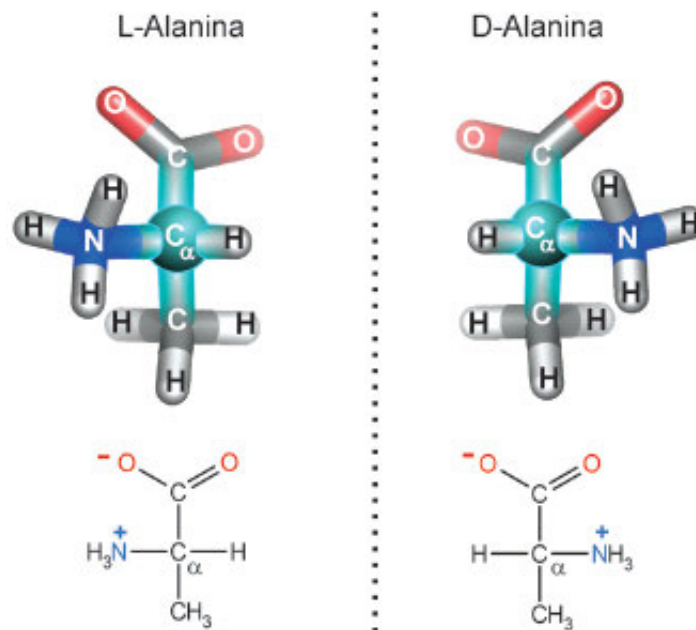
Os aminoácidos polares sem carga são *serina, treonina e tirosina*, com um grupo hidroxila na cadeia lateral; *asparagina e glutamina*, com um grupo amida; e *cisteína*, com um grupo sulfidríla. Em pH neutro, o grupo sulfidríla da cadeia lateral da cisteína ( $pK_a = 8,37$ ) está predominantemente protonado, sem carga; assim também se encontra o grupo fenólico da tirosina ( $pK_a = 10,46$ ).

O valor do  $pK_a$  de um grupo ionizável é determinado pela estrutura da molécula da qual faz parte, mas sofre influência de grupos adjacentes. Por isto, um mesmo grupo pode apresentar valores diferentes de  $pK_a$  dependendo da região da proteína em que se encontra. A Tabela 2.1 apresenta os valores mais usuais de  $pK_a$  que os aminoácidos podem apresentar quando estão unidos, formando proteínas.

Aminoácidos livres	pK <sub>1</sub> (α- COO <sup>-</sup> )	pK <sub>2</sub> (α- NH <sub>3</sub> <sup>+</sup> )	pK <sub>R</sub> (grupo R)			
Glicina	2,35	9,78				
Alanina	2,35	9,87				
Valina	2,29	9,74				
Leucina	2,33	9,74				
Isoleucina	2,32	9,76				
Metionina	2,13	9,28				
Prolina	1,95	10,64				
Fenilalanina	2,20	9,31				
Triptofano	2,46	9,41				
Serina	2,19	9,21				
Treonina	2,09	9,10				
				Aminoácidos em proteínas		
				COO <sup>-</sup> terminal	NH <sub>3</sub> <sup>+</sup> terminal	Grupos R
Asparagina	2,14	8,72				
Glutamina	2,17	9,13		3,5–4,0	7,6–9,0	
Cisteína	1,92	10,70	8,37			8,0–9,0
Tirosina	2,20	9,21	10,46			9,5–10,5
Lisina	2,16	9,06	10,54			9,5–10,5
Arginina	1,82	8,99	12,48			11,5–12,5
Histidina	1,80	9,33	6,04			6,0–7,4
Aspartato	1,99	9,90	3,90			4,0–5,5
Glutamato	2,10	9,47	4,07			4,0–5,5

### As proteínas são formadas por L-aminoácidos

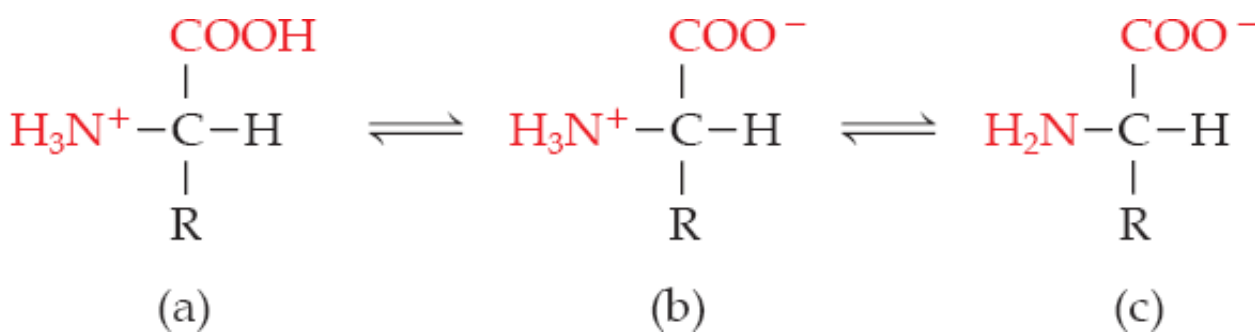
O carbono α de todos os aminoácidos, com exceção da glicina, é assimétrico, já que está ligado a quatro grupos diferentes: – NH<sub>3</sub><sup>+</sup>, – COO<sup>-</sup>, – H e – R. Na glicina, este carbono não é assimétrico porque o grupo R é constituído por – H. Os aminoácidos com carbono α assimétrico apresentam dois isômeros opticamente ativos, os isômeros D e L, que são imagens especulares um do outro (Figura 2.2). Todas as proteínas encontradas nos seres vivos são formadas por L-aminoácidos. Os D-aminoácidos aparecem somente em certos antibióticos e em peptídios componentes da parede de algumas bactérias. Os aminoácidos com configuração D, assim como outros aminoácidos exóticos encontrados nas células, são formados por modificações dos 20 aminoácidos proteicos ou são produtos intermediários das vias de síntese destes compostos.



**Figura 2.2** Isômeros L e D da alanina, em duas representações moleculares diferentes. A linha pontilhada indica o plano de um espelho.

## 2.2 Ionização dos aminoácidos

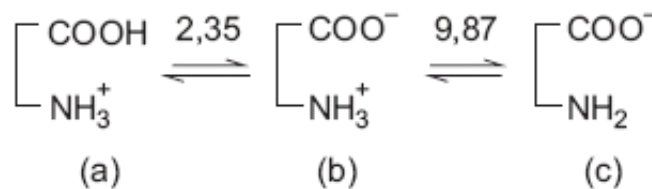
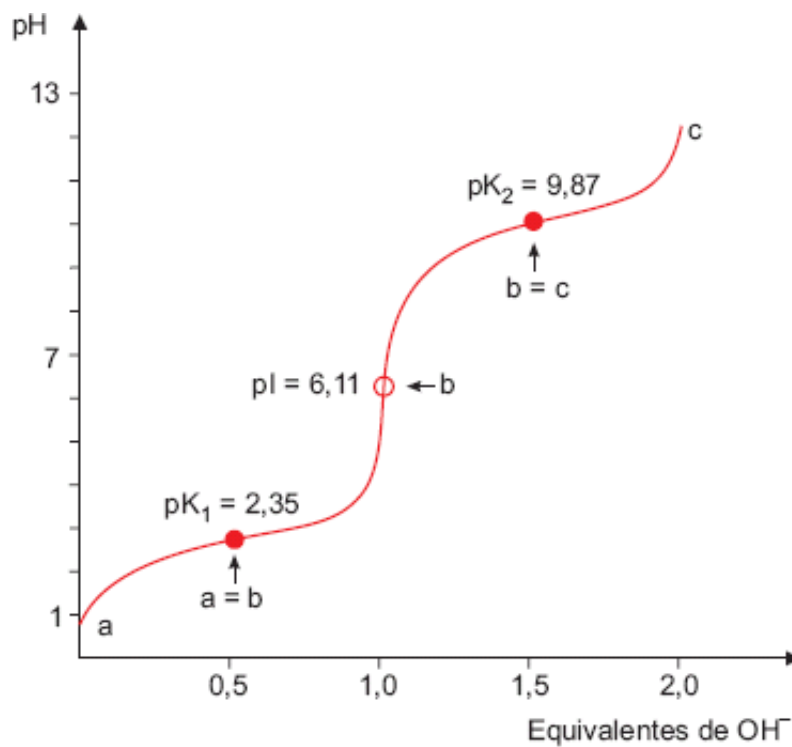
Dependendo do pH do meio, os aminoácidos podem existir nas formas seguintes:



Como se deduz pelos valores de seus  $\text{pK}_a$  (Tabela 2.1), em soluções muito ácidas, os dois grupos apresentam-se protonados (a); em soluções muito alcalinas, ambos apresentam-se desprotonados (c); e, em soluções neutras ou na forma cristalina, o aminoácido apresenta-se como um íon dipolar (b).

A conversão entre as formas *a*, *b* e *c* em função do pH do meio é refletida na curva de titulação do aminoácido. Quando o aminoácido tem apenas dois grupos ionizáveis, como a alanina, a sua curva de titulação (Figura 2.3) assemelha-se à composição das curvas de titulação de dois ácidos fracos com valores de  $\text{pK}_a$  muito diferentes, como, por exemplo, um ácido carboxílico e uma amina primária (Seção 1.3). Esta semelhança é resultado da presença do grupo carboxila e do grupo amino na mesma molécula. As duas regiões de tamponamento observadas na curva correspondem à ionização do grupo carboxila (conversão da forma *a* em *b*) e do grupo amino (conversão de *b* em *c*), com valores de  $\text{pK}_a$  de 2,35 e 9,87, respectivamente.

Outros aminoácidos monoamínicos e monocarboxílicos apresentam curvas de titulação similares, apenas variando a região de pH em que se dá o tamponamento. A curva de titulação de aminoácidos com cadeias laterais ionizáveis apresenta uma terceira região de tamponamento, correspondente ao seu terceiro  $\text{pK}_a$ . Isto ocorre com os aminoácidos ácidos, básicos, cisteína e tirosina.



**Figura 2.3** Curva de titulação da alanina e esquema de suas formas iônicas *a*, *b*, *c*.

### Ponto isoelétrico é um valor de pH

A carga elétrica total da molécula de um aminoácido resulta da soma algébrica das cargas apresentadas pelos seus grupos ionizáveis, os quais, por sua vez, dependem dos valores de seus  $pK_a$  e do pH do meio.

A curva de titulação de um aminoácido monoamínico e monocarboxílico, representada na Figura 2.3, inicia-se em pH muito baixo, menor do que o  $pK_a$  do grupo carboxila. Nesta situação, tanto a carboxila quanto o grupo amino estarão protonados, o que confere à maioria das moléculas do aminoácido uma carga líquida positiva (forma *a*). À medida que se adiciona álcali, o valor do pH sobe gradativamente, aumentando a dissociação do grupo carboxila e, conseqüentemente, a concentração da forma com uma carga negativa e uma positiva, a forma eletricamente neutra (forma *b*). Prosseguindo a adição de álcali, o valor do pH continua aumentando, promovendo a dissociação do grupo amino e o aumento da concentração da forma com carga negativa (forma *c*).

A forma eletricamente neutra (forma *b*) predomina nos valores de pH acima do  $pK_a$  do grupo carboxila e abaixo do  $pK_a$  do grupo amino e é mais abundante no pH equidistante dos dois valores de  $pK_a$ . Neste pH, há quantidades extremamente pequenas, mas iguais, da forma *a* (com carga positiva) e *c* (com carga negativa), ou seja, há perfeita compensação isoelétrica. Este valor de pH é definido como *ponto isoelétrico* (pI) do aminoácido. No pH equivalente ao pI, os aminoácidos comportam-se como moléculas neutras: não migram quando submetidos a um campo elétrico (Seção 2.10.2).

Os aminoácidos monoamínicos e dicarboxílicos (aspartato e glutamato) possuem um grupamento que pode apresentar carga positiva e dois grupamentos que podem apresentar carga negativa. Neste caso, a forma com carga líquida igual a zero será obtida quando um dos grupos carboxila estiver protonado (sem carga) e o outro desprotonado (com carga negativa): a carga negativa do grupo carboxila desprotonado será compensada pela carga positiva do grupamento amino protonado. O pH em que esta é a situação da enorme maioria das moléculas equidista dos valores de  $pK_a$  das duas carboxilas.

Por um raciocínio análogo, para os aminoácidos com um grupo carboxila e dois grupos que podem apresentar carga positiva (lisina, histidina e arginina), a forma eletricamente neutra será mais abundante em um valor de pH equidistante dos valores de  $pK_a$  dos dois grupamentos básicos do aminoácido.

Generalizando, o pH em que a forma eletricamente neutra do aminoácido é mais abundante, o ponto isoelétrico (pI) é a média aritmética de dois valores de  $pK_a$ :

$$pI = \frac{pK_{a1} + pK_{a2}}{2}$$

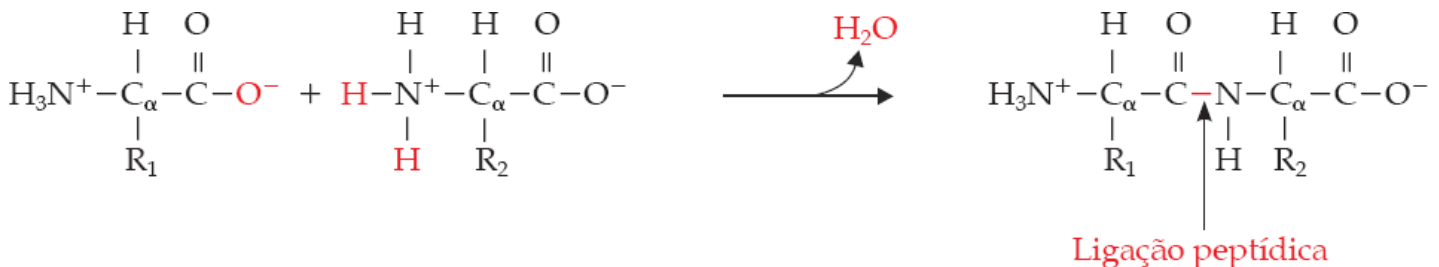
Os valores de  $pK_{a1}$  e  $pK_{a2}$  a serem considerados no cálculo do pI dependem do aminoácido. Para os que não contêm grupamentos ionizáveis na cadeia lateral, utilizam-se os valores de  $pK_a$  dos grupos amino e carboxila; para aminoácidos com três grupamentos ionizáveis, usam-se os valores de  $pK_a$  dos grupos com mesmo sinal de carga. A única exceção a esta regra é a *tirosina*, que apresenta valores de  $pK_a$  iguais a 2,20 (carboxila), 9,21 (amino) e 10,46 (fenólico). O grupo fenólico apresentará carga negativa em valores de pH maiores do que o  $pK_a$  do grupo amino, e o ponto isoelétrico é a média entre o  $pK_a$  do grupo carboxila e o do grupo amino.

### Os aminoácidos não constituem tampões fisiológicos importantes

Uma análise dos valores de  $pK_a$  dos aminoácidos revela que eles não apresentam poder tamponante significativo em pH fisiológico. Todavia, quando fazem parte de uma proteína, o valor de  $pK_a$  de suas cadeias laterais e dos grupos carboxila e amino terminais da proteína (Tabela 2.1) podem sofrer variações consideráveis, em relação ao seu valor no aminoácido livre. Essas alterações resultam das características do microambiente existente na região da proteína onde o aminoácido se encontra: polaridade, determinada pela proximidade de grupamentos com carga, exclusão de moléculas de água, como ocorre no interior da molécula proteica, e possibilidade de formação de ligações de hidrogênio (Seção 2.4). Um caso importante de alteração de  $pK_a$  de aminoácidos, que tem implicações decisivas para o tamponamento do sangue, é o da *hemoglobina*, analisado no Capítulo 3.

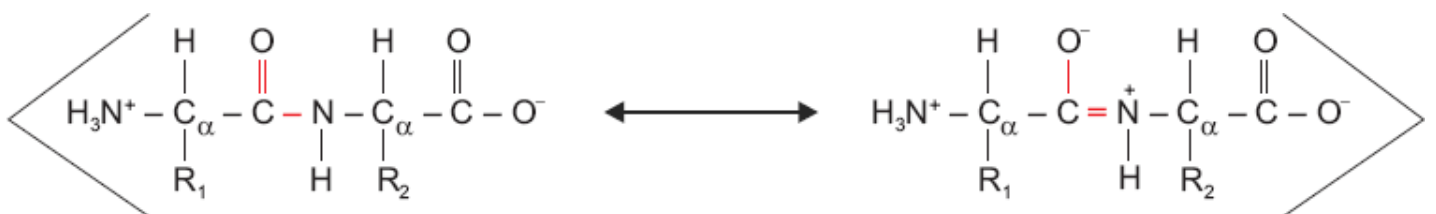
## 2.3 Polímeros de aminoácidos: peptídios e proteínas

Os aminoácidos podem formar polímeros lineares pela ligação do grupo  $\alpha$ -carboxila de um aminoácido com o grupo  $\alpha$ -amino de outro. Esta ligação carbono-nitrogênio é uma ligação *amídica*, chamada, no caso das proteínas, de *ligação peptídica*. É obtida, teoricamente, por exclusão de uma molécula de água e sua formação pode ser representada pela seguinte equação:



Esta reação, como está escrita, *jamais* ocorre, pois não é termodinamicamente viável. Nos seres vivos, a ligação peptídica não é feita por reação direta entre os aminoácidos, mas por um complexo aparato de síntese proteica, que inclui ribossomos, ácidos ribonucleicos, várias proteínas e enzimas. A síntese compreende uma sequência de etapas, envolvendo um expressivo gasto de ATP, que viabiliza termodinamicamente o processo. A equação é apenas um esquema didático para descrever a formação da ligação peptídica.

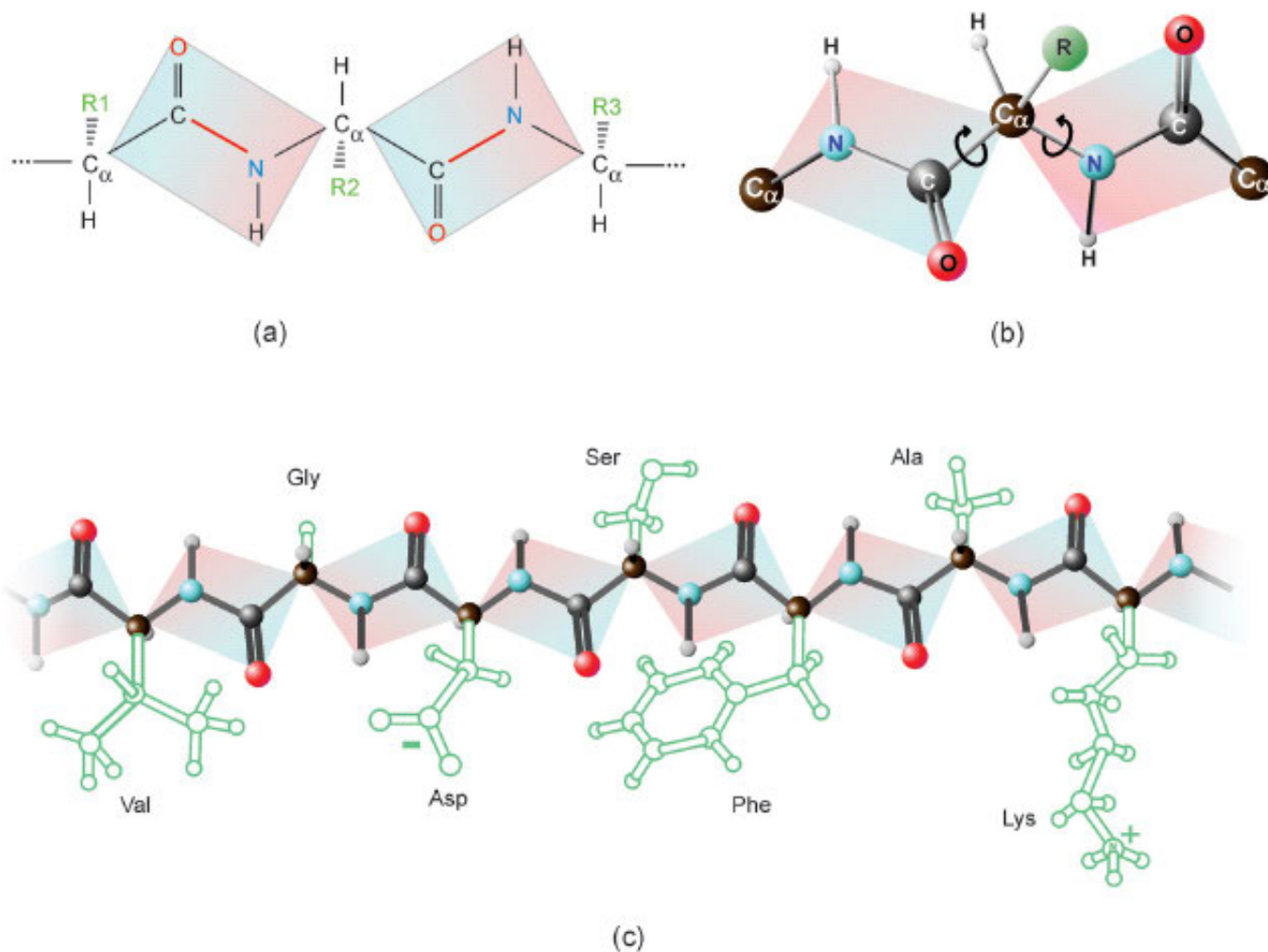
As propriedades da ligação peptídica impõem restrições ao dobramento do polímero formado. Apesar de ser representada por um único traço de ligação, ela tem características intermediárias entre uma ligação simples e uma dupla ligação, devido à ressonância entre duas formas:



A consequência desse caráter parcial de dupla ligação é não haver possibilidade de rotação em torno da ligação peptídica. Os quatro átomos dos grupamentos que participam da ligação peptídica – C, O, N, H – ficam dispostos em um plano rígido, constituindo o que se costuma chamar de *grupo peptídico* ou *unidade peptídica* (cada retângulo na Figura 2.4 a). Todavia, existem pontos de dobramento entre as unidades peptídicas, graças à possibilidade de rotação (indicada por setas na Figura 2.4 b) em torno das ligações com o carbono  $\alpha$  ( $\text{C}_\alpha - \text{C}$  e  $\text{C}_\alpha - \text{N}$ ), que são ligações efetivamente simples.



O polímero de aminoácidos pode, então, ser visualizado como uma cadeia constituída por unidades planares, as unidades peptídicas, ligadas por uma articulação flexível — o carbono  $\alpha$  (Figura 2.4 c). Esta cadeia chama-se *cadeia polipeptídica*, que, graças a esse arranjo estrutural, pode dobrar-se de muitas maneiras diferentes.



**Figura 2.4** Cadeia polipeptídica. a) Unidade peptídica, com a ligação peptídica em vermelho. b) Rotação das unidades peptídicas em torno do carbono  $\alpha$ . c) Segmento de uma cadeia polipeptídica, com as cadeias laterais dos aminoácidos em verde.

A cadeia polipeptídica pode conter um número variável de aminoácidos (ou, mais rigorosamente, *resíduos de aminoácidos*, já que, na formação de cada ligação peptídica, foi eliminada uma molécula de água). Quando o número de aminoácidos é igual a 2, o polímero é chamado de *dipeptídio*; com 3 é um *tripeptídio* e assim por diante. Polímeros contendo até 30 aminoácidos são chamados de *oligopeptídios*; quando o número é maior, podendo chegar a centenas ou milhares, são chamados de *polipeptídios*. As cadeias polipeptídicas que podem ser associadas a uma função recebem a designação de *proteínas*. Qualquer que seja o número de aminoácidos, os peptídios apresentam um grupamento amino livre em uma das extremidades — amino terminal — e um grupo carboxila livre na outra — carboxila terminal, além dos grupos R dos aminoácidos.

Muitos peptídios encontrados na natureza desempenham funções importantes, atuando como hormônios (encefalinas, oxitocina, vasopressina, glucagon), antibióticos (gramicidina), agentes redutores (glutaciona) etc. (Tabela 2.2). Peptídios sintéticos têm aplicações diversas; um exemplo é o aspartame, um adoçante artificial, com alto poder edulcorante. O aspartame é um dipeptídio modificado, formado por aspartato e fenilalanina esterificada a um grupo metila.

**Tabela 2.2** Peptídios de importância biológica.

Peptídios	Número de aminoácidos	Glândulas/células produtoras	Efeitos principais
Encefalinas	5	Hipófise anterior e medula adrenal	Analgesia
Oxitocina	9	Hipófise posterior	Contração da musculatura uterina no parto e de glândulas mamárias na lactação
Vasopressina	9	Hipófise posterior	Aumento da pressão sanguínea e da reabsorção de água pelo rim



Glucagon	29	Células $\alpha$ do pâncreas	Aumento da produção de glicose pelo fígado no jejum
Gramicidina	10	Cepas de <i>Bacillus brevis</i>	Antibiótico
Glutaciona	3	Maioria das células	Proteção de grupos SH de proteínas, manutenção do Fe <sup>2+</sup> da hemoglobina e dissipação de H <sub>2</sub> O <sub>2</sub>

As *proteínas* podem ser formadas por uma ou mais cadeias polipeptídicas e contêm, geralmente, mais de 50 aminoácidos. A maior cadeia polipeptídica conhecida de uma proteína é a *titina*, também chamada *conectina*, uma proteína estrutural de músculos estriados (Tabela 2.3); a fórmula molecular da isoforma mais comum é C<sub>169 723</sub>H<sub>270 464</sub>N<sub>45 688</sub>O<sub>52 243</sub>S<sub>912</sub>. As proteínas desempenham uma função específica e, com poucas exceções (colágeno, Seção 2.5), contêm todos os 20 aminoácidos, em proporções que variam muito de proteína para proteína (Tabela 2.4). Cada proteína apresenta uma estrutura tridimensional definida e característica. Apesar de existirem inúmeras conformações teoricamente possíveis, todas as moléculas de uma dada proteína assumem a mesma conformação espacial. Esta configuração, entretanto, não é permanentemente fixa, e, muitas vezes, alterações estruturais transitórias estão relacionadas com o controle da função desempenhada pela proteína.

**Tabela 2.3 Características de composição de algumas proteínas.**

Proteína	Número de aminoácidos	Número de cadeias polipeptídicas
Insulina (bovina)	51	2
Lisozima (clara de ovo)	129	1
Mioglobina (equina)	153	1
Hemoglobina (humana)	574	4
Aspartato transcarbamoilase ( <i>E. coli</i> )	2.700	12
RNA polimerase ( <i>E. coli</i> )	4.100	5
Titina (humana)	34.530	1

**Tabela 2.4 Composição em aminoácidos de três proteínas.**

Aminoácido	Número de aminoácidos por molécula de proteína		
	Quimotripsinogênio (bovino)	Lisozima (clara de ovo)	Citocromo <i>c</i> (humano)
Glicina	23	12	13
Alanina	22	12	6
Valina	23	6	3
Leucina	19	8	6
Isoleucina	10	6	8
Metionina	2	2	3
Prolina	9	2	4
Fenilalanina	6	3	3
Triptofano	8	6	1
Serina	28	10	2
Treonina	23	7	7
Asparagina	15	13	5
Glutamina	10	3	2
Tirosina	4	3	5
Cisteína	10	8	2

Lisina	14	6	18
Arginina	4	11	2
Histidina	2	1	3
Aspartato	8	8	3
Glutamato	5	2	8
<b>Total</b>	<b>245</b>	<b>129</b>	<b>104</b>

As proteínas são classificadas como globulares ou fibrosas, segundo sua forma. As *proteínas globulares* apresentam uma ou mais cadeias polipeptídicas organizadas em uma forma final aproximadamente esférica; são geralmente solúveis e desempenham funções dinâmicas. As *proteínas fibrosas* têm forma alongada, são geralmente insolúveis e desempenham um papel basicamente estrutural nos sistemas biológicos.

## 2.4 Estrutura das proteínas

### A sequência de aminoácidos determina a estrutura espacial da proteína

A organização espacial<sup>1</sup> da proteína é resultante do tipo de aminoácidos que a compõem e de como eles estão dispostos uns em relação aos outros. A sequência dos aminoácidos irá determinar o tipo de interação possível entre as cadeias laterais, que apresentam características de carga, volume e reatividade com a água muito variáveis. A organização tridimensional de uma proteína, desde a sequência de aminoácidos, passando pelo enrolamento da cadeia polipeptídica até a associação de várias cadeias, pode ser descrita em níveis estruturais de complexidade crescente.

Nesta seção, as estruturas tridimensionais descritas referem-se às proteínas globulares.

### A estrutura das proteínas pode ser descrita em quatro níveis

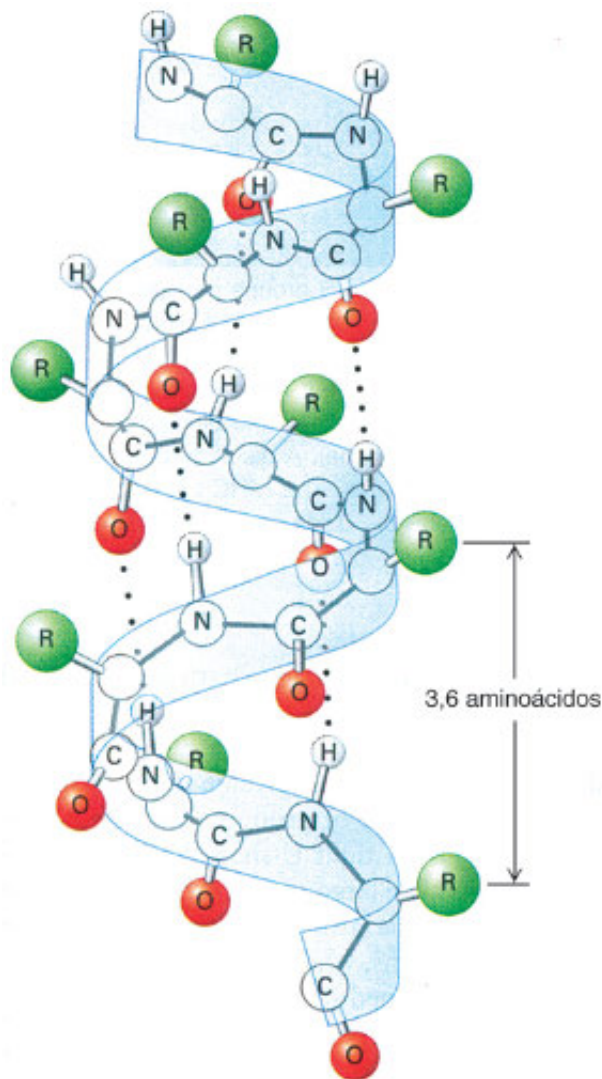
A **estrutura primária** é a sequência de aminoácidos da cadeia polipeptídica, determinada geneticamente e específica para cada proteína. Por convenção, a estrutura primária é escrita na direção amino terminal → carboxila terminal. Exemplificando, os peptídeos Ala-Ser-Lys e Lys-Ser-Ala são diferentes, porque, no primeiro caso é o grupo amino da alanina que está livre e, no segundo caso, o da lisina<sup>2</sup>.

A **estrutura secundária** descreve as estruturas tridimensionais regulares, formadas por segmentos da cadeia polipeptídica. Duas organizações são particularmente estáveis: o enrolamento da cadeia ao redor de um eixo e a interação lateral de segmentos de uma cadeia polipeptídica ou de cadeias diferentes. Estas conformações são denominadas, respectivamente,  $\alpha$ -hélice e folha  $\beta$  pregueada. A extensão do segmento da cadeia polipeptídica que se organiza nessas duas configurações pode variar de alguns a dezenas de aminoácidos, conforme a proteína. A  $\alpha$ -hélice e a folha  $\beta$  pregueada estabilizam-se por *ligações de hidrogênio*<sup>3</sup> entre o nitrogênio e o oxigênio dos grupos – NH e – C = O, constituintes das unidades peptídicas. Embora a ligação de hidrogênio seja uma interação fraca, o elevado número destas ligações confere grande estabilidade a essas estruturas.

A  $\alpha$ -hélice é mantida por ligações de hidrogênio formadas entre uma unidade peptídica e a quarta unidade peptídica subsequente; estas ligações dispõem-se paralelamente ao eixo da hélice. A  $\alpha$ -hélice tem um passo de 0,54 nm e apresenta 3,6 resíduos de aminoácidos por volta (Figura 2.5). As cadeias laterais dos aminoácidos estão projetadas para fora da hélice e, evidentemente, não participam das pontes de hidrogênio, estabelecidas unicamente entre os grupamentos das unidades peptídicas. Por esta razão, muitas sequências diferentes de aminoácidos podem adotar esta configuração e sua estabilidade independe do tipo de cadeia lateral, mas até certo ponto — certas sequências de aminoácidos não podem organizar-se em  $\alpha$ -hélice. Isto ocorre, por exemplo, quando a cadeia polipeptídica contém vários aminoácidos adjacentes de mesma carga, que se repelem fortemente. Assim, polilisina em solução a pH 7 não forma  $\alpha$ -hélice, pois apresenta as cadeias laterais carregadas positivamente; em pH 12, contudo, a maioria das cadeias laterais está desprotonada e a polilisina forma  $\alpha$ -hélice espontaneamente. A prolina, quando participa da ligação peptídica, não apresenta o átomo de hidrogênio no grupo imino e, portanto, não pode formar ligação de hidrogênio — este aminoácido é geralmente encontrado entre segmentos em  $\alpha$ -hélice.

A *folha  $\beta$ pregueada* ou *conformação  $\beta$*  é uma estrutura também mantida por ligações de hidrogênio entre as unidades peptídicas. Neste caso, entretanto, as ligações são estabelecidas entre cadeias polipeptídicas diferentes ou entre segmentos distantes de uma mesma cadeia. Na folha  $\beta$  pregueada, as cadeias apresentam uma conformação mais distendida que na  $\alpha$ -hélice e dispõem-se lado a lado, o que atribui a essa estrutura o aspecto de uma folha de papel pregueada. As ligações de hidrogênio são perpendiculares ao eixo das cadeias, e os grupos R dos aminoácidos projetam-se para cima e para baixo do

plano da folha pregueada (Figura 2.6).

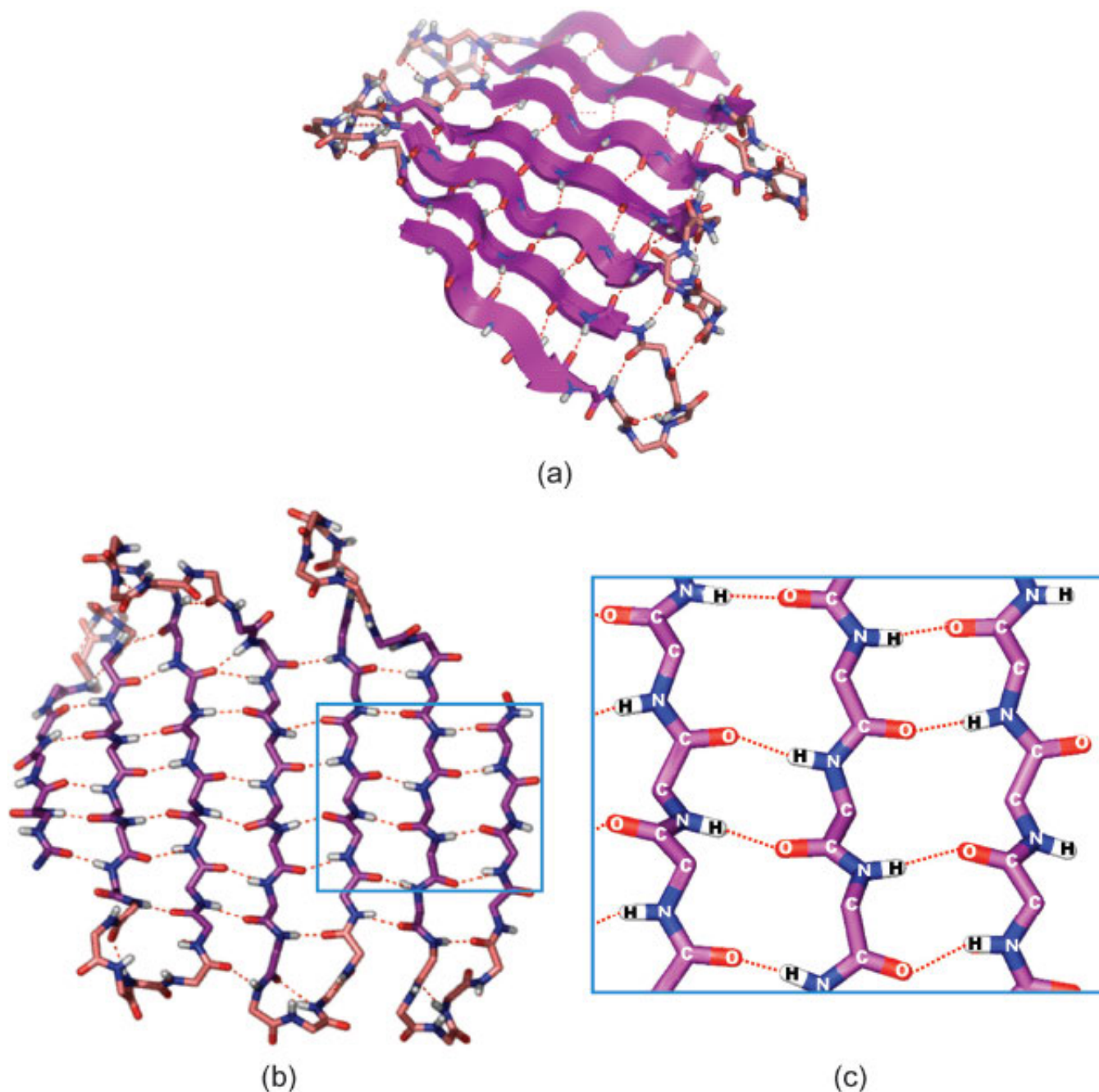


**Figura 2.5** Modelo da  $\alpha$ -hélice: a cadeia polipeptídica forma uma espiral, estabilizada por pontes de H entre os grupos  $-C=O$  e  $-NH$  das ligações peptídicas. As cadeias laterais dos resíduos de aminoácidos dispõem-se no exterior da hélice. (Adaptada de Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J: *Molecular Cell Biology*, 5th ed. W. H. Freeman and Company, 2004.)

Os dois tipos principais de estruturas secundárias regulares ( $\alpha$ -hélice e folha  $\beta$  pregueada) ocorrem nas proteínas em proporções muito diversas. Um exemplo extremo é a *mioglobina* — uma proteína de músculo transportadora de oxigênio — que apresenta cerca de 80% da cadeia polipeptídica organizada em  $\alpha$ -hélice. Sua molécula é formada por oito segmentos em  $\alpha$ -hélice, separados por trechos sem estrutura regular, que permitem o dobramento da cadeia (“cotovelos”), alguns deles resultantes da presença de prolina (Figura 2.7). No outro extremo aparecem proteínas como a *concanavalina A*<sup>4</sup>, que tem alto conteúdo de folha  $\beta$  pregueada e não forma  $\alpha$ -hélice (Figura 2.8). A maioria das proteínas exibe os dois tipos de estrutura secundária, como acontece com a *toxina diftérica* (Figura 2.9), produzida por uma bactéria que infecta o trato respiratório superior de seres humanos.

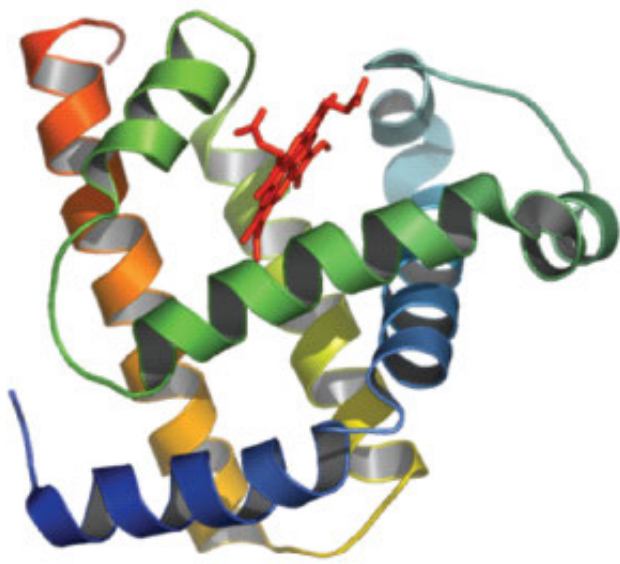
Cada proteína tem um conteúdo próprio de  $\alpha$ -hélice e de folha  $\beta$  pregueada, determinado pela sua estrutura primária. No entanto, um exemplo dramático e excepcional de alteração da estrutura de proteínas acontece no caso do *prion* (do inglês *prion*, derivado de *proteinaceous infectious particle*). Trata-se de uma partícula infecciosa constituída apenas de proteína — diferindo, portanto, dos agentes infecciosos conhecidos, capazes de se replicar porque contêm DNA ou RNA, como protozoários, bactérias e vírus —, responsável por encefalopatias transmissíveis em mamíferos. Alguns exemplos são: *encefalopatia espongiforme bovina* (*doença da vaca louca*), *scrapie*, uma moléstia semelhante que afeta ovelhas e a doença de Creutzfeldt-Jakob, em seres humanos. A incidência da moléstia no rebanho bovino atingiu proporções epidêmicas na Inglaterra na década de 1990, devido ao hábito de alimentar o gado com rações preparadas com carne e ossos derivados, inadvertidamente, de animais infectados. Estes animais apresentavam parte das moléculas da proteína *PrP* (de *Prion Protein*) com estrutura alterada; esta proteína, na sua forma normal, ocorre no cérebro de animais saudáveis e sua função ainda é desconhecida. As moléculas modificadas da proteína *PrP* são os príons, capazes de converter a proteína *PrP* normal em novos príons. As mudanças consistem em aumento do conteúdo de folha  $\beta$  pregueada e redução daquele de  $\alpha$ -

hélice, além de agregação das moléculas. Resultam os sintomas característicos — disfunções motoras e extensa degeneração do tecido nervoso — e a morte do animal infectado. A transmissão das encefalopatias causadas por príons ocorre interespecíes, como ficou demonstrado com a identificação de uma variante da doença de Creutzfeldt-Jakob, uma moléstia de ocorrência rara, em seres humanos que consumiram a carne de gado infectado. Outras doenças neurodegenerativas mais comuns, como as doenças de Alzheimer e Parkinson, estão igualmente associadas a agregados de proteínas anormais.



**Figura 2.6** Folha  $\beta$  pregueada. a) Esquema de parte da molécula de uma proteína — os segmentos da cadeia polipeptídica com este tipo de estrutura secundária são simbolizados por setas onduladas que apontam na direção amino terminal  $\rightarrow$  carboxila terminal. b) Representação plana dos dobramentos da cadeia polipeptídica e da disposição paralela dos diversos segmentos que, associados por ligações de H intracadeia, formam a folha  $\beta$  pregueada. c) Detalhamento mostrando os grupos que estabelecem as ligações de H. Não estão representadas as cadeias laterais dos aminoácidos.



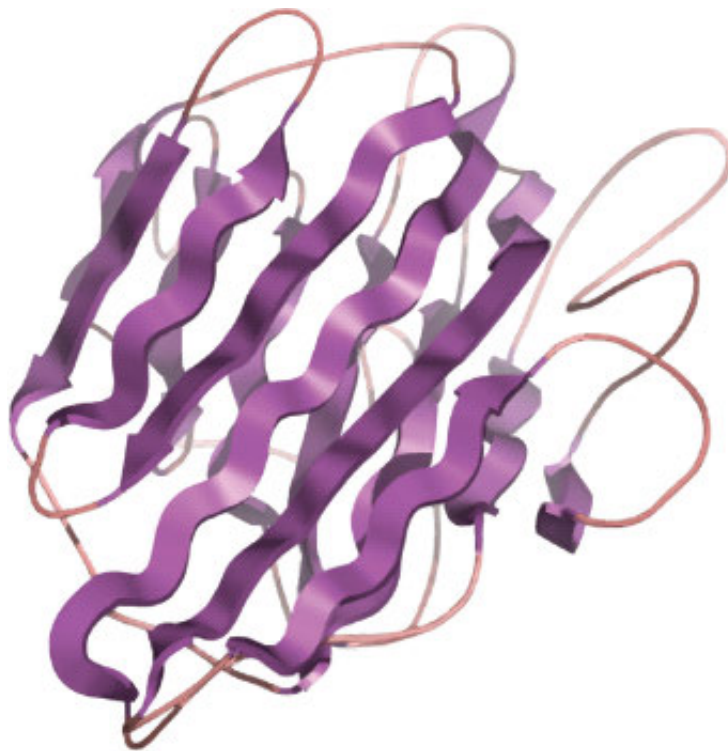


(a)

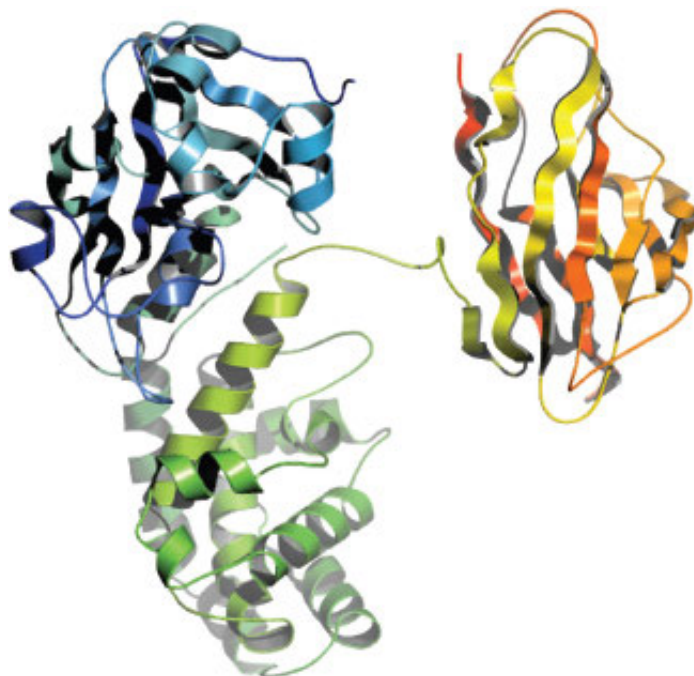


(b)

**Figura 2.7** Modelos da mioglobina mostrando: os diversos trechos em  $\alpha$ -hélice (representados por espirais), alternados por segmentos desenrolados (a); os dobramentos da cadeia da mioglobina, onde as esferas representam o carbono  $\alpha$  dos resíduos de aminoácidos (b). A cadeia polipeptídica liga-se ao grupo heme — vermelho em (a) e preto em (b) —, descrito no Capítulo 3.



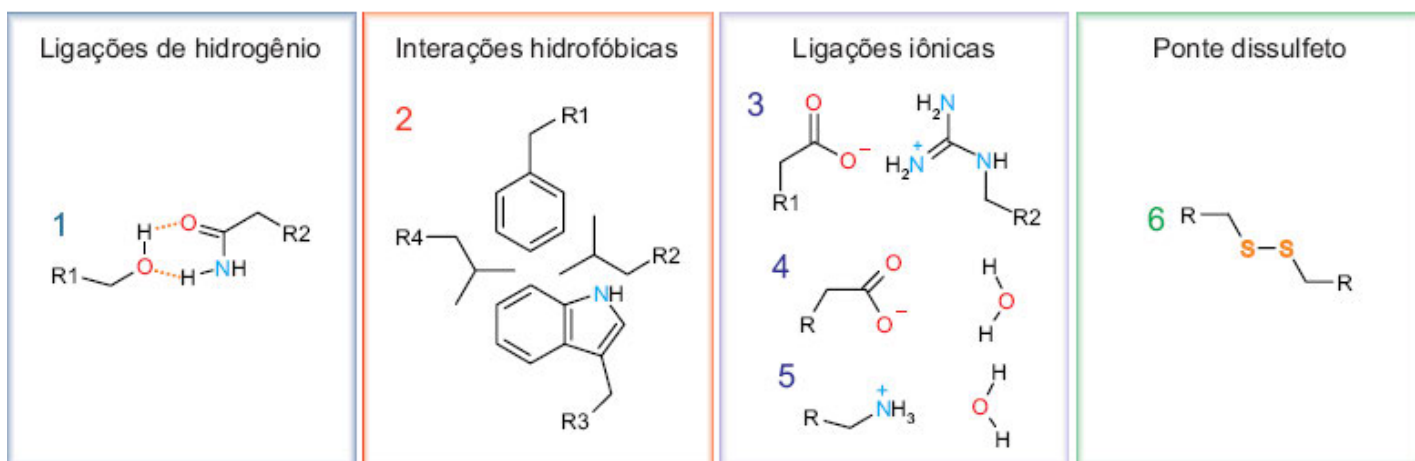
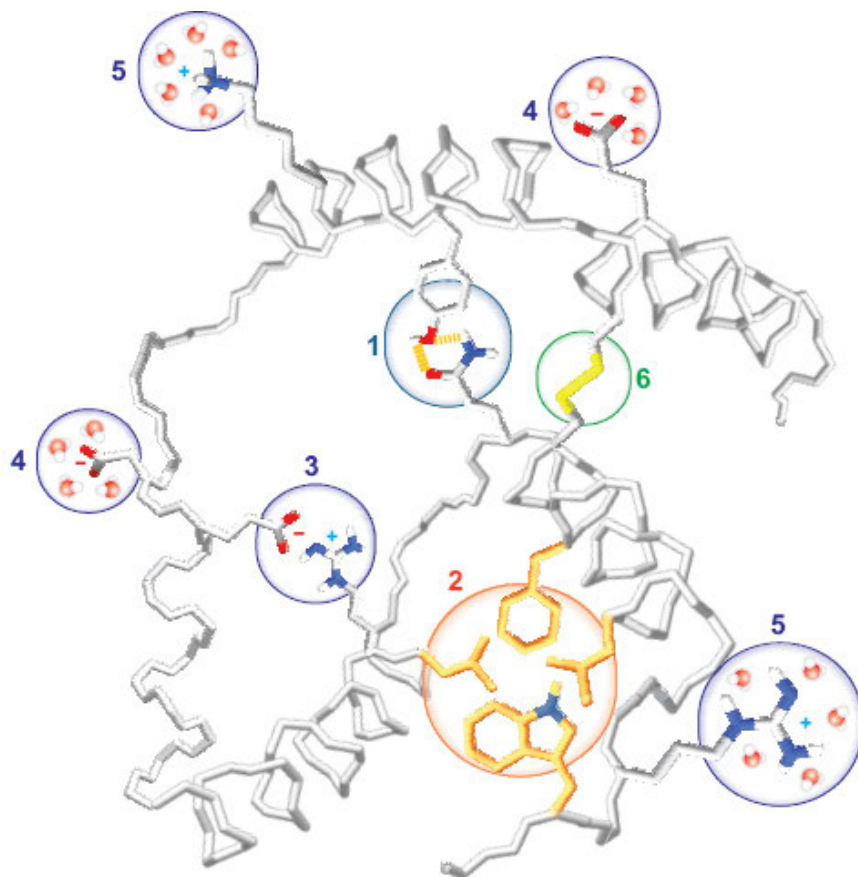
**Figura 2.9** Estrutura da toxina diftérica, que apresenta segmentos em  $\alpha$ -hélice, em folha  $\beta$  pregueada e sem estrutura regular, organizados em três domínios, representados em cores diferentes.



**Figura 2.8** Concanavalina A, uma proteína que se organiza, predominantemente, em folha  $\beta$  pregueada.

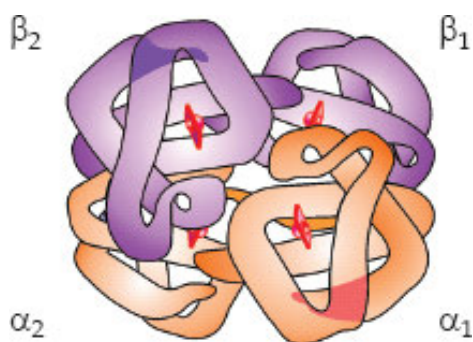
A **estrutura terciária** descreve o dobramento final da cadeia polipeptídica por interação de regiões com estrutura regular ( $\alpha$ -hélice ou folha  $\beta$  pregueada) ou de regiões sem estrutura definida. Neste nível de organização (Figuras 2.7 a 2.9), segmentos distantes da estrutura primária podem se aproximar e interagir, por intermédio de ligações não covalentes entre as cadeias laterais dos resíduos de aminoácidos. Estas ligações são consideradas fracas ( $4$  a  $30 \text{ kJ} \cdot \text{mol}^{-1}$ ), quando comparadas a ligações covalentes ( $200 \text{ kJ} \cdot \text{mol}^{-1}$ ). Como ocorre com as ligações de hidrogênio da estrutura secundária, é o grande número de ligações individualmente fracas que permite a manutenção dos dobramentos da estrutura terciária das proteínas. Estas ligações podem ser de diferentes tipos (Figura 2.10): ligações de hidrogênio, interações hidrofóbicas, ligações iônicas ou salinas e forças de van der Waals.

- *Ligações de hidrogênio.* Estabelecidas entre grupos R de aminoácidos polares com ou sem carga. Por exemplo, serina e treonina, que apresentam grupo hidroxila, podem formar ligações de hidrogênio com asparagina ou glutamina, que apresentam grupo carbonila. As ligações de hidrogênio da estrutura terciária, naturalmente, não apresentam um padrão regular de disposição, ao contrário do que ocorre com as ligações de hidrogênio da estrutura secundária, com as quais não devem ser confundidas.
- *Interações hidrofóbicas.* Formadas entre as cadeias laterais hidrofóbicas dos aminoácidos apolares. Estas cadeias não interagem com a água e aproximam-se, reduzindo a área apolar exposta ao solvente. As interações hidrofóbicas não resultam de qualquer atração entre os grupos apolares, mas são consequência da presença da molécula proteica no ambiente aquoso celular — a maioria das cadeias hidrofóbicas localiza-se no interior da molécula proteica. As interações hidrofóbicas são as mais importantes para a manutenção da conformação espacial das proteínas, dado o grande número (nove) de aminoácidos hidrofóbicos.
- *Ligações iônicas ou salinas.* Incluem interações de grupos com cargas opostas, como os presentes nos aminoácidos básicos (lisina, arginina e histidina) e ácidos (aspartato e glutamato). A energia de formação das ligações iônicas tem magnitude semelhante à das ligações dos grupos iônicos com a água, não contribuindo, portanto, para a conformação da molécula proteica quando estão localizados na sua superfície. Estas ligações, entretanto, têm importância fundamental para o dobramento da cadeia polipeptídica quando ocorrem no interior apolar da proteína. Todavia, esta situação não é muito frequente: a maioria dos grupos carregados de uma proteína localiza-se em sua superfície, estabelecendo interações íon-dipolo com a água, que forma uma camada organizada em volta da molécula proteica, a *camada de solvatação*.
- *Forças de van der Waals.* São a resultante das forças de atração e repulsão entre partes de moléculas. Inclui as forças entre dipolos permanentes e dipolos induzidos, encontrados nas cadeias laterais dos aminoácidos.



**Figura 2.10** Ligações da estrutura terciária de uma proteína globular: ligações não covalentes — ligações de hidrogênio (1), interações hidrofóbicas (2) e ligações iônicas (3) — e uma ligação covalente, a ponte dissulfeto (6). Estão mostradas, ainda, as ligações iônicas entre cadeias laterais dos aminoácidos com carga e dipolos da água (4, 5).

Além das ligações não covalentes, a estrutura proteica pode ser estabilizada por uma ligação covalente (Figura 2.10), a *ponte dissulfeto* ( $-S-S-$ ), formada entre dois resíduos de cisteína por uma reação de oxidação catalisada por enzimas específicas. Pontes dissulfeto são raramente encontradas em proteínas intracelulares, sendo mais frequentes em proteínas secretadas para o meio extracelular. Este é o caso da insulina, que apresenta três pontes dissulfeto (Figura 19.10 — Seção 19.4).





**Figura 2.11** A estrutura quaternária da hemoglobina consiste na associação de duas cadeias  $\alpha$  e duas cadeias  $\beta$ , cada uma associada a um grupo heme (em vermelho).

A **estrutura quaternária** descreve a associação de duas ou mais cadeias polipeptídicas (subunidades), para compor uma proteína funcional. A estrutura quaternária é mantida geralmente por ligações não covalentes entre as subunidades, dos mesmos tipos que mantêm a estrutura terciária. As subunidades que constituem uma proteína podem ser iguais ou diferentes. A molécula de *hemoglobina*, por exemplo, é formada por quatro cadeias polipeptídicas, iguais duas a duas, chamadas  $\alpha$  e  $\beta$ , associadas sobretudo por interações hidrofóbicas, com contribuição menor de ligações de hidrogênio e interações eletrostáticas (Figura 2.11).

### A estrutura terciária pode conter domínios e motivos

A estrutura terciária pode apresentar padrões de elementos estruturais, que se repetem em proteínas diferentes, chamados de domínios e motivos.

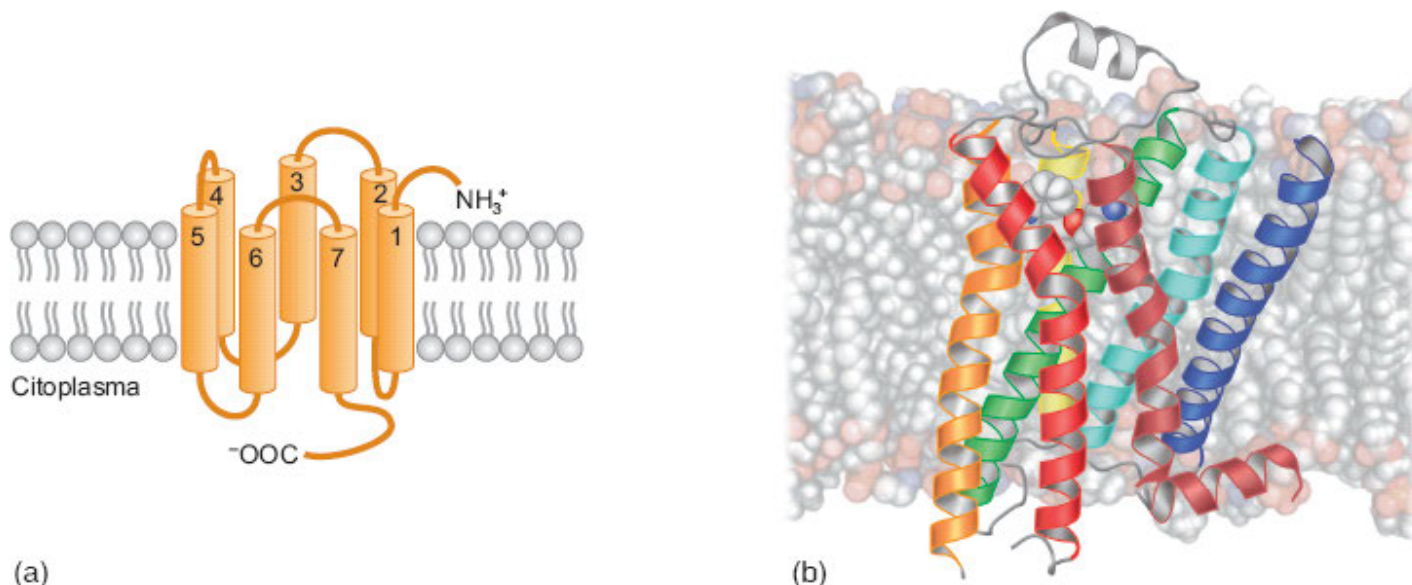
*Domínios* são regiões diferenciadas da molécula proteica, com organização espacial compacta; cada domínio é um conjunto estrutural definido, formado por dobramentos da cadeia polipeptídica. Geralmente, cadeias polipeptídicas longas, com centenas de resíduos de aminoácidos, são as que se organizam em domínios.

O grau de interação entre domínios pode variar desde domínios ligados por um segmento flexível da cadeia polipeptídica (Figura 2.9), até aqueles que estabelecem um contato muito íntimo, separados apenas por uma fenda estreita. Em qualquer um dos casos, os domínios podem movimentar-se, uns em relação aos outros. Esta flexibilidade é fundamental para que a molécula de proteína possa ligar-se eficientemente a outros compostos. Em muitas enzimas, a ligação com a molécula de substrato ocorre em fendas situadas entre domínios; estes se aproximam, encaixando o substrato na molécula da enzima (Figura 5.4, Seção 5.1).

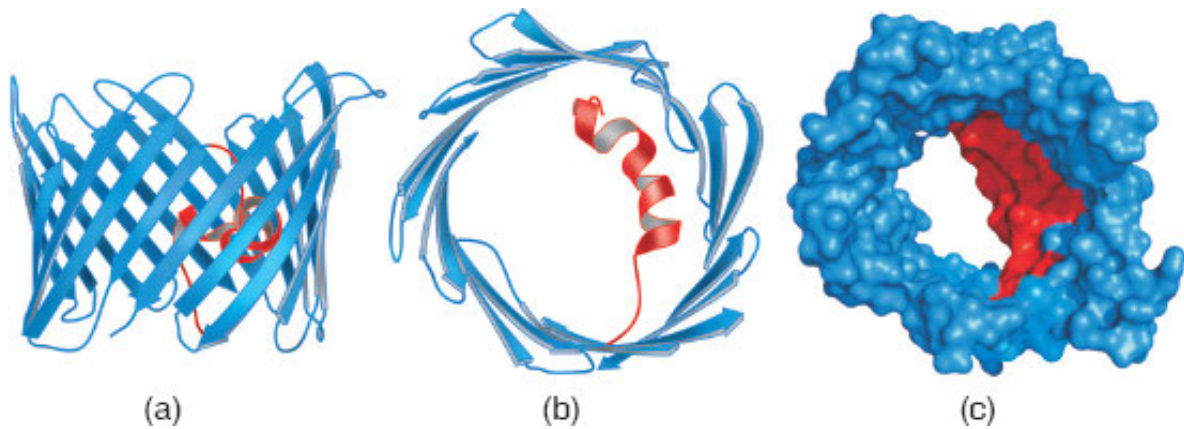
Os domínios frequentemente desempenham ações específicas; em inúmeras reações do metabolismo, o substrato liga-se a um dos domínios da enzima e a coenzima a outro. Em proteínas diferentes, domínios com a mesma função têm estruturas semelhantes, o que permite prever a atividade de uma proteína desconhecida a partir do conhecimento de seus domínios.

*Motivos* são diferentes formas de organização de elementos da estrutura secundária de proteínas globulares. Cada motivo tem um padrão de dobramento característico, que envolve interação entre segmentos da cadeia polipeptídica em  $\alpha$ -hélice e/ou folha  $\beta$  pregueada. Os mesmos motivos se repetem em proteínas de origens muito diferentes.

Os motivos podem ser constituídos por arranjos de  $\alpha$ -hélices, folhas  $\beta$  pregueadas ou por combinações das duas. Numerosos receptores da superfície celular, por exemplo, são compostos por sete  $\alpha$ -hélices que atravessam a membrana plasmática (Figura 2.12); são responsáveis por receber os sinais que iniciam o processo de transdução, envolvido em fenômenos tão diversos quanto a visão, o paladar, o olfato e a atividade hormonal (Seção 19.3). Outro motivo complexo, chamado  $\beta$  barril, resulta da associação de numerosos segmentos em folha  $\beta$  pregueada. É encontrado na família das *porinas* (Figura 2.13), que formam canais na membrana externa de bactérias gram-negativas e de mitocôndrias, destinados ao transporte de íons e moléculas pequenas, como nucleosídeos ou açúcares.



**Figura 2.12** Estrutura de um receptor com sete segmentos em  $\alpha$ -hélice (numeradas de 1 a 7), que atravessam a membrana plasmática. a) Representação esquemática. b) Estrutura tridimensional do receptor adrenérgico  $\beta_2$ .



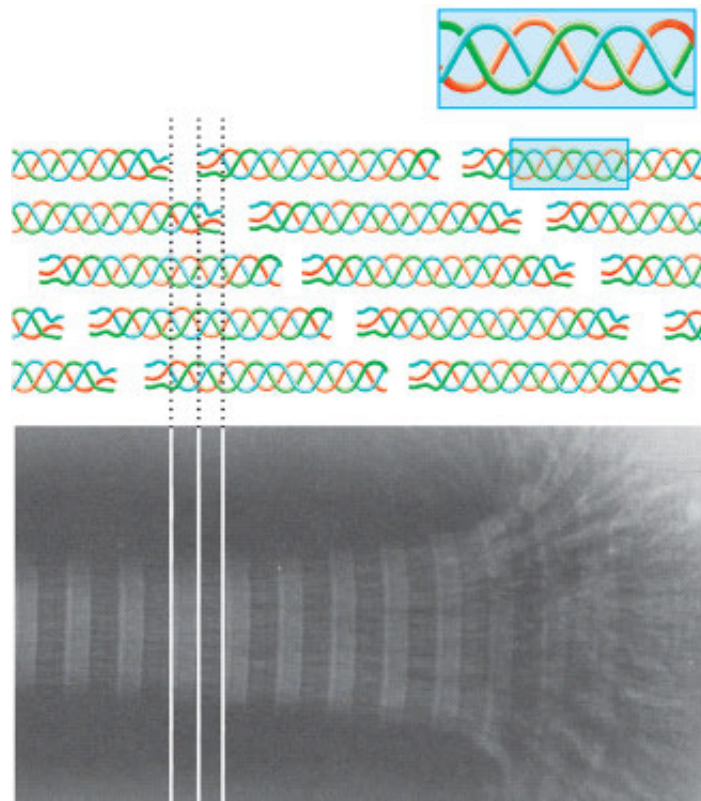
**Figura 2.13** Estrutura de uma porina mitocondrial, o canal iônico voltagem-dependente *mVDAC1*, contendo um motivo em  $\beta$  barril (em azul) e uma hélice na extremidade amino terminal (em vermelho). Vista lateral (a) e superior (b) da proteína na representação em fitas; no centro, vista superior da representação de volumes atômicos (c).

## 2.5 Proteínas fibrosas

As proteínas fibrosas têm forma alongada e, diferentemente das globulares, são formadas pela associação de módulos repetitivos, possibilitando a construção de grandes estruturas. O componente fundamental das proteínas fibrosas são cadeias polipeptídicas muito longas com estrutura secundária regular:  $\alpha$ -hélice nas  $\alpha$ -queratinas, folha  $\beta$  pregueada nas  $\beta$ -queratinas e uma hélice característica no colágeno.

Nas  $\alpha$ -queratinas, duas ou três cadeias em  $\alpha$ -hélice associam-se lateralmente, formando longos cabos helicoidais, que, reunidos, formam fibrilas e fibras. As  $\alpha$ -queratinas são o componente principal da pele dos vertebrados e de estruturas relacionadas, como cabelo, lã, chifres, unhas, cascos, bicos e penas. Nestas proteínas, são frequentes as pontes dissulfeto entre resíduos de cisteína de cadeias polipeptídicas ou fibrilas adjacentes, conferindo grande resistência às fibras. O padrão de distribuição dessas pontes determina o grau de ondulação do cabelo e da lã. Os tratamentos estéticos, visando tornar o cabelo mais liso ou mais crespo, utilizam o mesmo procedimento, embora pretendam resultados opostos: desfazer as pontes dissulfeto por tratamento com agentes redutores e refazê-las em novas posições por tratamento com agentes oxidantes, depois de posicionar os fios de cabelo na conformação desejada. A tropomiosina, uma proteína componente das miofibrilas também é formada por associação de cadeias polipeptídicas em  $\alpha$ -hélice. Nas  $\beta$ -queratinas, as fibras são formadas por empilhamento de folhas  $\beta$  pregueadas, como acontece na fibroína da seda e das teias de aranha.

No caso do colágeno, as cadeias polipeptídicas apresentam uma conformação helicoidal típica, derivada da sua composição peculiar em aminoácidos — alto conteúdo de glicina, prolina e de hidroxiprolina, um aminoácido derivado de prolina — e da grande regularidade na estrutura primária, sendo frequente a sequência glicina-prolina-hidroxiprolina. Estas características, diferentes da maioria das proteínas, permitem a associação íntima de três cadeias formando uma hélice tripla, o tropocolágeno, módulo estrutural básico do colágeno. As moléculas de tropocolágeno associam-se, formando as fibrilas de colágeno, que são estabilizadas por ligações covalentes entre as cadeias componentes do tropocolágeno e entre moléculas de tropocolágeno adjacentes (Figura 2.14). O colágeno é a proteína mais abundante dos vertebrados. Suas fibras são responsáveis pelas funções mecânicas e de sustentação do tecido conjuntivo, que se distribui por cartilagens, tendões, matriz óssea, córnea etc.; mantém, ainda, a estrutura e a elasticidade do sistema vascular e de todos os órgãos. O número de ligações covalentes do colágeno varia conforme o tecido e aumenta com a idade do animal (o que explica a maior rigidez da carne de animais mais velhos). A estrutura do colágeno é rompida por aquecimento, originando uma proteína desenrolada, mais solúvel, a gelatina. Este é o princípio da fabricação industrial desta proteína, muito frequente na dieta humana. O baixo valor nutricional da gelatina está analisado no Capítulo 18.



**Figura 2.14** Associação de moléculas de tropocolágeno para formar fibrilas de colágeno: as moléculas ficam deslocadas umas em relação às outras, o que atribui, à fibrila de colágeno, um aspecto estriado ao microscópio eletrônico. As ligações covalentes que estabilizam o colágeno foram omitidas.

## 2.6 Proteínas conjugadas

### As proteínas podem apresentar aminoácidos modificados e componentes não proteicos

Muitas proteínas contêm resíduos de determinados aminoácidos modificados, além dos 20 usuais. Esses aminoácidos exóticos não são codificados nos RNAs mensageiros; formam-se por alteração enzimática de aminoácidos comuns, após serem incorporados na cadeia polipeptídica que está sendo sintetizada.

No colágeno, é frequente a hidroxilação da cadeia lateral de prolina. A reação enzimática que produz esta modificação requer a presença de *ácido ascórbico (vitamina C)*. No homem, a deficiência de ácido ascórbico na dieta causa o *escorbuto*. As manifestações clínicas desta doença correlacionam-se com a síntese de moléculas de colágeno menos estáveis nos diferentes tecidos: interrupção do crescimento dos ossos em crianças, má cicatrização de ferimentos e aumento da fragilidade de vasos sanguíneos, resultando em hemorragias na pele e gengivas, principalmente.

Em outras proteínas, pode ocorrer acetilação do grupo amino de lisina ou fosforilação do grupo hidroxila de serina, treonina e tirosina. A adição e a remoção de grupos fosfato é um fenômeno cíclico frequentemente encontrado em proteínas com função enzimática. A fosforilação altera profundamente a atividade destas enzimas (Seção 19.2.2).

As proteínas podem, ainda, apresentar moléculas orgânicas não proteicas, ligadas à cadeia polipeptídica. Estes componentes são designados *grupos prostéticos* e as proteínas, neste caso, são chamadas *proteínas conjugadas*. O grupo prostético é de natureza variável, podendo ligar-se à cadeia polipeptídica covalente ou não covalentemente. As *hemoproteínas*, como mioglobina, hemoglobina, catalase e citocromos, apresentam um grupo prostético denominado *heme*. Na mioglobina (Figura 2.7), o heme liga-se não covalentemente à cadeia polipeptídica, o mesmo acontecendo com cada uma das quatro subunidades da hemoglobina (Figura 2.11); a estrutura do grupo heme está mostrada na Figura 3.2. O grupo prostético pode ser um carboidrato ou um lipídio, covalentemente ligados, e a proteína conjugada chama-se glicoproteína ou lipoproteína, respectivamente.

*Glicoproteínas* são encontradas em todos os compartimentos celulares, mas constituem, principalmente, as proteínas secretadas pelas células e aquelas localizadas na sua superfície externa. Exemplos de glicoproteínas secretadas são as mucinas das secreções mucosas e muitas proteínas do sangue, como as que participam da coagulação sanguínea e as imunoglobulinas. As glicoproteínas da membrana plasmática funcionam como marcadores biológicos, que permitem a comunicação entre as células (Seção 7.3). Outras proteínas importantes como o interferon e alguns hormônios também são glicoproteínas; até o colágeno contém pequenas porcentagens de açúcares covalentemente ligados. A hemoglobina, em determinadas condições, pode ligar-se à glicose, formando a *hemoglobina glicada* (Seção 3.5).

Quanto às *lipoproteínas*, devem-se distinguir as proteínas que apresentam algumas moléculas de lipídios

covalentemente ligados das *lipoproteínas plasmáticas*. As primeiras são proteínas conjugadas no sentido estrito, como as lipoproteínas da parede celular de certas bactérias que contêm moléculas de ácidos graxos unidos por ligações covalentes. As lipoproteínas plasmáticas (Seção 6.2.7), por sua vez, são partículas formadas por inúmeras moléculas de lipídios e algumas poucas moléculas de proteína, associadas por ligações não covalentes. Estas partículas atuam no transporte de lipídios pelo sangue (Seção 20.8).

## 2.7 Carga elétrica e solubilidade das proteínas

### O valor do pI de uma proteína reflete a sua composição em aminoácidos

A carga elétrica total de uma proteína é o somatório das cargas presentes nas cadeias laterais dos aminoácidos e nos grupos amino e carboxila terminais.

Para cada proteína existe um determinado valor de pH — *ponto isoelétrico (pI)* — no qual a molécula é eletricamente neutra. Neste pH, o número de cargas positivas (grupos básicos protonados) equivale ao número de cargas negativas (grupos ácidos desprotonados).

O pI de proteínas, diferentemente do pI dos aminoácidos, não pode ser calculado, de modo simples, a partir dos valores de  $pK_a$  dos aminoácidos componentes, devido ao seu grande número e, principalmente, porque o valor de  $pK_a$  dos aminoácidos varia conforme a sua localização na estrutura da proteína (Tabela 2.1). Programas de computador sofisticados permitem grande aproximação no cálculo do ponto isoelétrico de proteínas. O pI das proteínas é de fácil determinação experimental: é o valor de pH no qual elas não migram, quando submetidas a um campo elétrico (Seção 2.10.2).

As proteínas exibem valores de pI que refletem a proporção entre aminoácidos ácidos e básicos em sua composição (Tabela 2.5). Para a pepsina, por exemplo, que tem muito mais aminoácidos ácidos (28%), que aminoácidos básicos (2%), a equivalência de cargas é obtida quando a grande maioria das carboxilas dos aminoácidos ácidos está protonada (sem carga) e apenas uma pequena fração desprotonada (com carga negativa), compensando a carga positiva dos aminoácidos básicos. Esta situação é encontrada em um pH muito baixo: o pI da pepsina é igual a 1. No caso do citocromo *c*, cuja porcentagem de aminoácidos básicos é aproximadamente o dobro daquela de aminoácidos ácidos, o pI é o pH em que cerca da metade dos aminoácidos básicos estão desprotonados, ou seja, um pH alto: o pI do citocromo *c* é igual a 10,6. Proteínas com pI maior que 7 são chamadas *proteínas básicas* e aquelas com pI menor que 7, *proteínas ácidas*. As histonas são proteínas básicas, e no pH celular apresentam-se com carga positiva, o que permite sua ligação ao DNA na formação dos cromossomos.

O abaixamento de pH resulta sempre em aumento da carga positiva da proteína, por ganho efetivo de carga positiva (protonação de grupamentos básicos) ou perda de carga negativa (protonação de grupamentos ácidos); portanto, em pH menor do que o pI, a proteína apresenta uma carga líquida positiva, tanto maior quanto mais afastado do pI for o pH. Por um raciocínio análogo, conclui-se que, acima do pI, a proteína apresenta carga negativa porque grupos ácidos desprotonados adquirem carga negativa e grupos básicos desprotonados perdem carga positiva.

**Tabela 2.5 Ponto isoelétrico (pI) de algumas proteínas e sua composição em aminoácidos ácidos e básicos.**

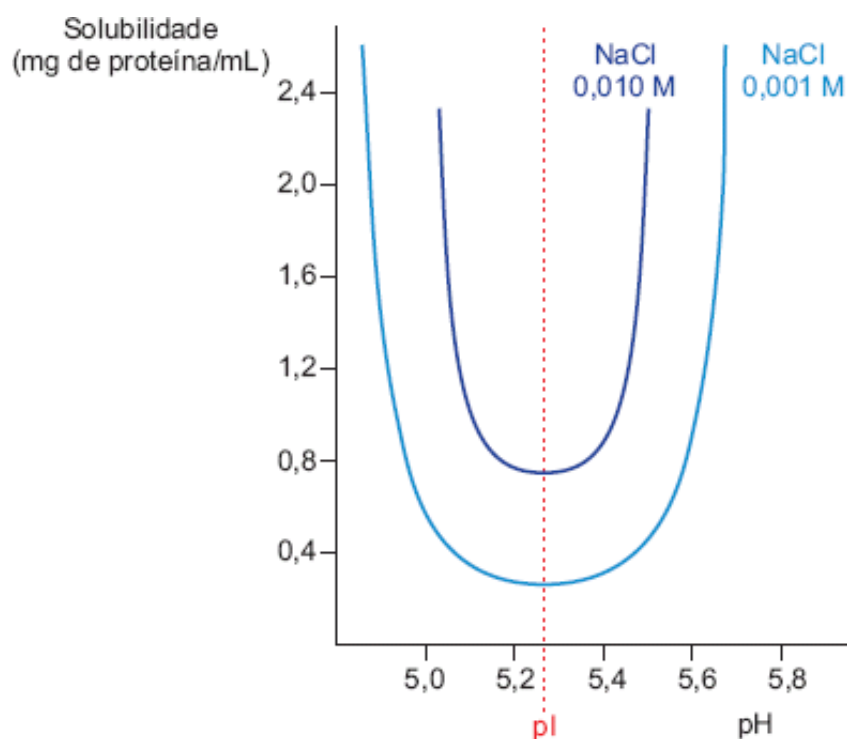
	Aminoácidos (%)						Ácidos Básicos
	pI	Ácidos		Básicos		Lys	
		Asp	Glu	Arg	His		
Pepsina	1,0	16,6	11,3	1,0	0,5	0,4	15
Albumina	4,8	10,4	17,4	6,2	3,5	12,3	1,3
Mioglobina	7,0	4,7	8,3	1,9	7,5	12,8	0,6
Citocromo <i>c</i>	10,6	3,6	5,9	2,2	2,5	15,2	0,5

### A solubilidade das proteínas é influenciada pela composição do meio aquoso

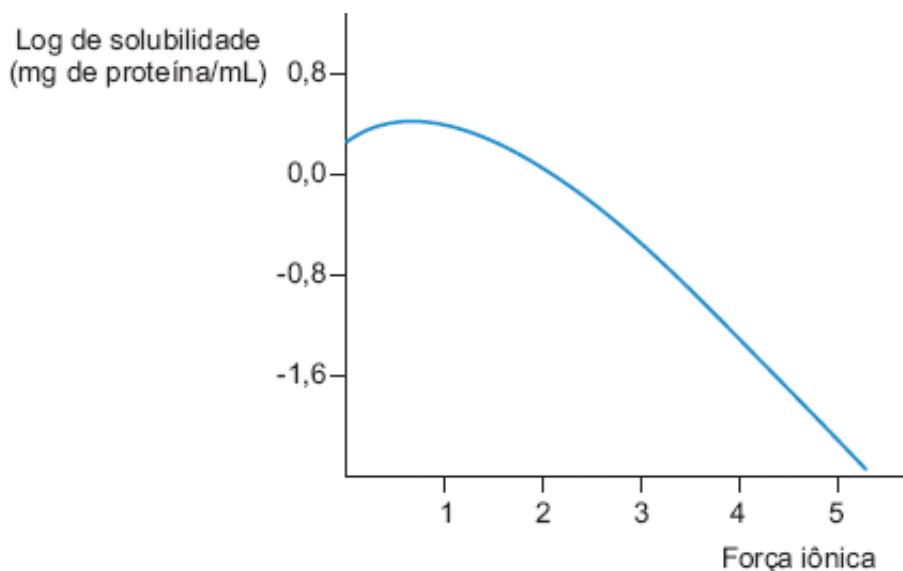
A solubilidade das proteínas é determinada, fundamentalmente, pela estrutura primária, que define a relação espacial entre os aminoácidos na estrutura tridimensional e sua interação com a água. Por outro lado, características do meio, tais como o pH, a concentração de sais e a constante dielétrica do solvente, interferem na solubilidade. A variação da carga líquida de uma proteína tem implicações na sua solubilidade. No pI a solubilidade é menor do que em outros valores de pH, nos quais as moléculas têm todas a mesma carga e se repelem eletrostaticamente, estabilizando-se em solução (Figura 2.15).



As proteínas apresentam também alteração da solubilidade em função da concentração de sais. Proteínas globulares pouco solúveis em água tornam-se cada vez mais solúveis à medida que aumenta a concentração de sal da solução (Figura 2.15), até certa concentração limitante, que depende da proteína e do tipo de sal escolhido (trecho ascendente da curva da Figura 2.16). Este fenômeno é chamado “*salting in*”. Acredita-se que os íons adicionais (positivos e negativos), presentes em solução, interagem com os grupos carregados das moléculas de proteína, atenuando a interação entre elas. Deste modo, o efeito eletrostático de íons em soluções salinas diluídas é um fator adicional para o aumento da solubilidade das proteínas, além da sua camada de solvatação<sup>5</sup>. Por outro lado, quando a concentração de sal atinge valores muito elevados, a solubilidade das proteínas diminui (trecho descendente da curva da Figura 2.16), até sua precipitação. Este efeito, chamado “*salting out*”, ocorre com sais di ou trivalentes, que competem com a proteína por moléculas de água para solvatação. Em altas concentrações desses sais, ocorre uma desorganização da camada de solvatação da proteína: há tantos íons solvatados que a quantidade de água disponível torna-se insuficiente para dissolver todos os solutos. As interações proteína-proteína tomam-se mais fortes que as interações proteína-solvente, a proteína sofre agregação e precipita. Como cada proteína precipita em uma concentração salina característica (que depende da extensão da sua camada de solvatação), o *salting out* pode ser empregado para separar proteínas. De fato, esta técnica costuma ser a etapa inicial de processos de purificação de proteínas. O sal mais utilizado é o sulfato de amônio —  $(\text{NH}_4)_2\text{SO}_4$  — devido à sua alta solubilidade, que permite obter soluções muito concentradas; além disso, este sal, por razões desconhecidas, estabiliza a estrutura nativa das proteínas, possibilitando que elas precipitem sem sofrer desnaturação.



**Figura 2.15** Solubilidade de uma proteína globular em função do pH, em duas concentrações de NaCl. Está assinalado o pH correspondente ao pI da proteína.



**Figura 2.16** Solubilidade da hemoglobina equina, em função da concentração de  $(\text{NH}_4)_2\text{SO}_4$ , expressa em termos da força iônica da solução.

Solventes orgânicos solúveis em água, como acetona e etanol, diminuem a solubilidade das proteínas, devido ao baixo valor de suas constantes dielétricas e porque eles também sofrem hidratação. Em soluções aquosas desses compostos, moléculas contendo grupos com carga, como as proteínas, sofrem menor hidratação e interagem tão fortemente que precipitam. Procedimentos experimentais que utilizam a adição de solventes orgânicos para precipitar proteínas devem ser realizados a  $0^\circ\text{C}$ , porque, em temperaturas mais altas, eles provocam a desnaturação das proteínas.

## 2.8 Alterações estruturais das proteínas

### A desnaturação da proteína consiste na perda de sua estrutura original

À medida que uma proteína é sintetizada, sua estrutura primária dobra-se espontaneamente, originando as estruturas secundária e terciária. Se a proteína em questão possuir estrutura quaternária, esta também se organiza espontaneamente, assim que a estrutura terciária das subunidades componentes é formada. A proteína assume a conformação denominada *nativa*. Esta é a conformação mais estável que a molécula pode assumir naquelas condições e reflete um equilíbrio delicado entre as interações ocorridas no interior da molécula proteica e entre esta e seu meio ambiente.

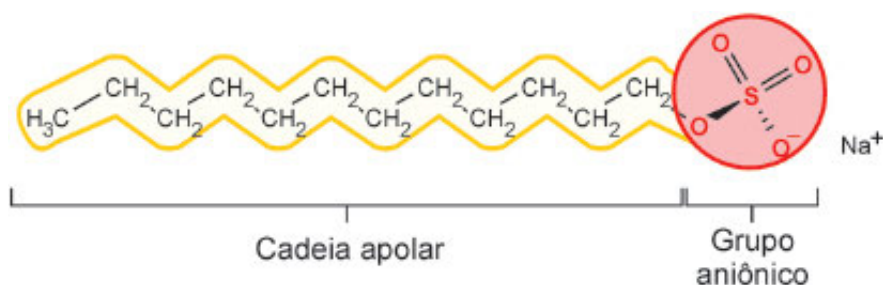
Ao proceder-se ao isolamento e purificação de uma proteína, a partir das células onde ela é encontrada, são introduzidas alterações físicas e químicas no seu meio ambiente, que podem afetar sua estrutura espacial a ponto de ocasionar a perda de sua função. A proteína é dita, então, *desnaturada*: sua conformação nativa é destruída devido ao rompimento de ligações não covalentes (as ligações peptídicas são mantidas) e o resultado é uma cadeia polipeptídica distendida.

A desnaturação pode ser provocada, experimentalmente, por tratamentos que ocasionam o rompimento de ligações não covalentes. O aquecimento a temperaturas altas, mas menores do que  $100^\circ\text{C}$ , provoca a desnaturação da maioria das proteínas. Há exceções, como as proteínas de *bactérias termofílicas* que habitam fontes termais, cujas temperaturas são próximas de  $100^\circ\text{C}$ . Uma destas proteínas, a DNA polimerase da bactéria *Thermus aquaticus*, a Taq polimerase, tem sido amplamente empregada na reação em cadeia da polimerase (PCR, de *Polymerase Chain Reaction*), um método laboratorial de amplificação de DNA. Graças à sua estabilidade térmica, a enzima mantém-se ativa a  $95^\circ\text{C}$ , temperatura em que as fitas de DNA separam-se, e podem, então, ser duplicadas pela ação da enzima. Um fato curioso é que a estrutura espacial dessas proteínas termoestáveis apresenta apenas diferenças sutis em relação à estrutura de proteínas com estabilidade térmica normal.

Outros agentes desnaturantes são os ácidos e os álcalis fortes: valores de pH muito baixos ou muito altos, afetando a ionização dos grupamentos da proteína, conferem à molécula uma elevada carga positiva, ou negativa, ocasionando repulsão intramolecular, com exposição do interior hidrofóbico. O tratamento com solventes orgânicos polares ou com compostos capazes de formar ligações de hidrogênio, como a ureia, determina a desnaturação da proteína; estes últimos compostos estabelecem ligações de hidrogênio com grupamentos da proteína, substituindo ligações que mantinham a estrutura nativa, e os solventes orgânicos por diminuir a constante dielétrica do meio.

A desnaturação também pode ser ocasionada por detergentes e sabões, que são compostos anfipáticos, ou seja, apresentam uma cadeia apolar longa, ligada a um grupo terminal carregado. Estes agentes são desnaturantes porque sua cauda hidrofóbica se introduz no interior da molécula proteica e se associa com radicais apolares, rompendo interações hidrofóbicas que mantêm a estrutura nativa. Um detergente frequentemente empregado é o dodecilsulfato de sódio (SDS) (Figura 2.17).

A desnaturação pode ser irreversível: algumas proteínas, quando desnaturadas, tomam-se insolúveis. É o caso da albumina do ovo, quando aquecida, e da caseína, quando o leite é acidificado, por exemplo, por crescimento bacteriano.



**Figura 2.17** Fórmula estrutural do detergente dodecilsulfato de sódio (SDS), ilustrando sua natureza anfipática.

## Proteínas desnaturadas podem renaturar

Retiradas as condições desnaturantes, algumas proteínas reassumem sua conformação nativa — este processo chama-se *renaturação*. A renaturação demonstra que a estrutura tridimensional de uma proteína é consequência de sua estrutura primária, ou seja, é determinada, unicamente, por sua sequência de aminoácidos.

Durante muito tempo, admitiu-se que o processo de enovelamento da cadeia polipeptídica recém-sintetizada nos ribossomos era autônomo, não requerendo nenhum fator adicional, nem fornecimento de energia. Esta visão passou a ser questionada diante da constatação de que a velocidade e a eficiência da renaturação *in vitro* são muito menores do que *in vivo*. Realmente, na década de 1990, demonstrou-se que as células dispõem de um aparato muito eficiente para assessorar a proteína durante a “montagem” de sua estrutura nativa. Este aparato é constituído por uma família de proteínas denominadas *chaperonas*<sup>6</sup>. Elas ligam-se a proteínas nascentes e, por meio de etapas cíclicas, sustentadas por hidrólise de ATP, impedem ou revertem interações inadequadas de regiões potencialmente complementares. Também facilitam o transporte de proteínas através de membranas desde o citosol até o destino final, seja a secreção ou a inclusão em uma organela.

As proteínas assessoras possibilitam, ainda, a estabilização de proteínas em condições desfavoráveis, como altas temperaturas etc. As chaperonas foram identificadas na década de 1970 em *Drosophila melanogaster* como proteínas de choque térmico, *Hsp* (de *Heat shock protein*), induzidas quando as células eram submetidas a altas temperaturas.

As proteínas assessoras são amplamente distribuídas na natureza, sendo encontradas desde em bactérias até em mamíferos. Distúrbios no processo de enovelamento da cadeia proteica, e a consequente agregação de moléculas, estão envolvidos em diversas condições patológicas, como as doenças de Alzheimer, de Parkinson, aquelas causadas por príons (Seção 2.4) etc.

### A substituição de aminoácidos pode alterar a função das proteínas

Uma mutação que resulte na substituição de um aminoácido em uma posição crítica na molécula da proteína pode ter consequências danosas para o desempenho da sua função. O exemplo clássico é a substituição, nas cadeias  $\beta$  da hemoglobina, de um resíduo de glutamato, cuja cadeia lateral polar negativa localiza-se na superfície externa da molécula, por valina, com grupo R apolar. As moléculas da hemoglobina substituída, quando desoxigenadas, agregam-se devido a ligações hidrofóbicas envolvendo as cadeias laterais apolares de valina. Os agregados formam um precipitado fibroso que distorce as hemácias, que adquirem forma de foice — *sickle*, em inglês — e, por isso, a hemoglobina alterada é chamada de *hemoglobina S*, em contraposição à hemoglobina normal, a *hemoglobina A*. Estas células deformadas obstruem os capilares, impedindo a oxigenação adequada dos tecidos; também são mais frágeis que as normais e sofrem hemólise facilmente (têm meia-vida de 16 a 20 dias em lugar dos 120 dias das hemácias normais), ocasionando anemia grave. A doença é conhecida como *anemia falciforme*, e manifesta-se somente quando a mutação ocorre em homozigose: indivíduos heterozigotos são normais. Outras hemoglobinas mutantes estão analisadas no próximo capítulo.

Término da leitura básica

## 2.9 Purificação de proteínas — Estratégia geral

A purificação de uma proteína inicia-se com a liberação da proteína do material biológico onde ela ocorre — órgãos, tecidos ou células isoladas como hemácias, bactérias e leveduras — pelo rompimento destas estruturas, que pode ser obtido por métodos diversos. A homogeneização mecânica em meio isotônico, por exemplo, produz o maceramento dos tecidos (ou órgãos) e a lise das células, originando um *extrato celular*, constituído por fragmentos de membrana plasmática, organelas (núcleos, mitocôndrias etc.) e o citosol. Os diversos componentes celulares podem ser separados por *fracionamento celular*, um procedimento que consiste em centrifugar o extrato celular em velocidades progressivamente maiores. Geralmente, quanto menor for uma estrutura, maior será a força centrífuga necessária para sedimentá-la; como os componentes celulares diferem em tamanho, eles sedimentarão em velocidades diferentes, podendo, assim, ser separados uns dos outros. Quando a proteína desejada localiza-se apenas em uma das frações obtidas — mitocôndrias, por exemplo — o fracionamento celular propicia uma purificação inicial considerável.

Uma vez conseguida uma preparação contendo a proteína, esta pode ser separada de outras proteínas e de outros tipos de moléculas por métodos que se baseiam em solubilidade, tamanho, carga elétrica ou afinidade por determinados compostos. O método de purificação a ser escolhido dependerá da particular proteína que se pretende isolar e, usualmente, empregam-se combinações sequenciais de diferentes métodos. Frequentemente, o primeiro passo empregado para a separação de proteínas de extratos brutos é a precipitação por adição de sais (sulfato de amônio é o mais comumente usado) ou solventes orgânicos miscíveis com água — a separação, neste caso, baseia-se em diferenças de solubilidade apresentadas pelas proteínas. Estas técnicas permitem uma purificação parcial e devem ser seguidas de outras, mais seletivas, como cromatografia e eletroforese.

A purificação da proteína deve ser monitorada ao longo das diferentes etapas de separação e, para isso, é necessário



# Aminoácidos, Peptídeos e Proteínas

- 3.1 Aminoácidos 76
- 3.2 Peptídeos e proteínas 85
- 3.3 Trabalhando com proteínas 89
- 3.4 A estrutura de proteínas: estrutura primária 96

**P**roteínas controlam praticamente todos os processos que ocorrem em uma célula, exibindo uma quase infinita diversidade de funções. Para explorar o mecanismo molecular de um processo biológico, um bioquímico estuda quase que inevitavelmente uma ou mais proteínas. Proteínas são as macromoléculas biológicas mais abundantes, ocorrendo em todas as células e em todas as partes das células. As proteínas também ocorrem em grande variedade; milhares de diferentes tipos podem ser encontrados em uma única célula. Como os árbitros da função molecular, as proteínas são os produtos finais mais importantes das vias de informação discutidas na Parte III deste livro. As proteínas são os instrumentos moleculares pelos quais a informação genética é expressa.

Subunidades monoméricas relativamente simples fornecem a chave da estrutura de milhares de proteínas diferentes. As proteínas de cada organismo, da mais simples das bactérias aos seres humanos, são construídas a partir do mesmo conjunto onipresente de 20 aminoácidos. Como

cada um desses aminoácidos tem uma cadeia lateral com propriedades químicas características, esse grupo de 20 moléculas precursoras pode ser considerado o alfabeto no qual a linguagem da estrutura proteica é lida.

Para gerar uma determinada proteína, os aminoácidos se ligam de modo covalente em uma sequência linear característica. O mais marcante é que as células produzem proteínas com propriedades e atividades completamente diferentes ligando os mesmos 20 aminoácidos em combinações e sequências muito diferentes. A partir desses blocos de construção, diferentes organismos podem gerar produtos tão diversos como enzimas, hormônios, anticorpos, transportadores, fibras musculares, proteínas das lentes dos olhos, penas, teias de aranha, chifres de rinocerontes, proteínas do leite, antibióticos, venenos de cogumelos e uma miríade de outras substâncias com atividades biológicas distintas (**Figura 3-1**). Entre esses produtos de proteínas, as enzimas são as mais variadas e especializadas. Como catalisadoras de quase todas as reações celulares, as enzimas são uma das chaves para compreensão da química da vida e, assim, fornecem um ponto central para qualquer curso de bioquímica.

Estruturas e funções de proteínas são os tópicos deste e dos próximos três capítulos. Aqui, primeiro é feita uma descrição das propriedades químicas fundamentais dos aminoácidos, peptídeos e proteínas. Também é abordado como um bioquímico trabalha com proteínas.



**FIGURA 3-1 Algumas funções de proteínas.** (a) A luz produzida por vaga-lumes é o resultado de uma reação envolvendo a proteína luciferina e ATP, catalisada pela enzima luciferase (ver Quadro 13-1). (b) Eritrócitos contêm grandes quantidades da proteína transportadora de oxigênio hemoglobina. (c) A proteína queratina, produzida por todos os vertebrados, é o componente estrutural principal de pelos, escamas, chifres, lâ, unhas e penas. O

rinoceronte preto está próximo da extinção em ambiente natural devido à crença encontrada em algumas partes do mundo de que o pó do seu chifre tem propriedades afrodisíacas. Na verdade, as propriedades químicas do pó de chifre de rinoceronte não são diferentes daquelas do pó dos cascos de bovinos e das unhas humanas.

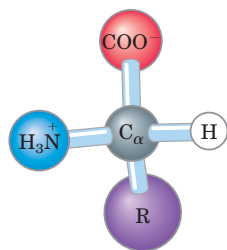
### 3.1 Aminoácidos

**Arquitetura proteica – aminoácidos** Proteínas são polímeros de aminoácidos, com cada **resíduo de aminoácido** unido ao seu vizinho por um tipo específico de ligação covalente (o termo “resíduo” reflete a perda de elementos de água quando um aminoácido é unido a outro). As proteínas podem ser degradadas (hidrolisadas) em seus aminoácidos constituintes por vários métodos, e os estudos mais iniciais de proteínas naturalmente se concentraram nesses aminoácidos livres delas derivados. Vinte aminoácidos diferentes são comumente encontrados em proteínas. O primeiro a ser descoberto foi a asparagina, em 1806. O último dos 20 a ser descoberto (treonina) não havia sido identificado até 1938. Todos os aminoácidos têm nomes comuns ou triviais, em alguns casos derivados da fonte da qual foram primeiramente isolados. A asparagina foi descoberta pela primeira vez no aspargo e o glutamato no glúten do trigo; a tirosina foi isolada a primeira vez a partir do queijo (seu nome é derivado do grego *tyros*, “queijo”); e a glicina (do grego *glykos*, “doce”) foi assim denominada devido ao seu sabor adocicado.

#### Aminoácidos compartilham características estruturais comuns

Todos os 20 tipos de aminoácidos comuns são  $\alpha$ -aminoácidos. Eles têm um grupo carboxila e um grupo amino ligados ao mesmo átomo de carbono (o carbono  $\alpha$ ) (**Figura 3-2**). Diferem uns dos outros em suas cadeias laterais ou **grupos R**, que variam em estrutura, tamanho e carga elétrica, e que influenciam a solubilidade dos aminoácidos em água. Além desses 20 aminoácidos, há muitos outros menos comuns. Alguns são resíduos modificados após a síntese de uma proteína; outros são aminoácidos presentes em organismos vivos, mas não como constituintes de proteínas. Foram atribuídas aos aminoácidos comuns das proteínas abreviações de três letras e símbolos de uma letra (Tabela 3-1), utilizados como abreviaturas para indicar a composição e a sequência de aminoácidos polimerizados em proteínas.

**CONVENÇÃO-CHAVE:** O código de três letras é transparente; as abreviações em geral consistem nas três primeiras letras do nome do aminoácido. O código de uma letra foi concebido por Margaret Oakley Dayhoff, considerada por muitos a fundadora do campo da bioinformática. O código de uma letra reflete uma tentativa de reduzir o tamanho dos arquivos de dados (em uma época da computação de cartões perfurados) utilizados para descrever as sequências de aminoácidos. Foi desenvolvido para ser facilmente memorizado, e a compreensão de sua origem pode ajudar os estudantes a fazer exatamente isso. Para seis aminoácidos (CHIMSV), a primeira letra do nome do aminoácido é única e, portanto, utilizada como o símbolo.



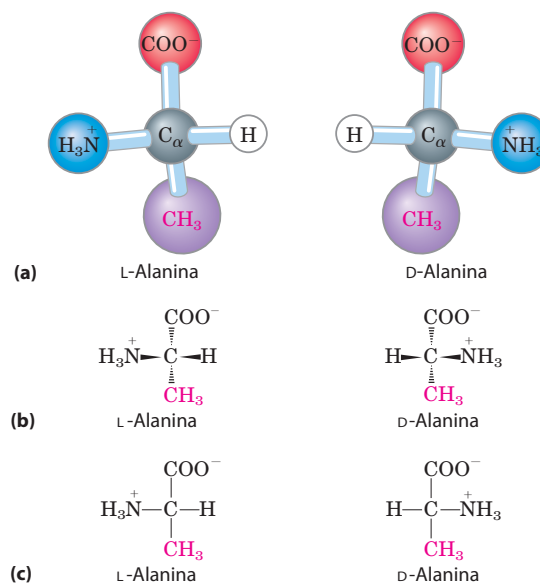
**FIGURA 3-2** Estrutura geral de um aminoácido. Esta estrutura é comum a todos os tipos de  $\alpha$ -aminoácidos, exceto um (a prolina, aminoácido cíclico, é a exceção). O grupo R, ou cadeia lateral (roxo), ligado ao carbono  $\alpha$  (cinza) é diferente em cada aminoácido.



Margaret Oakley Dayhoff, 1925–1983

Para cinco outros (AGLPT), a primeira letra não é única, mas é atribuída ao aminoácido mais comum em proteínas (por exemplo, leucina é mais comum do que lisina). Para outros quatro, a letra utilizada é foneticamente sugestiva (RFYW: aRginina, Fenilalanina, tirosina [do inglês *tYrosine*], triptofano [do inglês *tWiptophan*]). Os demais foram mais difíceis de nomear. Para quatro (DNEQ), foram atribuídas letras encontradas em seus nomes ou sugeridas por eles (aspartático [do inglês *asparDic*], asparagiNa, glutâmico [do inglês *glutamEke*], glutamina [do inglês *Q-tamine*]). Faltava a lisina. Sobravam poucas letras no alfabeto, e a letra K foi escolhida porque era a mais próxima de L. ■

Para todos os aminoácidos comuns, exceto a glicina, o carbono  $\alpha$  está ligado a quatro grupos diferentes: um grupo carboxila, um grupo amino, um grupo R e um átomo de hidrogênio (Figura 3-2; na glicina, o grupo R é outro átomo de hidrogênio). O átomo de carbono  $\alpha$  é, portanto, um **centro quiral** (p. 17). Em decorrência do arranjo tetraédrico dos orbitais de ligação em volta do átomo de carbono  $\alpha$ , os quatro grupos diferentes podem ocupar dois arranjos espaciais únicos e, portanto, os aminoácidos têm dois estereoisômeros possíveis. Uma vez que elas são imagens especulares não sobreponíveis uma da outra (**Figura 3-3**),



**FIGURA 3-3** Estereoisomerismo em  $\alpha$ -aminoácidos. (a) Os dois estereoisômeros da alanina, L- e D-alanina, são imagens especulares não sobrepostas um do outro (enantiômeros). (b, c) Duas convenções diferentes para representar as configurações espaciais dos estereoisômeros. Em fórmulas de perspectiva (b), as ligações sólidas em forma de cunha se projetam para fora do plano do papel, com as ligações tracejadas por trás dele. Em fórmulas de projeção (c), assume-se que as ligações horizontais se projetam para fora do plano do papel e as ligações verticais para trás. Entretanto, fórmulas de projeção muitas vezes são usadas casualmente e nem sempre pretendem representar uma configuração estereoquímica específica.

**TABELA 3-1** Propriedades e convenções associadas a aminoácidos comuns encontrados em proteínas

Aminoácido	Abreviação/ símbolo	$M_r^*$	Valores de $pK_a$			pI	Índice de hidropatia <sup>†</sup>	Ocorrência em proteínas (%) <sup>‡</sup>
			$pK_1$ (—COOH)	$pK_2$ (—NH <sub>3</sub> <sup>+</sup> )	$pK_R$ (grupo R)			
<b>Grupos R alifáticos, apolares</b>								
Glicina	Gly G	75	2,34	9,60		5,97	-0,4	7,2
Alanina	Ala A	89	2,34	9,69		6,01	1,8	7,8
Prolina	Pro P	115	1,99	10,96		6,48	-1,6	5,2
Valina	Val V	117	2,32	9,62		5,97	4,2	6,6
Leucina	Leu L	131	2,36	9,60		5,98	3,8	9,1
Isoleucina	Ile I	131	2,36	9,68		6,02	4,5	5,3
Metionina	Met M	149	2,28	9,21		5,74	1,9	2,3
<b>Grupos R aromáticos</b>								
Fenilalanina	Phe F	165	1,83	9,13		5,48	2,8	3,9
Tirosina	Tyr Y	181	2,20	9,11	10,07	5,66	-1,3	3,2
Triptofano	Trp W	204	2,38	9,39		5,89	-0,9	1,4
<b>Grupos R polares, não carregados</b>								
Serina	Ser S	105	2,21	9,15		5,68	-0,8	6,8
Treonina	Thr T	119	2,11	9,62		5,87	-0,7	5,9
Cisteína <sup>¶</sup>	Cys C	121	1,96	10,28	8,18	5,07	2,5	1,9
Asparagina	Asn N	132	2,02	8,80		5,41	-3,5	4,3
Glutamina	Gln Q	146	2,17	9,13		5,65	-3,5	4,2
<b>Grupos R carregados positivamente</b>								
Lisina	Lys K	146	2,18	8,95	10,53	9,74	-3,9	5,9
Histidina	His H	155	1,82	9,17	6,00	7,59	-3,2	2,3
Arginina	Arg R	174	2,17	9,04	12,48	10,76	-4,5	5,1
<b>Grupos R carregados negativamente</b>								
Aspartato	Asp D	133	1,88	9,60	3,65	2,77	-3,5	5,3
Glutamato	Glu E	147	2,19	9,67	4,25	3,22	-3,5	6,3

\*Os valores de  $M_r$  refletem as estruturas como mostradas na Figura 3-5. Os elementos da água ( $M_r$ , 18) são removidos quando o aminoácido é incorporado a um polipeptídeo.

<sup>†</sup>Uma escala combinando hidrofobicidade e hidrofiliabilidade de grupos R. Os valores refletem a energia livre ( $\Delta G$ ) de transferência da cadeia lateral do aminoácido de um solvente hidrofóbico para a água. Esta transferência é favorável ( $\Delta G < 0$ ; valor negativo no índice) para cadeias laterais de aminoácidos carregados ou polares, e desfavorável ( $\Delta G > 0$ ; valor positivo no índice) para aminoácidos com cadeias laterais apolares ou mais hidrofóbicas. Ver o Capítulo 11. A partir de Kyte, J. & Doolittle, R.F. (1982) *A simple method for displaying the hydropathic character of a protein*. *J. Mol. Biol.* **157**, 105-132.

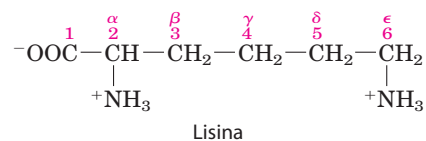
<sup>‡</sup>Ocorrência média em mais de 1.150 proteínas. De Doolittle, R.F. (1989) *Redundancies in protein sequences*. Em *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed.), pp. 599-623, Plenum Press, New York.

<sup>¶</sup>Em geral, a cisteína é classificada como polar apesar de apresentar um índice hidropático positivo. Isso reflete a capacidade do grupo sulfidril em atuar como ácido fraco e formar uma fraca ligação de hidrogênio com o oxigênio ou nitrogênio.

as duas formas representam uma classe de estereoisômeros denominada **enantiômeros** (ver Figura 1-20). Todas as moléculas com um centro quiral também são **ópticamente ativas** – isto é, elas giram o plano da luz polarizada (ver Quadro 1-2).

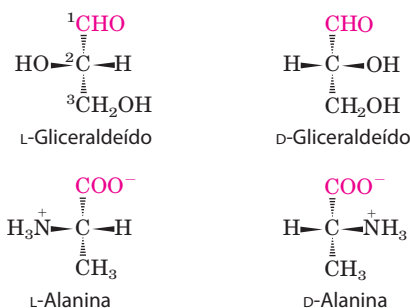
**CONVENÇÃO-CHAVE:** Duas convenções são utilizadas para identificar os carbonos em um aminoácido – prática que pode ser confusa. Os carbonos adicionais em um grupo R são comumente designados como  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , e assim por diante, a partir do carbono  $\alpha$ . Para a maioria das outras moléculas orgânicas, os átomos de carbono são simplesmente nume-

rados a partir de uma extremidade, conferindo a mais alta prioridade (C-1) ao carbono com o substituinte contendo o átomo de maior número atômico. Nessa última convenção, o carbono carboxílico de um aminoácido seria o C-1 e o carbono  $\alpha$  seria o C-2.



Em alguns casos, como aminoácidos com grupos R heterocíclicos (tal como a histidina), o sistema de letras gregas é ambíguo e a convenção numérica é então utilizada. Para aminoácidos de cadeias laterais ramificadas, carbonos equivalentes recebem números após as letras gregas. Leucina, portanto, tem carbonos  $\delta 1$  e  $\delta 2$  (ver a estrutura na Figura 3-5). ■

Uma nomenclatura especial foi desenvolvida para especificar a **configuração absoluta** dos quatro substituintes dos átomos de carbono assimétricos. As configurações absolutas de açúcares simples e de aminoácidos são especificadas pelo **sistema D, L** (Figura 3-4), com base na configuração absoluta do açúcar de três carbonos gliceraldeído, uma convenção proposta por Emil Fischer em 1891 (Fischer sabia que grupos circundavam o carbono assimétrico do gliceraldeído, mas teve de supor sua configuração absoluta; ele supôs corretamente, como foi confirmado posteriormente por análises de difração de raios x). Para todos os compostos quirais, os estereoisômeros com configuração relacionada àquela do L-gliceraldeído são designados L, e os estereoisômeros relacionados ao D-gliceraldeído foram designados D. Os grupos funcionais de L-alanina são combinados com aqueles de L-gliceraldeído pelo alinhamento daqueles que podem ser interconvertidos por reações químicas simples, de etapa única. Portanto, o grupo carboxila de L-alanina ocupa a mesma posição ao redor do carbono quiral que o grupo aldeído de L-gliceraldeído, porque um aldeído é prontamente convertido em um grupo carboxila por meio de uma oxidação de etapa única. Historicamente, as designações semelhantes L e D eram utilizadas para levorrotatória (rotação da luz polarizada à esquerda) e dextrorrotatória (rotação da luz polarizada à direita). Entretanto, nem todos os L-aminoácidos são levorrotatórios, e a convenção mostrada na Figura 3-4 foi necessária para evitar potenciais ambiguidades sobre a configuração absoluta. Pela convenção de Fischer, L e D se referem *apenas* à configuração absoluta dos quatro substituintes em torno do carbono quiral, e não às propriedades ópticas da molécula.



**FIGURA 3-4** Relação estérica dos estereoisômeros de alanina à configuração absoluta do L-gliceraldeído e do D-gliceraldeído. Nestas fórmulas em perspectiva, os carbonos são alinhados verticalmente, com o átomo quiral no centro. Os carbonos nestas moléculas são numerados de 1 a 3, de cima para baixo, começando com o carbono do aldeído ou carboxiterminal (vermelho), como mostrado. Quando apresentado desta maneira, o grupo R do aminoácido (nesse caso o grupo metil da alanina) está sempre abaixo do carbono  $\alpha$ . Os L-aminoácidos são aqueles com o grupo  $\alpha$ -amino na esquerda e os D-aminoácidos com esse grupo na direita.

Outro sistema para especificar a configuração ao redor de um centro quiral é o **sistema RS**, utilizado na nomenclatura sistemática da química orgânica para descrever, com mais exatidão, a configuração das moléculas com mais de um centro quiral (p. 18).

## Os resíduos de aminoácidos em proteínas são estereoisômeros L

Quase todos os compostos biológicos com centro quiral ocorrem naturalmente em apenas uma forma estereoisomérica, D ou L. Os resíduos de aminoácidos em moléculas proteicas são exclusivamente estereoisômeros L. Os resíduos de D-aminoácidos foram encontrados apenas em alguns peptídeos, geralmente pequenos, incluindo alguns peptídeos de paredes celulares bacterianas e certos antibióticos peptídicos.

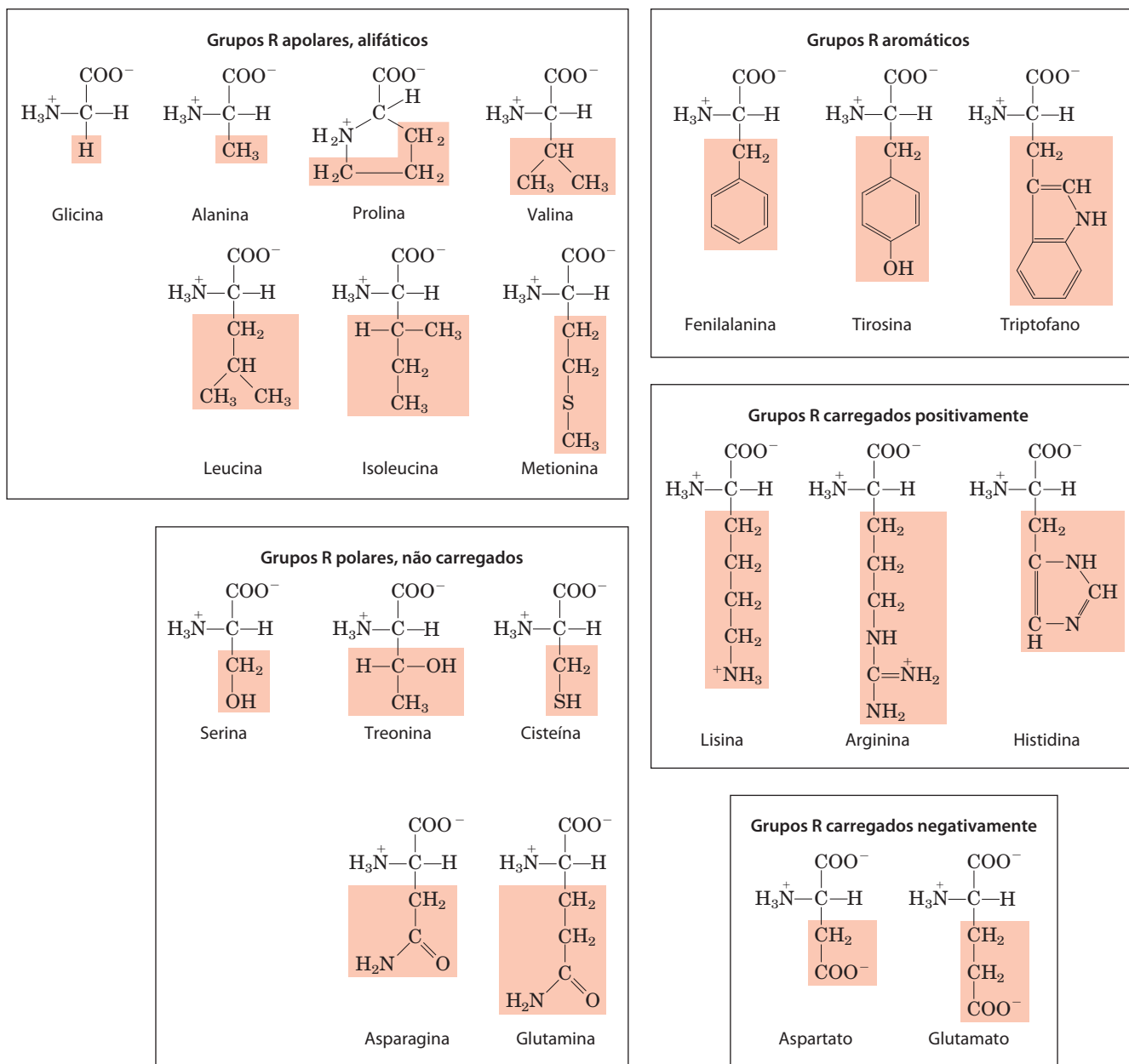
É notável que praticamente todos os resíduos de aminoácidos em proteínas sejam estereoisômeros L. Quando compostos quirais são formados em reações químicas comuns, o resultado é uma mistura racêmica de isômeros D e L, os quais são difíceis para um químico distinguir e separar. Contudo, para um sistema vivo, os isômeros D e L são tão diferentes entre si quanto a mão direita é diferente da esquerda. A formação de subestruturas repetidas estáveis em proteínas (Capítulo 4) geralmente exige que seus aminoácidos constituintes sejam de uma série estereoquímica. As células são capazes de sintetizar especificamente os isômeros L de aminoácidos porque os sítios ativos de enzimas são assimétricos, tornando estereoespecíficas as reações por elas catalisadas.

## Aminoácidos podem ser classificados pelo grupo R

O conhecimento das propriedades químicas dos aminoácidos comuns é fundamental para a compreensão da bioquímica. O tópico pode ser simplificado agrupando-se os aminoácidos em cinco classes principais com base nas propriedades dos seus grupos R (Tabela 3-1), particularmente sua **polaridade** ou tendência para interagir com a água em pH biológico (próximo do pH 7,0). A polaridade dos grupos R varia amplamente, de apolar e hidrofóbico (não hidrossolúvel) ao altamente polar e hidrofílico (hidrossolúvel). Alguns aminoácidos são um pouco difíceis de caracterizar ou não se encaixam perfeitamente em qualquer grupo, particularmente glicina, histidina e cisteína. Suas atribuições a determinados grupos são o resultado de avaliações ponderadas em vez de absolutas.

As estruturas dos 20 aminoácidos comuns são mostradas na Figura 3-5, e algumas de suas propriedades são listadas na Tabela 3-1. Em cada classe há gradações de polaridade, tamanho e forma dos grupos R.

**Grupos R apolares, alifáticos** Os grupos R nesta classe de aminoácidos são apolares e hidrofóbicos. As cadeias laterais de **alanina, valina, leucina** e **isoleucina** tendem a se agrupar no interior de proteínas, estabilizando a estrutura proteica por meio de interações hidrofóbicas. A **glicina** tem a estrutura mais simples. Embora seja mais facilmente agrupada com os aminoácidos apolares, sua cadeia lateral



**FIGURA 3-5 Os 20 aminoácidos comuns de proteínas.** As fórmulas estruturais mostram o estado de ionização que predomina em pH 7,0. As porções não sombreadas são aquelas comuns a todos os aminoácidos; aquelas sombreadas são os grupos R. Embora o grupo R da histidina seja mostrado

sem carga, seu  $pK_a$  (ver a Tabela 3-1) é tal que uma pequena mas significativa fração desses grupos seja positivamente carregada em pH 7,0. A forma protonada da histidina é mostrada acima do gráfico na Figura 3-12b.

muito pequena não contribui realmente para interações hidrofóbicas. A **metionina**, um dos dois aminoácidos que contém enxofre, tem um grupo tioéter ligeiramente apolar em sua cadeia lateral. A **prolina** tem cadeia lateral alifática com estrutura cíclica distinta. O grupo amino secundário (imino) de resíduos de prolina é mantido em uma configuração rígida que reduz a flexibilidade estrutural de regiões polipeptídicas contendo prolina.

**Grupos R aromáticos Fenilalanina, tirosina e triptofano**, com suas cadeias laterais aromáticas, são relativamente apolares (hidrofóbicos). Todos podem participar em in-

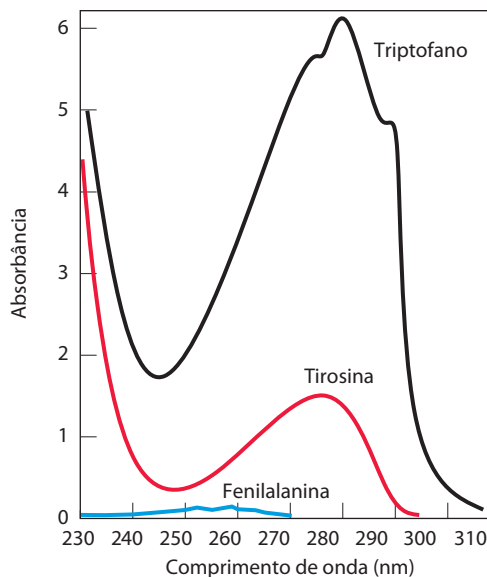
terações hidrofóbicas. O grupo hidroxila da tirosina pode formar ligações de hidrogênio e é um importante grupo funcional em algumas enzimas. A tirosina e o triptofano são significativamente mais polares do que a fenilalanina, devido ao grupo hidroxila da tirosina e ao nitrogênio do anel indol do triptofano.

O triptofano, a tirosina e, em menor extensão, a fenilalanina, absorvem a luz ultravioleta (**Figura 3-6**; ver também Quadro 3-1). Isso explica a forte absorbância de luz com comprimento de onda de 280 nm característica da maior parte das proteínas, propriedade explorada por pesquisadores na caracterização de proteínas.



**Grupos R polares, não carregados** Os grupos R desses aminoácidos são mais solúveis em água, ou mais hidrofílicos do que aqueles dos aminoácidos apolares, porque eles contêm grupos funcionais que formam ligações de hidrogênio com a água. Essa classe de aminoácidos inclui a **serina, treonina, cisteína, asparagina e glutamina**. Os grupos hidroxila da serina e treonina e os grupos amida da asparagina e glutamina contribuem para suas polaridades. A cisteína é um caso isolado aqui porque sua polaridade, devida ao seu grupo sulfidril, é bastante modesta. A cisteína é um ácido

**FIGURA 3-6 Absorção da luz ultravioleta por aminoácidos aromáticos.** Comparação dos espectros de absorção de luz dos aminoácidos aromáticos triptofano, tirosina e fenilalanina em pH 6,0. Os aminoácidos estão presentes em quantidades equimolares ( $10^{-3}M$ ) sob condições idênticas. A absorbância medida do triptofano é mais do que quatro vezes aquela da tirosina em um comprimento de onda de 280 nm. Observe que a absorção luminosa máxima tanto para o triptofano quanto para a tirosina ocorre próxima de 280 nm. A absorção luminosa pela fenilalanina geralmente contribui pouco para as propriedades espectroscópicas das proteínas.



### QUADRO 3-1 MÉTODOS Absorção de luz por moléculas: a Lei de Lambert-Beer

Uma ampla variedade de biomoléculas absorve a luz em comprimentos de onda característicos, como o triptofano, que absorve a luz em 280 nm (ver Figura 3-6). A medida da absorção da luz por um espectrofotômetro é utilizada para detectar e identificar moléculas e para determinar suas concentrações em solução. A fração da luz incidente absorvida por uma solução em um determinado comprimento de onda está relacionada à espessura da camada de absorção (comprimento do caminho) e à concentração da substância que absorve (Figura 1). Essas duas relações são combinadas na lei de Lambert-Beer,

$$\log \frac{I_0}{I} = \epsilon cl$$

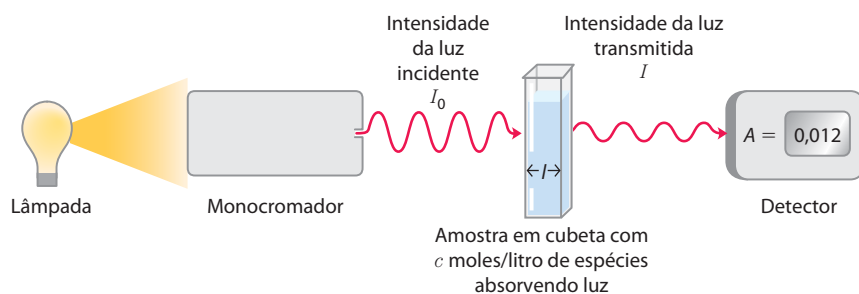
em que  $I_0$  é a intensidade da luz incidente,  $I$  é a intensidade da luz transmitida, a relação  $I/I_0$  (o inverso da razão na equação) é a transmitância,  $\epsilon$  é o coeficiente de extinção molar (em unidades de litros por mol por centímetro),  $c$  é a concentração da substância absorvida (em mols por

litro), e  $l$  é o comprimento do caminho de luz da amostra absorvente de luz (em centímetros).

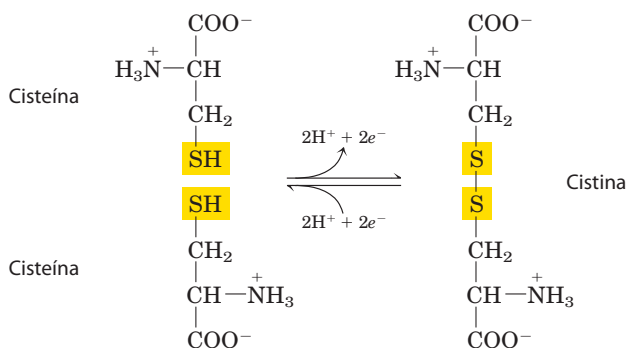
A lei de Lambert-Beer pressupõe que a luz incidente é paralela e monocromática (de um único comprimento de onda) e que as moléculas de solvente e soluto são orientadas aleatoriamente. A expressão  $\log(I_0/I)$  é denominada **absorbância** e designada  $A$ .

É importante observar que cada milímetro sucessivo do comprimento do caminho da solução absorvente em uma célula de 1,0 cm não absorve uma quantidade constante, mas uma fração constante da luz que incide sobre ela. Entretanto, com uma camada absorvente de comprimento de caminho fixo, a **absorbância**,  $A$ , é **diretamente proporcional à concentração do soluto absorvente**.

O coeficiente de extinção molar varia com a natureza do composto absorvente, do solvente e do comprimento de onda, e também com o pH se a substância que absorve a luz está em equilíbrio com um estado de ionização que possui diferentes propriedades de absorção.



**FIGURA Q-1** Os principais componentes de um espectrofotômetro. A fonte de luz emite luz em um amplo espectro, quando o monocromador seleciona e transmite luz de um comprimento de onda específico. A luz monocromática passa através da amostra em uma cubeta de tamanho  $l$  e é absorvida pela amostra como uma proporção da concentração das espécies absorvendo luz. A luz transmitida é medida por um detector.



**FIGURA 3-7** Formação reversível de uma ligação dissulfeto pela oxidação de duas moléculas de cisteína. Ligações de dissulfeto entre resíduos Cys estabilizam as estruturas de muitas proteínas.

fraco e pode fazer fracas ligações de hidrogênio com o oxigênio ou nitrogênio.

A asparagina e a glutamina são as amidas de dois outros aminoácidos também encontrados em proteínas – aspartato e glutamato, respectivamente – nos quais a asparagina e a glutamina são facilmente hidrolisadas por ácido ou base. A cisteína é prontamente oxidada para formar um aminoácido dimérico ligado de modo covalente chamado **cistina**, no qual duas moléculas ou resíduos de cisteína são ligadas por uma ligação dissulfeto (**Figura 3-7**). Os resíduos ligados a dissulfetos são fortemente hidrofóbicos (apolares).

As ligações dissulfeto desempenham um papel especial nas estruturas de muitas proteínas pela formação de ligações covalentes entre partes de uma molécula polipeptídica ou entre duas cadeias polipeptídicas diferentes.

**Grupos R carregados positivamente (básicos)** Os grupos R mais hidrofílicos são aqueles carregados positivamente ou negativamente. Os aminoácidos nos quais os grupos R têm uma carga positiva significativa em pH 7,0 são a **lisina**, com um segundo grupo amino primário na posição  $\epsilon$  em sua cadeia alifática; a **arginina**, com um grupo guanidínio positivamente carregado; e a **histidina**, com um grupo imidazol aromático. Como o único aminoácido comum que tem uma cadeia lateral ionizável com  $pK_a$  próximo da neutralidade, a histidina pode ser positivamente carregada (forma protonada) ou não carregada em pH 7,0. Seus resíduos facilitam muitas reações catalisadas por enzimas, funcionando como doadores/aceptores de prótons.

**Grupos R carregados negativamente (ácidos)** Os dois aminoácidos que apresentam grupos R com carga negativa final em pH 7,0 são o **aspartato** e o **glutamato**, cada um dos quais tem um segundo grupo carboxila.

### Aminoácidos incomuns também têm funções importantes

Além dos 20 aminoácidos comuns, as proteínas podem conter resíduos criados por modificações de resíduos comuns já incorporados em um polipeptídeo (**Figura 3-8a**). Entre esses aminoácidos incomuns estão a **4-hidroxiprolina**, um derivado da prolina, e a **5-hidroxilisina**, derivada da lisina.

O primeiro é encontrado em proteínas da parede celular de células vegetais e ambos são encontrados no colágeno, proteína fibrosa de tecidos conectivos. A **6-N-metil-lisina** é um constituinte da miosina, uma proteína contrátil do músculo. Outro aminoácido incomum importante é o  **$\gamma$ -carboxiglutamato**, encontrado na proteína de coagulação protrombina e em algumas outras proteínas que se ligam ao  $Ca^{2+}$  como parte de suas funções biológicas. Mais complexa é a **desmosina**, derivada de quatro resíduos Lys, encontrada na proteína fibrosa elastina.

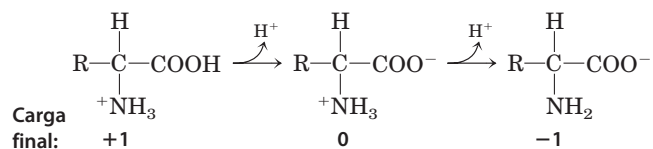
A **selenocisteína** é um caso especial. Esse raro resíduo de aminoácido é introduzido durante a síntese proteica, em vez de criado por uma modificação pós-sintética. Contém selênio em vez do enxofre da cisteína. Na verdade derivada de serina, a selenocisteína é um constituinte de apenas algumas poucas proteínas conhecidas.

Alguns resíduos de aminoácidos em uma proteína podem ser modificados transitariamente para alterar as funções da proteína. A adição de grupos fosforil, metil, acetil, adenilil, ADP-ribosil ou outros grupos a resíduos de aminoácidos específicos pode aumentar ou diminuir a atividade de uma proteína (**Figura 3-8b**). A fosforilação é uma modificação reguladora particularmente comum. A modificação covalente como uma estratégia reguladora em uma proteína é discutida com mais detalhe no Capítulo 6.

Cerca de 300 aminoácidos adicionais foram encontrados nas células. Eles têm várias funções, mas não são todos constituintes de proteínas. A **ornitina** e a **citrulina** (**Figura 3-8c**) merecem atenção especial porque são intermediários-chave (metabólitos) na biossíntese de arginina (Capítulo 22) e no ciclo da ureia (Capítulo 18).

### Aminoácidos podem agir como ácidos e bases

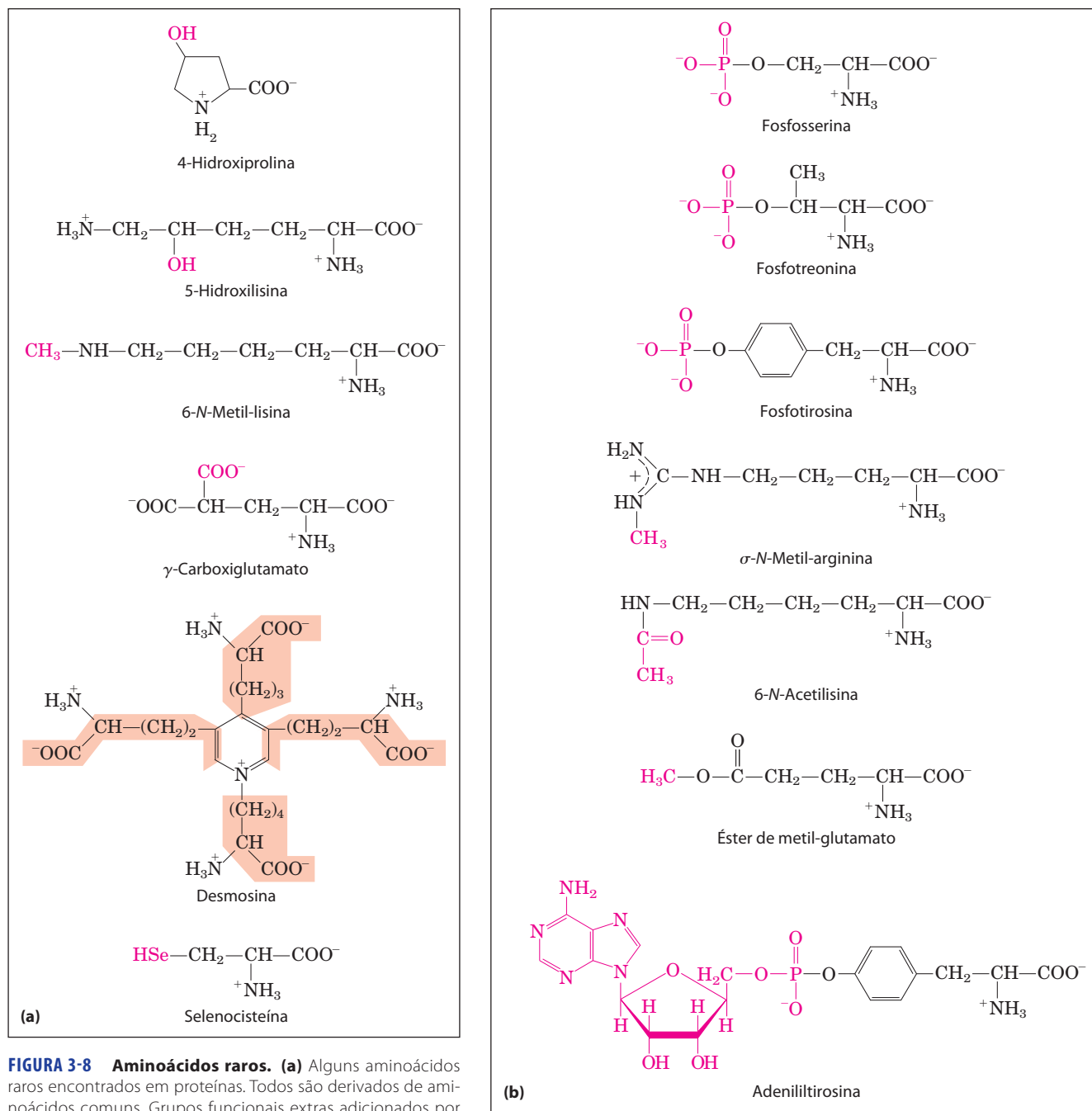
Os grupos amino e carboxila de aminoácidos, em conjunto com os grupos ionizáveis R de alguns aminoácidos, funcionam como ácidos e bases fracos. Quando um aminoácido sem um grupo R ionizável é dissolvido em água em pH neutro, ele permanece na solução como um íon bipolar, ou **zwitterion** (do alemão “íon híbrido”), que pode agir como ácido ou base (**Figura 3-9**). Substâncias com essa natureza dupla (ácido-base) são **anfotéricas** e são frequentemente chamadas de **anfólitos** (a partir de “eletrólitos anfotéricos”). Um simples  $\alpha$ -aminoácido monoamino monocarboxílico, como a alanina, é um ácido diprótico quando completamente protonado; ele tem dois grupos, o grupo  $-COOH$  e o grupo  $-NH_3^+$ , que pode produzir dois prótons:



### Aminoácidos têm curvas de titulação características

A titulação ácido-base envolve a adição ou remoção gradual de prótons (Capítulo 2). A **Figura 3-10** mostra a curva de titulação de uma forma diprótica de glicina.

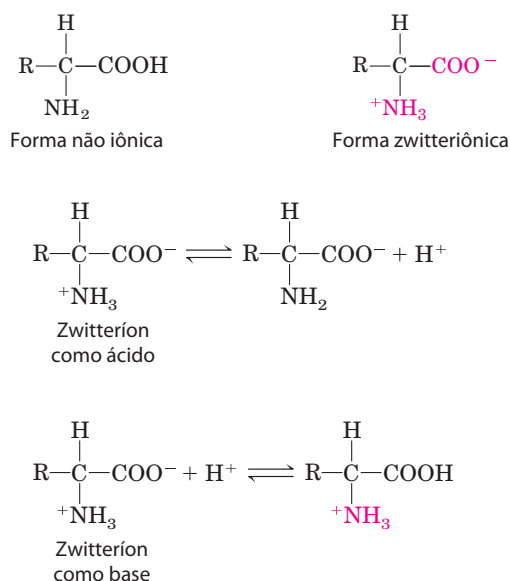




**FIGURA 3-8 Aminoácidos raros.** (a) Alguns aminoácidos raros encontrados em proteínas. Todos são derivados de aminoácidos comuns. Grupos funcionais extras adicionados por reações de modificação são mostrados em vermelho. A desmosina é formada a partir de quatro resíduos Lys (os esqueletos de carbono estão sombreados). Observe o uso tanto de números quanto de letras gregas nos nomes dessas estruturas para identificar os átomos de carbono alterados. (b) Modificações dos aminoácidos reversíveis envolvidos na regulação da atividade proteica. A fosforilação é o tipo mais comum de modificação regulatória. (c) Ornitina e citrulina, não encontrados em proteínas, são intermediários na biossíntese de arginina e no ciclo da ureia.

Os dois grupos ionizáveis de glicina, o grupo carboxila e o grupo amino, são titulados com uma base forte, como NaOH. O gráfico tem duas fases distintas, correspon-

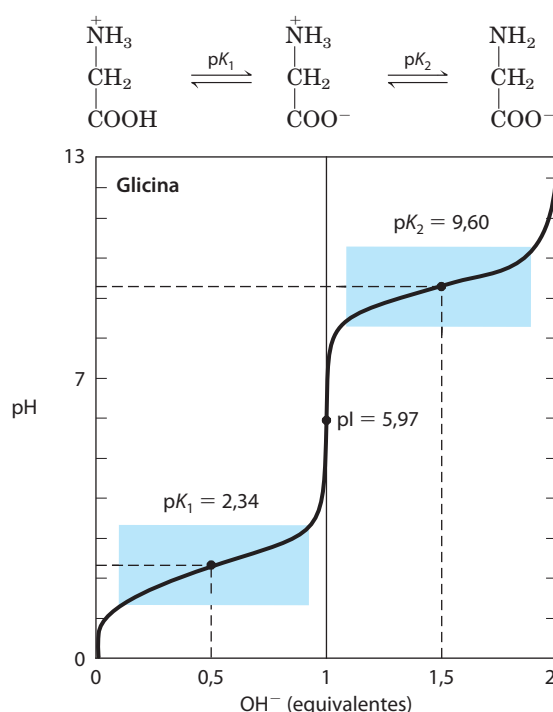
dendo à desprotonação de dois grupos diferentes na glicina. Cada uma das duas fases se assemelha ao formato da curva de titulação de um ácido monoprótico, como o



**FIGURA 3-9 Formas não iônicas e zwitteriônicas de aminoácidos.** A forma não iônica não ocorre em quantidades significativas em soluções aquosas. O zwitterion predomina em pH neutro. Um zwitterion pode atuar tanto como ácido (doador de prótons) quanto como base (aceptor de prótons).

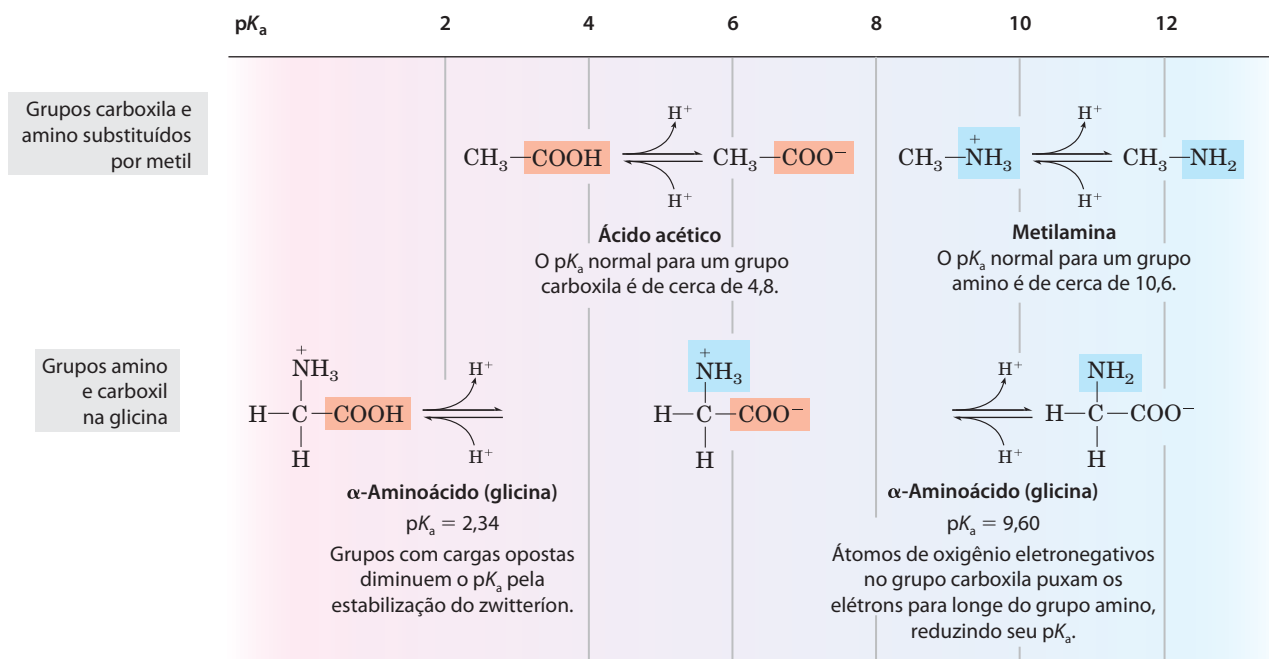
ácido acético (ver Figura 2-17), e pode ser analisada do mesmo modo. Em pH muito baixo, a espécie iônica predominante de glicina é a forma completamente protonada,  $^+\text{H}_3\text{N}-\text{CH}_2-\text{COOH}$ . No primeiro estágio da titulação, o grupo  $-\text{COOH}$  de glicina perde seu próton. No ponto médio desse estágio, estão presentes concentrações equimolares de espécies doadoras ( $^+\text{H}_3\text{N}-\text{CH}_2-\text{COOH}$ ) e aceptoras ( $^+\text{H}_3\text{N}-\text{CH}_2-\text{COO}^-$ ) de prótons. Como na titulação de qualquer ácido fraco, um ponto de inflexão é alcançado nesse ponto médio onde o pH é igual ao  $\text{pK}_a$  do grupo protonado que está sendo titulado (ver Figura 2-18). Para a glicina, o pH no ponto médio é 2,34, portanto seu grupo  $-\text{COOH}$  tem um  $\text{pK}_a$  (marcado  $\text{pK}_1$  na Figura 3-10) de 2,34 (lembre-se do Capítulo 2 que pH e  $\text{pK}_a$  são simplesmente notações convenientes para concentração de prótons e a constante de equilíbrio para ionização, respectivamente. O  $\text{pK}_a$  é uma medida da tendência de um grupo doar um próton, com essa tendência diminuindo dez vezes à medida que o  $\text{pK}_a$  aumenta em uma unidade). À medida que a titulação da glicina prossegue, outro ponto importante é alcançado no pH 5,97. Aqui há outro ponto de inflexão, no qual a remoção do primeiro próton está completa e a remoção do segundo apenas começou. Nesse pH, a glicina está presente em grande parte como o íon bipolar (zwitterion)  $^+\text{H}_3\text{N}-\text{CH}_2-\text{COO}^-$ . Em breve será analisado o significado desse ponto de inflexão na curva de titulação (marcado como pI na Figura 3-10).

O segundo estágio da titulação corresponde à remoção de um próton do grupo  $-\text{NH}_3^+$  da glicina. O pH no ponto médio dessa fase é 9,60, igual ao  $\text{pK}_a$  (marcado  $\text{pK}_2$  na Figura 3-10) para o grupo  $-\text{NH}_3^+$ . A titulação está completa em um pH de cerca de 12, no ponto em que a forma predominante de glicina é  $\text{H}_2\text{N}-\text{CH}_2-\text{COO}^-$ .



**FIGURA 3-10 Titulação de um aminoácido.** Aqui é mostrada a curva de titulação de 0,1 M de glicina a 25°C. As espécies iônicas que predominam em pontos-chave na titulação são mostradas acima do gráfico. Os retângulos sombreados, centrados em torno de  $\text{pK}_1 = 2,34$  e  $\text{pK}_2 = 9,60$ , indicam as regiões de maior poder de tamponamento. Observe que 1 equivalente de  $\text{OH}^- = 0,1 \text{ M}$  de NaOH foi adicionado.

A partir da curva de titulação da glicina, é possível obter várias informações importantes. Em primeiro lugar, ela fornece uma medida quantitativa do  $\text{pK}_a$  de cada um dos dois grupos ionizáveis: 2,34 para o grupo  $-\text{COOH}$  e 9,60 para o grupo  $-\text{NH}_3^+$ . Observe que o grupo carboxila da glicina é mais de cem vezes mais ácido (mais facilmente ionizado) do que o grupo carboxila do ácido acético, que, como foi visto no Capítulo 2, tem um  $\text{pK}_a$  de 4,76 – próximo da média para um grupo carboxila ligado a um hidrocarboneto alifático não substituído. O  $\text{pK}_a$  alterado da glicina é provocado pela repulsão entre o próton que está saindo e o grupo amino próximo positivamente carregado no átomo de carbono  $\alpha$ , como descrito na Figura 3-11. As cargas opostas no zwitterion resultante estão estabilizadas. De modo semelhante, o  $\text{pK}_a$  do grupo amino na glicina é alterado para baixo em relação ao  $\text{pK}_a$  médio de um grupo amino. Esse efeito se deve parcialmente aos átomos de oxigênio eletronegativos nos grupos carboxila, que tendem a puxar os elétrons na direção deles, aumentando a tendência do grupo amino em abrir mão de um próton. Assim, o grupo  $\alpha$ -amino tem um  $\text{pK}_a$  menor do que o de um de uma amina alifática, como a metilamina (Figura 3-11). Em resumo, o  $\text{pK}_a$  de qualquer grupo funcional é em grande parte afetado por seu ambiente químico, fenômeno algumas vezes explorado nos sítios ativos de enzimas para promover mecanismos de reação extraordinariamente adaptados que dependem dos valores de  $\text{pK}_a$  perturbados de grupos doadores/aceptores de prótons de resíduos específicos.



**FIGURA 3-11** Efeito do ambiente químico no  $pK_a$ . Os valores de  $pK_a$  para os grupos ionizáveis na glicina são mais baixos do que aqueles dos grupos simples de carboxila e o amino substituídos por metil. Essas perturbações do  $pK_a$  se devem a interações intramoleculares. Efeitos semelhantes podem ser causados por grupos químicos que possam estar posicionados próximos – por exemplo, no sítio ativo de uma enzima.

bações do  $pK_a$  se devem a interações intramoleculares. Efeitos semelhantes podem ser causados por grupos químicos que possam estar posicionados próximos – por exemplo, no sítio ativo de uma enzima.

A segunda informação fornecida pela curva de titulação da glicina é que esse aminoácido tem duas regiões com poder de tamponamento. Uma delas está na parte relativamente achatada da curva, se estendendo por aproximadamente 1 unidade de pH de cada lado do primeiro  $pK_a$  de 2,34, indicando que a glicina é um bom tampão próxima desse pH. A outra zona de tamponamento está centrada em volta do pH 9,60 (observe que a glicina não é um bom tampão no pH do líquido intracelular ou do sangue, em torno de 7,4). Dentro das faixas de tamponamento da glicina, a equação de Henderson-Hasselbalch (p. 64) pode ser utilizada para calcular as proporções de espécies de glicina próton-doadoras e próton-aceptoras necessárias para preparar um tampão em um determinado pH.

### Curvas de titulação predizem a carga elétrica dos aminoácidos

Outra importante peça de informação derivada da curva de titulação de um aminoácido é a relação entre a sua carga final e o pH da solução. No pH de 5,97, o ponto de inflexão entre os dois estágios na sua curva de titulação, a glicina está presente predominantemente em sua forma bipolar, totalmente ionizada, mas sem carga elétrica final (Figura 3-10). O pH característico no qual a carga elétrica final é zero é chamado de **ponto isoelétrico** ou **pH isoelétrico**, designado por **pI**. Para a glicina, que não possui qualquer grupo ionizável em sua cadeia lateral, o ponto isoelétrico é simplesmente a média aritmética dos dois valores de  $pK_a$ :

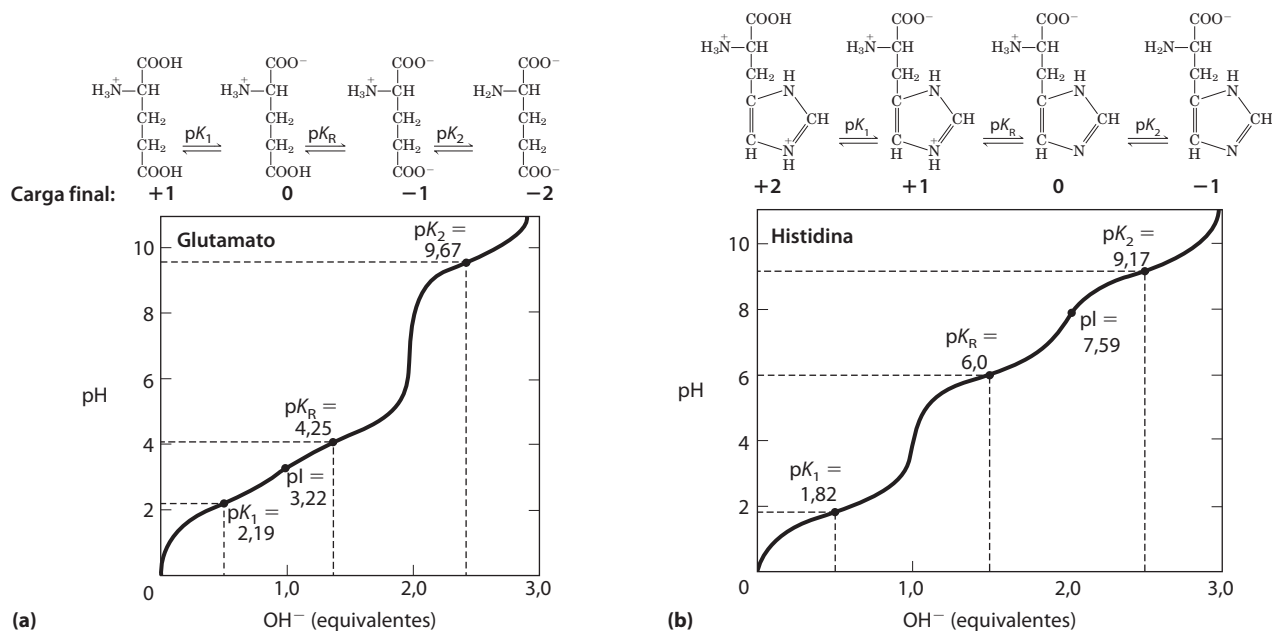
$$pI = \frac{1}{2}(pK_1 + pK_2) = \frac{1}{2}(2,34 + 9,60) = 5,97$$

Como evidenciado na Figura 3-10, a glicina tem uma carga final negativa em qualquer pH acima do seu pI e, portanto, irá se deslocar na direção do eletrodo positivo (o ânodo) quando colocada em um campo elétrico. Em qualquer pH abaixo do seu pI, a glicina tem uma carga final positiva e irá se deslocar em direção ao eletrodo negativo (o cátodo). Quanto mais distante for o pH de uma solução de glicina de seu ponto isoelétrico, maior será a carga elétrica final da população de moléculas de glicina. Em um pH igual a 1,0, por exemplo, a glicina existe quase totalmente na forma  $^+H_3N-CH_2-COOH$  com uma carga positiva final igual a 1,0. Em um pH de 2,34, onde há uma igual mistura de  $^+H_3N-CH_2-COOH$  e  $^+H_3N-CH_2-COO^-$ , a média ou a carga final positiva é igual a 0,5. O sinal e a magnitude da carga final de qualquer aminoácido em qualquer pH podem ser previstos do mesmo modo.

### Aminoácidos diferem em suas propriedades acidobásicas

As propriedades compartilhadas de muitos aminoácidos permitem algumas generalizações simplificadas sobre seu comportamento acidobásico. Em primeiro lugar, todos os aminoácidos com um único grupo  $\alpha$ -amino, um único grupo  $\alpha$ -carboxila e um grupo R não ionizável têm curvas de titulação semelhantes à da glicina (Figura 3-10). Esses aminoácidos têm valores de  $pK_a$  muito semelhantes, mas não idênticos:  $pK_a$  do grupo  $-COOH$  na faixa de 1,8 a 2,4, e  $pK_a$  do grupo  $-NH_3^+$  na faixa de 8,8 a 11,0 (Tabela 3-1).

As diferenças nesses valores de  $pK_a$  refletem os ambientes químicos impostos por seus grupos R. Em segundo lugar, os aminoácidos com um grupo R ionizável têm curvas de titulação mais complexas, com *três* estágios correspondendo às três etapas possíveis de ionização; as-



**FIGURA 3-12** Curvas de titulação para (a) glutamato e (b) histidina. O grupo R do  $pK_a$  é designado aqui como  $pK_R$ .

sim, eles possuem três valores de  $pK_a$ . O estágio adicional para a titulação do grupo R ionizável se funde, em algum grau, com aquele para a titulação do grupo  $\alpha$ -carboxila, para a titulação do grupo  $\alpha$ -amino, ou ambos. As curvas de titulação para dois aminoácidos desse grupo, glutamato e histidina, são mostradas na **Figura 3-12**. Os pontos isoeletrônicos refletem a natureza dos grupos R ionizáveis presentes. Por exemplo, o glutamato tem um pI de 3,22, consideravelmente mais baixo do que o da glicina. Isso se deve à presença de dois grupos carboxila, que, na média de seus valores de  $pK_a$  (3,22), contribuem para uma carga final de  $-1$  que equilibra o  $+1$  proveniente do grupo amina. Do mesmo modo, o pI da histidina, com dois grupos positivamente carregados quando protonados, é de 7,59 (a média dos valores de  $pK_a$  dos grupos amina e imidazol), muito mais alto do que aquele da glicina.

Por fim, como apontado anteriormente, sob a condição geral de exposição livre e aberta ao ambiente aquoso, apenas a histidina tem um grupo R ( $pK_a = 6,0$ ) que fornece um poder de tamponamento significativo próximo do pH neutro normalmente encontrado nos líquidos intracelulares e extracelulares da maior parte dos animais e bactérias (Tabela 3-1).

### RESUMO 3.1 Aminoácidos

► Os 20 aminoácidos comumente encontrados como resíduos em proteínas contêm um grupo  $\alpha$ -carboxila, um grupo  $\alpha$ -amino e um grupo R característico substituído no átomo do carbono  $\alpha$ . O átomo de carbono  $\alpha$  de todos os aminoácidos, exceto a glicina, é assimétrico e, portanto, os aminoácidos podem existir em pelo menos duas formas estereoisoméricas. Apenas os estereoisômeros L, com uma configuração relacionada à configuração absoluta da molécula de referência L gliceraldeído, são encontrados em proteínas.

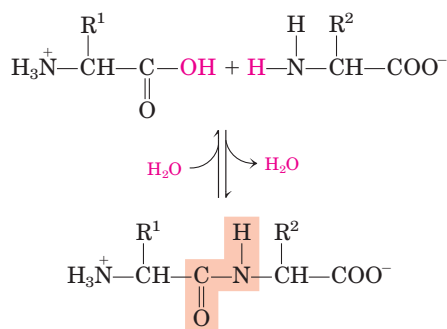
- Outros aminoácidos menos comuns também ocorrem, tanto como constituintes de proteínas (pela modificação de resíduos de aminoácidos comuns após a síntese proteica) quanto como metabólitos livres.
- Os aminoácidos podem ser classificados em cinco tipos com base na polaridade e carga (em pH 7) de seus grupos R.
- Os aminoácidos variam em suas propriedades ácido-básicas e têm curvas de titulação características. Aminoácidos monoamino monocarboxílicos (com grupos R não ionizáveis) são ácidos dipróticos ( $^+H_3NCH(R)COOH$ ) em pH baixo e existem em várias formas iônicas diferentes à medida que o pH aumenta. Aminoácidos com grupos R ionizáveis têm espécies iônicas adicionais, dependendo do pH do meio e do  $pK_a$  do grupo R.

## 3.2 Peptídeos e proteínas

Agora o foco passa a ser os polímeros de aminoácidos, os **peptídeos** e as **proteínas**. Os polipeptídeos que ocorrem biologicamente variam em tamanho de pequenos a muito grandes, consistindo em dois ou três a milhares de resíduos de aminoácidos ligados. Aqui, serão focalizadas as propriedades químicas fundamentais desses polímeros.

### Peptídeos são cadeias de aminoácidos

Duas moléculas de aminoácidos podem ser ligadas de modo covalente por meio de uma ligação amida substituída, denominada **ligação peptídica**, a fim de produzir um dipeptídeo. Tal ligação é formada pela remoção de elementos de água (desidratação) do grupo  $\alpha$ -carboxila de um aminoácido e do grupo  $\alpha$ -amino do outro (**Figura 3-13**). A formação da ligação peptídica é um exemplo de



**FIGURA 3-13** Formação de uma ligação peptídica por condensação.

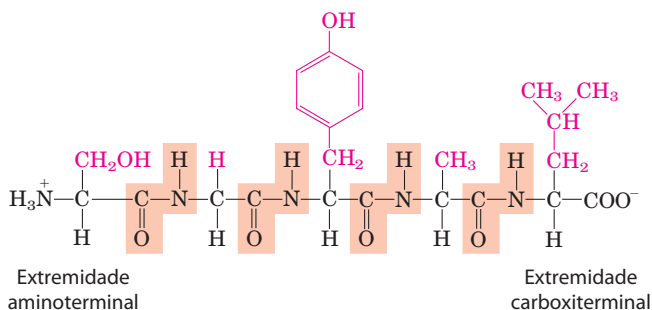
O grupo  $\alpha$ -amino de um aminoácido (com grupo  $\text{R}^2$ ) atua como nucleófilo para deslocar o grupo hidroxila de outro aminoácido (com grupo  $\text{R}^1$ ), formando uma ligação peptídica (sombreada). Os grupos amino são bons nucleófilos, mas o grupo hidroxila é um grupo de saída fraco e não prontamente deslocado. No pH fisiológico, a reação mostrada aqui não ocorre em grau apreciável.

uma reação de condensação, uma classe comum de reações nas células vivas. Em condições bioquímicas padrão, o equilíbrio para a reação mostrada na Figura 3-13 favorece os aminoácidos em relação ao dipeptídeo. Para tornar a reação mais favorável termodinamicamente, o grupo carboxila deve ser modificado ou ativado quimicamente, de modo que o grupo hidroxila possa ser mais rapidamente eliminado. Uma abordagem química para esse problema será destacada posteriormente neste capítulo. A abordagem biológica para a formação de ligações peptídicas é o tópico principal do Capítulo 27.

Três aminoácidos podem ser unidos por duas ligações peptídicas para formar um tripeptídeo; do mesmo modo, quatro aminoácidos podem ser unidos para formar um tetrapeptídeo, cinco para formar um pentapeptídeo, e assim por diante. Quando alguns aminoácidos se ligam desse modo, a estrutura é chamada de **oligopeptídeo**. Quando muitos aminoácidos se ligam, o produto é chamado de **polipeptídeo**. As proteínas podem ter milhares de resíduos de aminoácidos. Embora os termos “proteína” e “polipeptídeo” sejam algumas vezes intercambiáveis, as moléculas chamadas de polipeptídeos têm massas moleculares abaixo de 10.000, e as chamadas de proteínas têm massas moleculares mais elevadas.

A **Figura 3-14** mostra a estrutura de um pentapeptídeo. Como já observado, uma unidade de aminoácido em um peptídeo é frequentemente chamada de resíduo (a parte restante após a perda de elementos de água – um átomo de hidrogênio de seu grupo amino e a metade hidroxila de seu grupo carboxila). Em um peptídeo, o resíduo de aminoácido na extremidade com um grupo  $\alpha$ -amino livre é chamado de resíduo **aminoterminal** (ou *N*-terminal); o resíduo na outra extremidade, que tem um grupo carboxila livre, é o resíduo **carboxiterminal** (*C*-terminal).

**CONVENÇÃO-CHAVE:** Quando uma sequência de aminoácidos de um peptídeo, polipeptídeo ou proteína é exibida, a extremidade aminoterminal é localizada à esquerda e a extremidade carboxiterminal à direita. A sequência é lida da esquerda para a direita, começando com a extremidade aminoterminal. ■



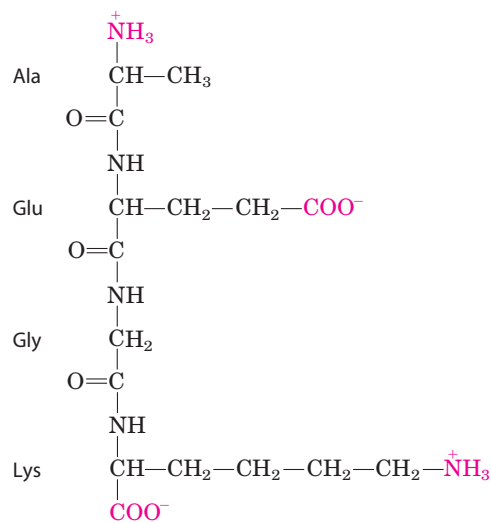
**FIGURA 3-14** O pentapeptídeo seril-glicil-tirosil-alanil-leucina, Ser-Gly-Tyr-Ala-Leu, ou SGYAL.

Os peptídeos são nomeados a partir do resíduo aminoterminal, que por convenção é colocado à esquerda. As ligações peptídicas são sombreadas; os grupos R estão em cor-de-rosa.

Embora a hidrólise de uma ligação peptídica seja uma reação exergônica, ela só ocorre lentamente porque tem uma elevada energia de ativação (p. 27). Como resultado, as ligações peptídicas em proteínas são muito estáveis, com meia-vida média ( $t_{1/2}$ ) de cerca de 7 anos na maioria das condições intracelulares.

### Peptídeos podem ser diferenciados por seus comportamentos de ionização

Peptídeos contêm apenas um grupo  $\alpha$ -amino e um grupo  $\alpha$ -carboxila livres, em extremidades opostas da cadeia (**Figura 3-15**). Esses grupos se ionizam como nos aminoácidos livres, embora as constantes de ionização sejam diferentes porque um grupo de carga oposta não é mais ligado ao carbono  $\alpha$ . Os grupos  $\alpha$ -amino e  $\alpha$ -carboxila de todos os aminoácidos não terminais são ligados covalentemente nas ligações peptídicas, que não se ionizam e, portanto, não contribuem para o comportamento ácido-básico total dos peptídeos. Entretanto, os grupos R de



**FIGURA 3-15** Alanil-glutamil-glicil-lisina. Este tetrapeptídeo tem um

grupo  $\alpha$ -amino livre, um grupo  $\alpha$ -carboxila livre e dois grupos R ionizáveis. Os grupos ionizados em pH 7,0 estão em cor-de-rosa.

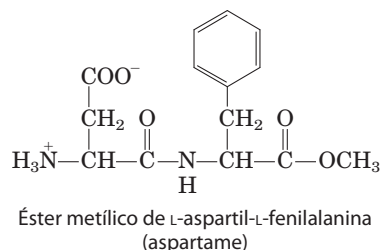


alguns aminoácidos podem se ionizar (Tabela 3-1), e em um peptídeo, esses contribuem para as propriedades acidobásicas gerais da molécula (Figura 3-15). Assim, o comportamento acidobásico de um peptídeo pode ser previsto a partir de seus grupos  $\alpha$ -amino e  $\alpha$ -carboxila livres combinado com a natureza e o número de seus grupos R ionizáveis.

Como os aminoácidos livres, os peptídeos têm curvas de titulação características e um pH isoeletrico característico (pI) que não se desloca em um campo isoeletrico. Essas propriedades são exploradas em algumas das técnicas utilizadas para separar peptídeos e proteínas, como será visto mais adiante neste capítulo. Deve ser enfatizado que o valor do  $pK_a$  para um grupo R ionizável pode se alterar um pouco quando um aminoácido se torna um resíduo em um peptídeo. A perda da carga nos grupos  $\alpha$ -carboxila e  $\alpha$ -amino, as interações com outros grupos R do peptídeo e outros fatores ambientais podem afetar o  $pK_a$ . Os valores de  $pK_a$  para os grupos R listados na Tabela 3-1 podem ser um guia útil para a variação do pH em que um determinado grupo irá se ionizar, mas eles não podem ser estritamente aplicados aos peptídeos.

### Peptídeos e polipeptídeos biologicamente ativos ocorrem em uma ampla variação de tamanhos e composições

Nenhuma generalização pode ser feita sobre as massas moleculares de peptídeos e proteínas biologicamente ativos em relação às suas funções. Peptídeos que ocorrem naturalmente variam em comprimento de dois a muitos milhares de resíduos de aminoácidos. Mesmo os menores peptídeos podem ter efeitos biologicamente importantes. Considere o dipeptídeo sintetizado comercialmente éster metílico de L-aspartil-L-fenilalanina, o adoçante artificial mais conhecido como aspartame ou *NutraSweet*.



Muitos peptídeos pequenos exercem seus efeitos em concentrações muito baixas. Por exemplo, vários hormônios de vertebrados (Capítulo 23) são peptídeos pequenos. Esses incluem a ocitocina (nove resíduos de aminoácidos), secretada pela glândula neuro-hipófise, que estimula as contrações uterinas, e o fator de liberação de tireotropina (três resíduos), formado no hipotálamo e que estimula a liberação de outro hormônio, tireotropina, da glândula adeno-hipófise. Alguns venenos extremamente tóxicos de cogumelos, como a amanitina, também são peptídeos pequenos, assim como muitos antibióticos.

Quão longo é o comprimento das cadeias polipeptídicas em proteínas? Como a Tabela 3-2 mostra, os comprimentos variam consideravelmente. O citocromo *c* humano tem 104 resíduos de aminoácidos ligados em uma única cadeia; o quimotripsinogênio bovino tem 245 resíduos. No extremo está a titina, constituinte dos músculos de vertebrados, que tem aproximadamente 27.000 resíduos de aminoácidos e massa molecular de cerca de 3.000.000. A grande maioria das proteínas que ocorrem naturalmente é muito menor do que ela, contendo menos de 2.000 resíduos de aminoácidos.

Algumas proteínas consistem em apenas uma única cadeia polipeptídica, porém outras, chamadas de proteínas **multissubunidade**, têm dois ou mais polipeptídeos associados de modo não covalente (Tabela 3-2). As cadeias polipeptídicas individuais em uma proteína multissubunidade podem ser idênticas ou diferentes. Se pelo menos duas

**TABELA 3-2** Dados moleculares de algumas proteínas

	Massa molecular	Número de resíduos	Número de cadeias polipeptídicas
Citocromo <i>c</i> (humano)	12.400	104	1
Ribonuclease A (pâncreas bovino)	13.700	124	1
Lisozima (clara de ovo de galinha)	14.300	129	1
Mioglobina (coração de equinos)	16.700	153	1
Quimotripsina (pâncreas bovino)	25.200	241	3
Quimotripsinogênio (bovinos)	25.700	245	1
Hemoglobina (humana)	64.500	574	4
Albumina sérica (humana)	66.000	609	1
Hexocinase (levedura)	107.900	972	2
RNA-polimerase ( <i>E. coli</i> )	450.000	4.158	5
Apolipoproteína B (humana)	513.000	4.536	1
Glutamina-sintetase ( <i>E. coli</i> )	619.000	5.628	12
Titina (humana)	2.993.000	26.926	1

são idênticas, a proteína é chamada de **oligomérica**, e as unidades idênticas (consistindo em uma ou mais cadeias polipeptídicas) são chamadas de **protômeros**. A hemoglobina, por exemplo, tem quatro subunidades polipeptídicas: duas cadeias  $\alpha$  idênticas e duas cadeias  $\beta$  idênticas, todas as quatro mantidas unidas por interações não covalentes. Cada subunidade  $\alpha$  é pareada de modo idêntico com uma subunidade  $\beta$  dentro da estrutura dessa proteína multissubunidade, de modo que a hemoglobina pode ser considerada tanto um tetrâmero de quatro subunidades de polipeptídeos quanto um dímero de protômeros  $\alpha\beta$ .

Algumas proteínas contêm duas ou mais cadeias polipeptídicas ligadas covalentemente. Por exemplo, as duas cadeias polipeptídicas da insulina são unidas por ligações dissulfeto. Em tais casos, os polipeptídeos individuais não são considerados subunidades, mas são comumente chamados simplesmente de cadeias.

A composição de aminoácidos das proteínas também é muito variável. Os 20 aminoácidos comuns quase nunca ocorrem em quantidades iguais em uma proteína. Alguns aminoácidos podem ocorrer apenas uma vez ou estar au-

sentes em determinado tipo de proteína; outros podem ocorrer em grande número. A Tabela 3-3 mostra a composição de aminoácidos do citocromo *c* e do quimotripsinogênio bovino, o último sendo o precursor inativo da enzima digestiva quimotripsina. Essas duas proteínas, com funções muito diferentes, também diferem significativamente em números relativos de cada tipo de resíduo de aminoácido.

É possível calcular o número aproximado de resíduos de aminoácidos em uma simples proteína que não contenha quaisquer outros constituintes químicos dividindo a sua massa molecular por 110. Embora a massa molecular média dos 20 aminoácidos comuns seja de cerca de 138, os aminoácidos menores predominam na maioria das proteínas. Levando em conta as proporções nas quais os vários aminoácidos ocorrem em uma proteína média (Tabela 3-1; as médias são determinadas pela pesquisa da composição dos aminoácidos de mais de 1.000 proteínas diferentes), a massa molecular média dos aminoácidos de uma proteína é mais próxima de 128. Como uma molécula de água ( $M_r$ , 18) é removida para criar cada ligação peptídica, a massa molecular média de um resíduo de aminoácido em uma proteína é de cerca de  $128 - 18 = 110$ .

**TABELA 3-3** Composição de aminoácidos de duas proteínas

Aminoácido	Citocromo <i>c</i> bovino		Quimotripsinogênio bovino	
	Número de resíduos por molécula	Porcentagem do total*	Número de resíduos por molécula	Porcentagem do total*
Ala	6	6	22	9
Arg	2	2	4	1,6
Asn	5	5	14	5,7
Asp	3	3	9	3,7
Cys	2	2	10	4
Gln	3	3	10	4
Glu	9	9	5	2
Gly	14	13	23	9,4
His	3	3	2	0,8
Ile	6	6	10	4
Leu	6	6	19	7,8
Lys	18	17	14	5,7
Met	2	2	2	0,8
Phe	4	4	6	2,4
Pro	4	4	9	3,7
Ser	1	1	28	11,4
Thr	8	8	23	9,4
Trp	1	1	8	3,3
Tyr	4	4	4	1,6
Val	3	3	23	9,4
Total	104	102	245	99,7

**Nota:** Em algumas análises usuais, como a hidrólise ácida, Asp e Asn não são distinguidos um do outro, sendo designados em conjunto como Asx (ou B). De forma semelhante, quando Glu e Gln não podem ser distinguidos, eles são designados juntos como Glx (ou Z). Adicionalmente, Trp é destruído por hidrólise ácida. Métodos adicionais devem ser utilizados para se obter uma avaliação precisa do conteúdo completo de aminoácidos.

\*Porcentagens não somam 100% em razão de arredondamentos.



**TABELA 3-4** Proteínas conjugadas

Classe	Grupo prostético	Exemplo
Lipoproteínas	Lipídeos	$\beta_1$ -Lipoproteína sanguínea
Glicoproteínas	Carboidratos	Imunoglobulina G
Fosfoproteínas	Grupos fosfato	Caseína do leite
Hemoproteínas	Heme (porfirina férrica)	Hemoglobina
Flavoproteínas	Nucleotídeos de flavina	Succinato-desidrogenase
Metaloproteínas	Ferro	Ferritina
	Zinco	Álcool-desidrogenase
	Cálcio	Calmodulina
	Molibdênio	Dinitrogenase
	Cobre	Plastocianina

### Algumas proteínas contêm outros grupos químicos além dos aminoácidos

Muitas proteínas, como, por exemplo, as enzimas ribonuclease A e a quimotripsina, contêm apenas resíduos de aminoácidos e nenhum outro constituinte químico; elas são consideradas proteínas simples. Entretanto, algumas proteínas contêm componentes químicos permanentemente associados além dos aminoácidos; elas são chamadas de **proteínas conjugadas**. A parte não aminoácido de uma proteína conjugada é normalmente chamada de **grupo prostético**. As proteínas conjugadas são classificadas com base na natureza química de seus grupos prostéticos (Tabela 3-4); por exemplo, **lipoproteínas** contêm lipídeos, **glicoproteínas** contêm grupos de açúcares e **metaloproteínas** contêm um metal específico. Algumas proteínas contêm mais de um grupo prostético. Normalmente o grupo prostético desempenha um papel importante na função biológica da proteína.

#### RESUMO 3.2 Peptídeos e proteínas

- ▶ Aminoácidos podem ser unidos de modo covalente por meio de ligações peptídicas para formar peptídeos e proteínas. As células geralmente contêm milhares de proteínas diferentes, cada uma com uma atividade biológica diferente.
- ▶ Proteínas podem ser cadeias peptídicas muito longas de 100 a muitos milhares de resíduos de aminoácidos. Entretanto, alguns peptídeos que ocorrem naturalmente possuem apenas alguns poucos resíduos de aminoácidos. Algumas proteínas são compostas por várias cadeias polipeptídicas associadas de modo não covalente, chamadas de subunidades.

Proteínas simples produzem, por hidrólise, apenas aminoácidos; proteínas conjugadas contêm além deles, alguns outros componentes, tais como um metal ou um grupo prostético.

Início leitura complementar aula 4

### 3.3 Trabalhando com proteínas

A compreensão da estrutura e função de proteínas pelos bioquímicos derivou de estudos de muitas proteínas indivi-

duais. Para estudar em detalhe uma proteína, o pesquisador deve ser capaz de separá-la de outras proteínas na forma pura e deve dominar as técnicas para determinar suas propriedades. Os métodos necessários vêm da química de proteínas, disciplina tão antiga quanto a própria bioquímica e que mantém uma posição central na pesquisa bioquímica.

### Proteínas podem ser separadas e purificadas

Uma preparação pura é essencial para a determinação das propriedades e atividades de uma proteína. Visto que as células contêm milhares de diferentes tipos de proteínas, como uma proteína pode ser purificada? Métodos clássicos para separação de proteínas se aproveitam das propriedades que variam de uma proteína para outra, incluindo o tamanho, a carga e as propriedades de ligação. Eles foram complementados nas últimas décadas por outros métodos, envolvendo a clonagem do DNA e o sequenciamento do genoma, que podem simplificar o processo de purificação de proteínas. Os métodos mais recentes, apresentados no Capítulo 9, frequentemente modificam artificialmente a proteína que está sendo purificada, adicionando poucos ou muitos resíduos de aminoácidos a uma ou ambas as extremidades. A conveniência, portanto, paga o preço de alterar potencialmente a atividade da proteína purificada. A purificação de proteínas em seus estados nativos (a forma como funcionam nas células) depende geralmente dos métodos descritos aqui.

A fonte de uma proteína é geralmente um tecido ou uma célula microbiana. A primeira etapa de qualquer procedimento de purificação de proteína é romper essas células, liberando suas proteínas em uma solução chamada de **extrato bruto**. Se necessário, pode ser utilizada centrifugação diferencial para preparar frações subcelulares ou para isolar organelas específicas (ver Figura 1-8).

Uma vez prontos o extrato ou a preparação de organelas, vários métodos estão disponíveis para purificar uma ou mais das proteínas neles contidas. Em geral, o extrato é submetido a tratamentos para separar as proteínas em diferentes **frações** com base em uma propriedade, tal como tamanho ou carga, em um processo chamado de **fracionamento**. Etapas iniciais de fracionamento em uma purificação utilizam diferenças na solubilidade de proteínas, que são uma função complexa do pH, temperatura, concentra-

Término leituras complementar e avançada aula 3

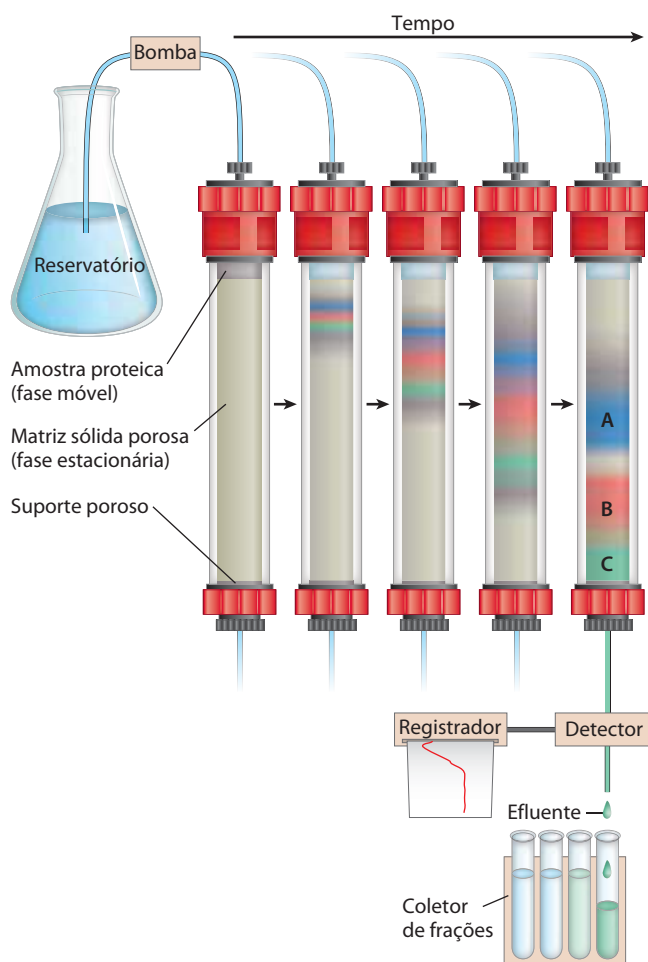
ção de sais e outros fatores. A solubilidade de proteínas é reduzida em presença de alguns sais, um efeito chamado de *salting out*. A adição de certos sais na quantidade correta pode precipitar seletivamente algumas proteínas, enquanto outras permanecem em solução. Particularmente eficaz, o sulfato de amônio ((NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>) é muitas vezes utilizado para precipitar proteínas. As proteínas assim precipitadas são removidas daquelas que permanecem em solução por centrifugação em baixa rotação.

Uma solução contendo a proteína de interesse geralmente precisa ser modificada adicionalmente antes que as etapas de purificação subsequentes sejam possíveis. Por exemplo, a **diálise** é um procedimento que separa proteínas de solutos pequenos se aproveitando do tamanho maior das proteínas. O extrato parcialmente purificado é colocado em uma bolsa ou tubo composto por uma membrana semipermeável. Quando este é suspenso em um volume muito maior de uma solução tamponada de força iônica adequada, a membrana permite a troca de sal e de solução tampão, mas não de proteínas. Assim, a diálise retém as proteínas grandes no interior da bolsa membranosa ou tubo, permitindo que a concentração de outros solutos na preparação de proteínas se altere até ficarem em equilíbrio com a solução fora da membrana. A diálise pode ser utilizada, por exemplo, para remover o sulfato de amônio da preparação proteica.

Os métodos mais eficientes para fracionar proteínas utilizam a **cromatografia em coluna**, que se utiliza das diferenças na carga das proteínas, tamanho, afinidade de ligação e outras propriedades (Figura 3-16). Um material sólido poroso com propriedades químicas adequadas (fase estacionária) é mantido em uma coluna, e uma solução tamponada (fase móvel) migra através dela. A proteína, dissolvida na mesma solução tampão que foi utilizada para estabelecer a fase móvel, é colocada no topo da coluna. A proteína então atravessa a matriz sólida como uma banda que se expande cada vez mais no interior da fase móvel maior. Proteínas individuais migram com mais rapidez ou lentidão através da coluna, dependendo de suas propriedades.

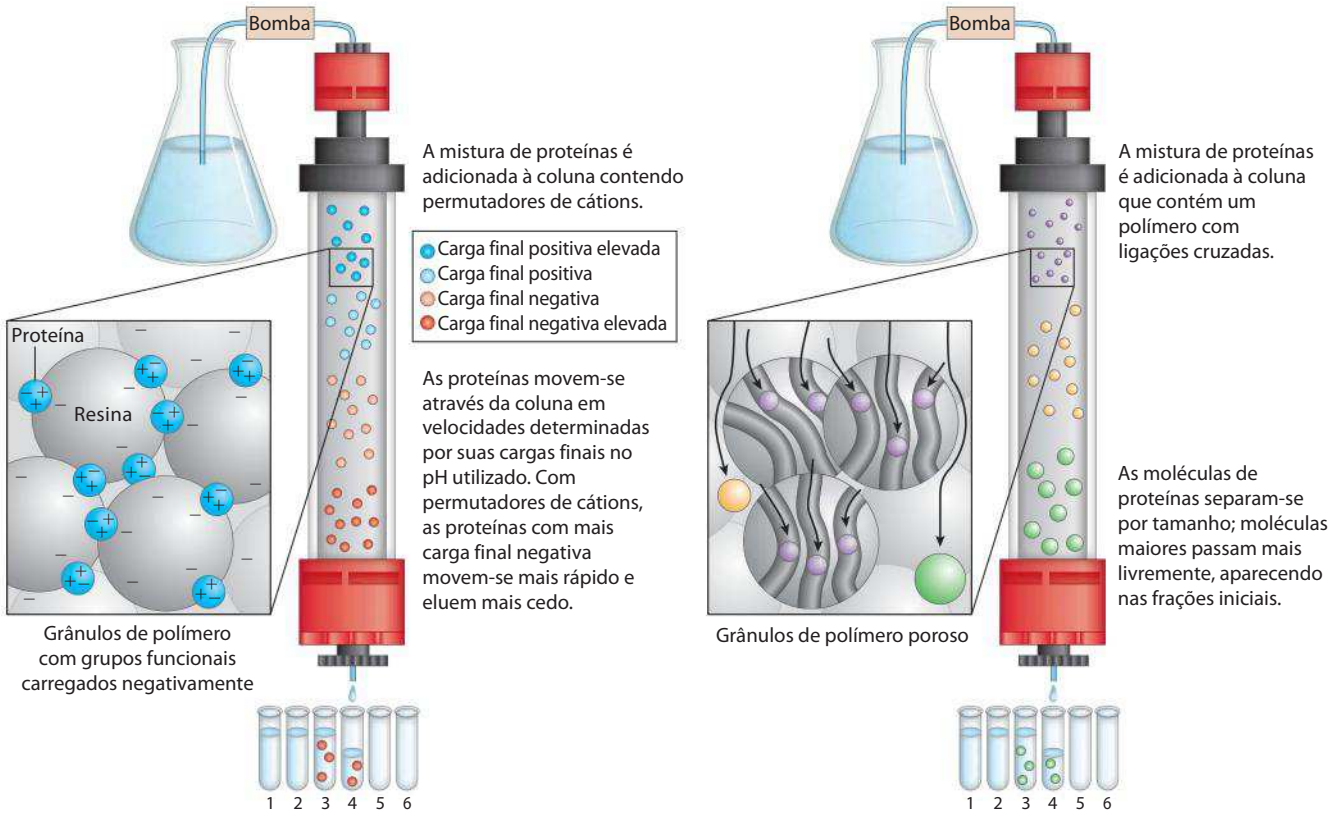
A **cromatografia de troca iônica** explora as diferenças no sinal e magnitude da carga elétrica final de proteínas em um determinado pH (Figura 3-17a). A matriz da coluna é um polímero sintético (resina) que contém grupos carregados ligados; aqueles ligados a grupos aniônicos são chamados de **permutadores de cátions**, e aqueles ligados a grupos catiônicos são chamados de **permutadores de ânions**. A afinidade de cada proteína pelos grupos carregados na coluna é afetada pelo pH (que determina o estado de ionização da molécula) e a concentração de íons de sais livres competindo na solução circundante. A separação pode ser otimizada por mudanças graduais no pH e/ou na concentração de sal da fase móvel de modo a criar um gradiente de pH ou de sal. Na **cromatografia de troca catiônica**, a matriz sólida tem grupos carregados negativamente. Na fase móvel, as proteínas com uma carga final positiva migram através da matriz mais lentamente que aquelas proteínas com uma carga final negativa, porque a migração das primeiras é mais retardada por sua interação com a fase estacionária.

Em colunas de troca iônica, a expansão da banda de proteína na fase móvel (a solução proteica) é causada



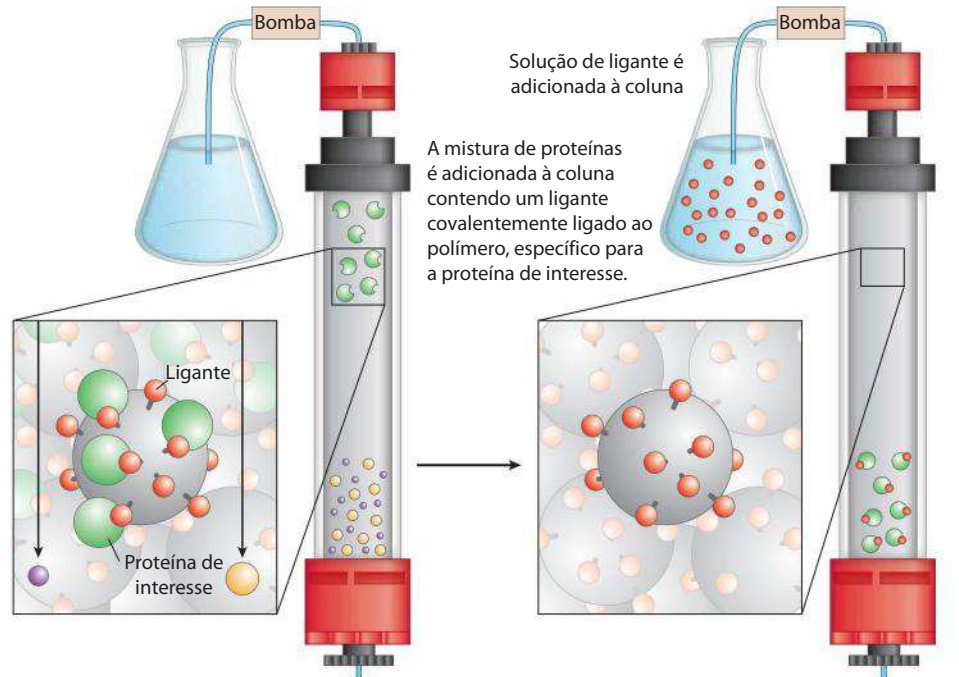
**FIGURA 3-16 Cromatografia em coluna.** Os elementos padrão de uma coluna cromatográfica incluem um material poroso (matriz) sólido apoiado no interior de uma coluna, geralmente feita de plástico ou vidro. Uma solução, a fase móvel, flui através da matriz, a fase estacionária. A solução que sai da coluna (o efluente) é constantemente substituída pela solução fornecida por um reservatório no topo. A solução de proteína a ser separada é colocada no topo da coluna e deixada percolar pela matriz sólida. Mais solução é adicionada no topo. A solução proteica forma uma banda no interior da fase móvel que tem inicialmente a profundidade da solução de proteína aplicada à coluna. À medida que as proteínas migram através da coluna (mostrada aqui em cinco momentos diferentes), elas são retardadas em diferentes graus por suas diferentes interações com o material da matriz. A banda total de proteína, portanto, se amplia à medida que se move através da coluna. Tipos individuais de proteínas (como A, B e C, mostradas em azul, vermelho e verde) se separam gradualmente umas das outras, formando bandas no interior da banda proteica mais larga. A separação melhora (i.e., aumenta a resolução) à medida que o comprimento da coluna aumenta. Entretanto, cada banda proteica individual também se alarga com o tempo devido à dispersão por difusão, processo que diminui a resolução. Nesse exemplo, a proteína A está bem separada da B e C, mas a dispersão por difusão impede a separação completa de B e C sob essas condições.

tanto pela separação de proteínas com diferentes propriedades quanto pela dispersão por difusão. À medida que o comprimento da coluna aumenta, a resolução de dois tipos de proteínas com diferentes cargas finais geralmente melhora. Entretanto, a velocidade na qual a solução proteica pode fluir através da coluna geralmente diminui com o comprimento da coluna. E, à medida que a duração do



(a) Cromatografia de troca iônica

(b) Cromatografia de exclusão por tamanho



1 2 3 4 5 6 7 8 9  
Proteínas não desejadas são lavadas da coluna.

1 2 3 4 5 6 7 8 9  
A proteína de interesse é eluída pela solução de ligante.

(c) Cromatografia de afinidade

**FIGURA 3-17** Três métodos cromatográficos usados na purificação de proteínas. (a) A cromatografia de troca iônica explora diferenças no sinal e na magnitude das cargas elétricas finais de proteínas em um determinado pH. (b) A cromatografia de exclusão por tamanho, também chamada de filtração em gel, separa proteínas de acordo com o tamanho. (c) A cromatografia de afinidade separa proteínas por suas especificidades de ligação. Detalhes adicionais desses métodos são fornecidos no texto.



tempo dispendido na coluna aumenta, a resolução pode diminuir como resultado da dispersão por difusão no interior de cada banda proteica. À medida que o conteúdo da solução proteica sai de uma coluna, porções sucessivas (frações) desse efluente são coletadas em tubos de ensaio. Cada fração pode ser testada para a presença da proteína de interesse, assim como outras propriedades, tais como a força iônica ou a concentração total de proteínas. Todas as frações positivas para a proteína de interesse podem ser reunidas como o produto dessa etapa cromatográfica da purificação de proteínas.

### PROBLEMA RESOLVIDO 3-1 Troca iônica de peptídeos

Um bioquímico deseja separar dois peptídeos por cromatografia por troca iônica. No pH da fase móvel a ser utilizado na coluna, um peptídeo (A) possui uma carga final de  $-3$ , em decorrência da presença de mais resíduos de Glu e Asp do que de Arg, Lys e His. O peptídeo B tem carga final de  $+1$ . Qual peptídeo irá eluir primeiro a partir de uma resina de troca catiônica? Qual irá eluir primeiro a partir da resina de troca aniônica?

**Solução:** Uma resina de troca catiônica possui cargas negativas e se liga a moléculas carregadas positivamente, retardando seu progresso pela coluna. O peptídeo B, com sua carga final positiva, interagirá mais fortemente do que o peptídeo A com a resina de troca catiônica e, portanto, o peptídeo A irá eluir primeiro. Na resina de troca aniônica, o peptídeo B irá eluir primeiro. O peptídeo A, sendo carregado negativamente, será retardado por sua interação com a resina positivamente carregada.

A Figura 3-17 mostra duas outras variações da cromatografia em coluna além da troca iônica. A **cromatografia de exclusão por tamanho**, também chamada de filtração em gel (Figura 3-17b), separa as proteínas de acordo com o tamanho. Neste método, as proteínas grandes emergem da coluna mais cedo do que as proteínas menores – resultado um tanto contrário ao esperado intuitivamente. A fase sólida consiste em grânulos de polímeros reticulados com poros ou cavidades projetados com um determinado tamanho. As proteínas grandes não podem entrar nas cavidades e, assim, tomam um caminho mais curto (e mais rápido) através da coluna, ao redor dos grânulos. Proteínas pequenas penetram nas cavidades e são retardadas em seu caminho de labirintos através da coluna. A cromatografia de exclusão por tamanho também pode ser utilizada para estimar o tamanho de uma proteína que está sendo purificada, utilizando métodos semelhantes aos descritos na Figura 3-19.

A **cromatografia de afinidade** se baseia na afinidade de ligação (Figura 3-17c). Os grânulos na coluna têm um grupo químico covalentemente ligado chamado de ligante – um grupo ou molécula que se liga a uma macromolécula, tal como uma proteína. Quando uma mistura de proteínas é adicionada à coluna, qualquer proteína com afinidade para esse ligante se liga aos grânulos, e sua migração através da matriz é retardada. Por exemplo, se a função biológica de uma proteína envolve a ligação ao ATP, então, ligando-se uma molécula que se assemelha ao ATP a esses grânulos

na coluna cria-se uma matriz de afinidade que pode ajudar a purificar a proteína. À medida que a solução proteica se desloca através da coluna, as proteínas ligadoras de ATP (incluindo a proteína de interesse) se ligam à matriz. Após a lavagem das proteínas que não se ligam na coluna, a proteína ligada é eluída por uma solução contendo uma alta concentração de sal ou um ligante livre – nesse caso o ATP ou um análogo do ATP. O sal enfraquece a ligação da proteína ao ligante imobilizado, interferindo com as interações iônicas. O ligante livre compete com o ligante ligado aos grânulos, liberando a proteína da matriz; o produto proteico que elui da coluna é com frequência ligado ao ligante utilizado para eluí-lo.

Métodos cromatográficos são aperfeiçoados com a utilização de **HPLC**, ou **cromatografia líquida de alto desempenho**. A HPLC faz uso de bombas de alta pressão que aceleram o movimento das moléculas de proteína coluna abaixo, bem como materiais cromatográficos de maior qualidade que podem suportar a força de esmagamento do fluxo pressurizado. Reduzindo o tempo de trânsito na coluna, a HPLC pode limitar a dispersão por difusão das bandas proteicas e, assim, melhorar muito a resolução.

A abordagem para purificação de uma proteína que não tenha sido previamente isolada é guiada tanto pelos precedentes estabelecidos quanto pelo senso comum. Na maioria dos casos, vários métodos diferentes devem ser utilizados sequencialmente para purificar uma proteína completamente, cada método separando as proteínas com base em propriedades diferentes. Por exemplo, se uma etapa separa as proteínas ligadoras de ATP daquelas que não se ligam a ele, então, a próxima etapa deve separar as várias proteínas ligadoras de ATP com base no tamanho ou na carga para isolar a proteína específica que é desejada. A escolha dos métodos é um tanto empírica, e muitas estratégias podem ser tentadas antes que a mais eficaz seja encontrada. Tentativas e erros podem ser frequentemente minimizados baseando-se o novo procedimento em técnicas de purificação desenvolvidas para proteínas semelhantes. Protocolos de purificação publicados estão disponíveis para muitos milhares de proteínas. O senso comum determina que procedimentos mais baratos, tal como o *salting out*, devam ser utilizados primeiro, quando o volume total e o número de contaminantes são maiores. Métodos cromatográficos são frequentemente impraticáveis nas fases iniciais porque a quantidade de meio cromatográfico necessário aumenta com o tamanho da amostra. À medida que cada etapa de purificação se completa, o tamanho da amostra geralmente se torna menor (Tabela 3-5), tornando possível utilizar procedimentos cromatográficos mais sofisticados (e caros) em fases posteriores.

### Proteínas podem ser separadas e caracterizadas por eletroforese

Outra técnica importante para separação de proteínas se baseia na migração de proteínas carregadas em um campo elétrico, um processo chamado de **eletroforese**. Em geral, esses procedimentos não são utilizados para purificar proteínas, pois alternativas mais simples estão disponíveis e métodos eletroforéticos com frequência afetam adversa-

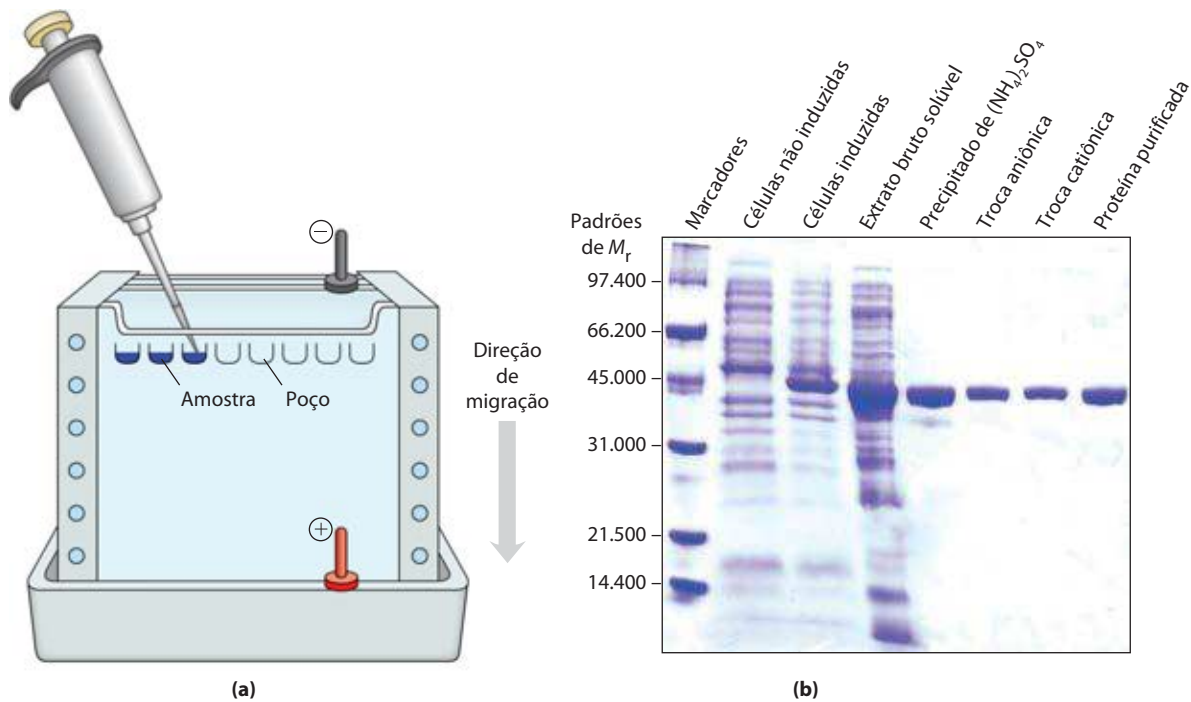
**TABELA 3-5** Tabela de purificação para uma enzima hipotética

Procedimento ou etapa	Volume da fração (mL)	Proteína total (mg)	Atividade (unidades)	Atividade específica (unidades/mg)
1. Extrato celular bruto	1.400	10.000	100.000	10
2. Precipitação com sulfato de amônio	280	3.000	96.000	32
3. Cromatografia de troca iônica	90	400	80.000	200
4. Cromatografia de exclusão por tamanho	80	100	60.000	600
5. Cromatografia de afinidade	6	3	45.000	15.000

**Nota:** Todos os dados representam o estado da amostra após a realização do procedimento designado. A atividade e a atividade específica são definidas na página 95.

mente a estrutura e, desse modo, a função das proteínas. Entretanto, como um método analítico, a eletroforese é extremamente importante. Sua vantagem é que as proteínas podem ser visualizadas, bem como separadas, permitindo ao pesquisador estimar rapidamente o número de proteínas diferentes em uma mistura ou o grau de pureza de uma preparação proteica específica. A eletroforese também pode ser utilizada para determinar propriedades cruciais de uma proteína, tal como seu ponto isoelétrico, e estimar sua massa molecular.

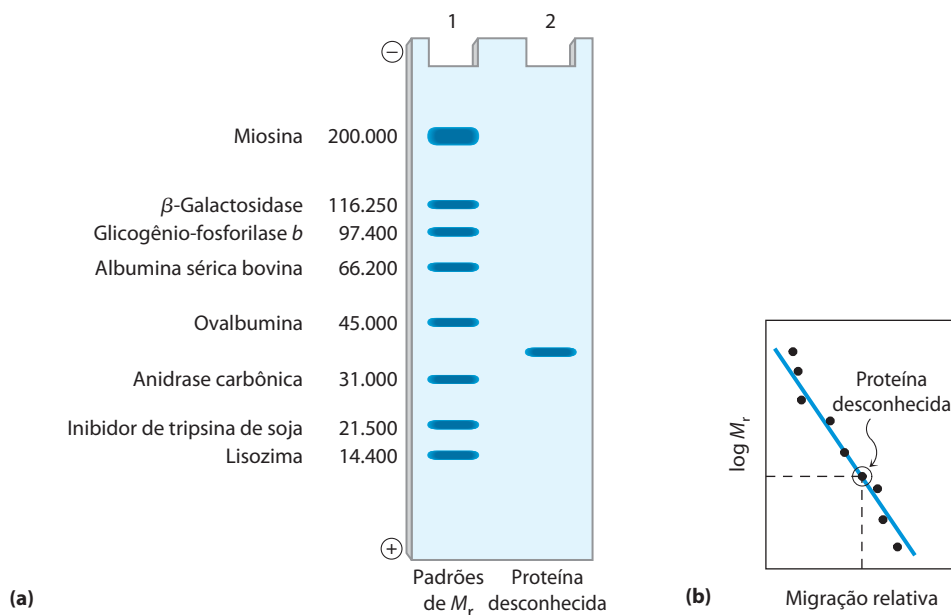
Em geral, a eletroforese de proteínas é realizada em géis compostos de polímeros reticulados de poli(acrilamida) (**Figura 3-18**). O gel de poli(acrilamida) age como uma peneira molecular, retardando a migração de proteínas aproximadamente em proporção à sua razão carga-massa. A migração também pode ser afetada pela forma da proteína. Na eletroforese, a força que move a macromolécula é o potencial elétrico  $E$ . A mobilidade eletroforética,  $\mu$ , de uma molécula é a razão de sua velocidade  $V$ , em relação ao seu potencial elétrico. A mobilidade eletroforética é também igual à carga



**FIGURA 3-18 Eletroforese.** (a) Diferentes amostras são colocadas em poços ou depressões no topo do gel de SDS-poli(acrilamida). As proteínas se movem para o gel quando um campo elétrico é aplicado. O gel minimiza as correntes de convecção causadas pelos pequenos gradientes de temperatura, bem como movimentos proteicos além daqueles induzidos pelo campo elétrico. (b) Proteínas podem ser visualizadas após eletroforese tratando o gel com um corante como o azul Coomassie, que se liga às proteínas, mas não ao gel em si. Cada banda no gel representa uma proteína diferente (ou subunidade de proteína); proteínas menores se movem através do gel mais rapidamente que as maiores e, portanto, são encontradas mais

próximas da base do gel. Esse gel ilustra a purificação da proteína RecA de *Escherichia coli* (descrita no Capítulo 25). O gene para a proteína RecA foi clonado (Capítulo 9) para que sua expressão (síntese da proteína) pudesse ser controlada. A primeira canaleta mostra um conjunto de proteínas padrão (de  $M_r$  conhecido), servindo como marcadores de massa molecular. As duas canaletas seguintes mostram proteínas de células de *E. coli* antes e depois que a síntese da proteína RecA foi induzida. A quarta canaleta mostra as proteínas presentes após sucessivas etapas de purificação. A proteína purificada é uma cadeia polipeptídica única ( $M_r$ , 38.000), como mostrado na canaleta mais à direita.





**FIGURA 3-19** Estimando a massa molecular de uma proteína. A mobilidade eletroforética de uma proteína em gel de SDS-poliacrilamida está relacionada à sua massa molecular,  $M_r$ . **(a)** Proteínas-padrão de massa molecular conhecida são sujeitas à eletroforese (calha 1). Estas proteínas marcadoras podem ser usadas para estimar a massa molecular de uma proteína desconhecida (calha 2). **(b)** Um gráfico do  $\log M_r$  das proteínas marcadoras versus migração relativa durante a eletroforese é linear, permitindo que

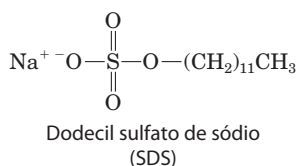
a massa molecular da proteína desconhecida seja lido a partir do gráfico. (De maneira semelhante, um conjunto de proteínas padrão com tempos de retenção reproduzíveis em uma coluna de exclusão por tamanho pode ser usado para criar uma curva padrão de tempo de retenção versus  $\log M_r$ . O tempo de retenção de uma substância desconhecida na coluna pode ser comparado com sua curva padrão para obter um  $M_r$  aproximado.)

final  $Z$  da molécula dividida por seu coeficiente de fricção  $f$ , que reflete em parte a forma de uma proteína. Portanto:

$$\mu = \frac{V}{E} = \frac{Z}{f}$$

A migração de uma proteína em um gel durante a eletroforese é, portanto, uma função do seu tamanho e formato.

Um método eletroforético comumente empregado para estimar a pureza e a massa molecular utiliza o detergente **dodecil sulfato de sódio (SDS)** (“dodecil” significa uma cadeia de 12 carbonos).

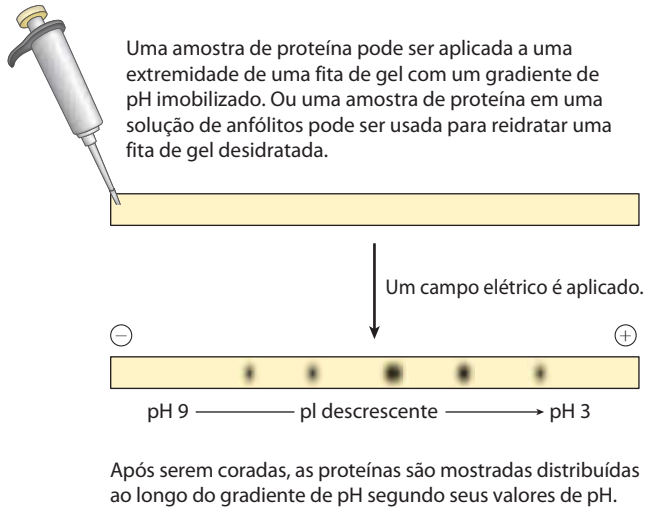


Uma proteína se ligará cerca de 1,4 vez sua massa de SDS, aproximadamente uma molécula de SDS para cada resíduo de aminoácido. Um SDS ligado contribui com uma grande carga final negativa, tornando a carga intrínseca da proteína insignificante e conferindo a cada proteína uma razão carga-massa semelhante. Além disso, a ligação de SDS desdobra parcialmente as proteínas, de modo que a maior parte das proteínas ligadas ao SDS assume uma forma semelhante a bastonetes. A eletroforese na presença de SDS, portanto, separa proteínas quase que exclusivamente com base em sua massa (massa molecular), com os peptídeos menores migrando mais rapidamente. Após a eletroforese, as proteínas são visualizadas pela adição de um corante, como o azul de Co-

massie, que se liga às proteínas, mas não ao gel em si (Figura 3-18b). Assim, um pesquisador pode monitorar o progresso de um procedimento de purificação de proteínas à medida que o número de bandas de proteínas visíveis no gel diminui após cada nova fase de fracionamento. Quando comparada às posições para as quais as proteínas de massa molecular conhecido migram no gel, a posição de uma proteína não identificada pode fornecer uma boa estimativa de sua massa molecular (Figura 3-19). Se a proteína tem duas ou mais subunidades diferentes, as subunidades são geralmente separadas por tratamento com SDS, e uma banda separada aparece para cada uma delas. **Eletroforese em gel com SDS**

A **focalização isoeétrica** é um procedimento utilizado para determinar o ponto isoeétrico (pI) de uma proteína (Figura 3-20). Um gradiente de pH é estabelecido permitindo-se que uma mistura de ácidos e bases orgânicas de baixo peso molecular (anfólitos; p. 81) se distribua em um campo elétrico gerado ao longo do gel. Quando uma mistura de proteínas é aplicada, cada proteína migra até alcançar o pH correspondente ao seu pI. Proteínas com pontos isoeétricos diferentes são, assim, distribuídas de modo diferente ao longo do gel.

A combinação da focalização isoeétrica com a eletroforese em SDS sequencialmente em um processo chamado de **eletroforese bidimensional** permite a resolução de misturas complexas de proteínas (Figura 3-21). Esse é um método analítico mais sensível do que qualquer método eletroforético sozinho. A eletroforese bidimensional separa proteínas de massa molecular idêntica que diferem em seu pI, ou proteínas com valores de pI semelhantes, mas com massas moleculares diferentes.

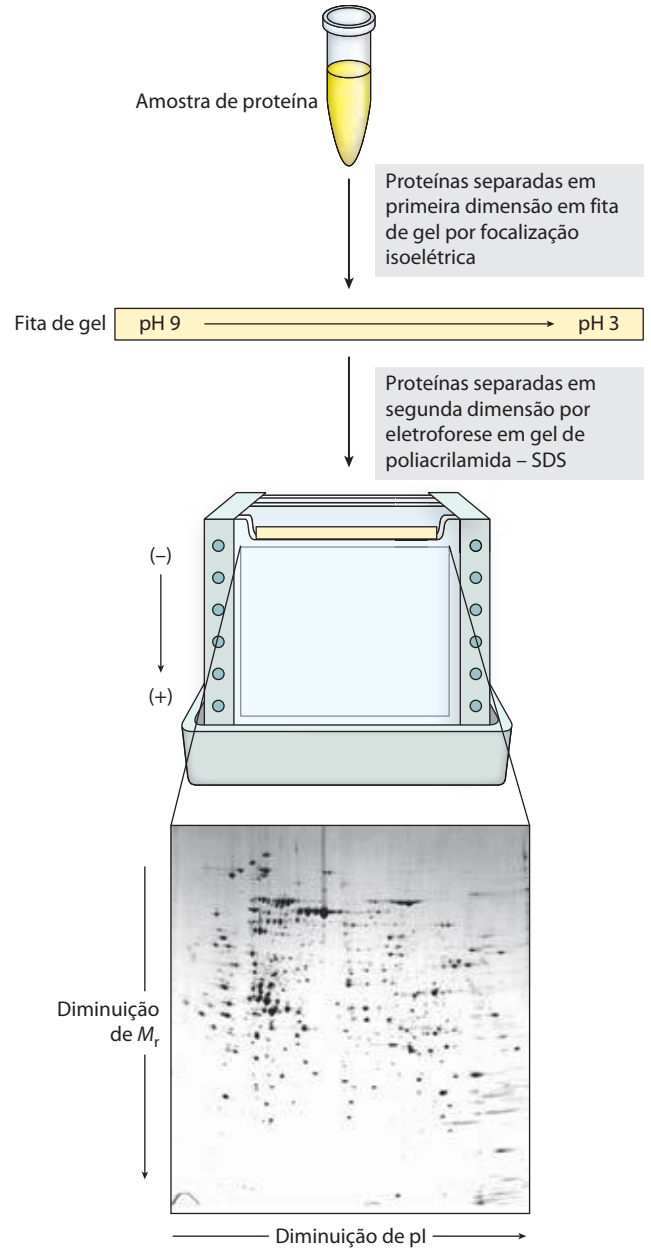


**FIGURA 3-20 Focalização isoeletrica.** Essa técnica separa proteínas de acordo com seus pontos isoelétricos. Uma mistura de proteínas é colocada em uma fita de gel contendo um gradiente de pH imobilizado. Com a aplicação de um campo elétrico, as proteínas entram no gel e migram até que cada uma atinja um pH equivalente ao seu pI. Lembre-se que quando o pH = pI, a carga final de uma proteína é zero.

**Proteínas não separadas podem ser quantificadas**

Para purificar uma proteína, é essencial possuir um meio para detectar e quantificar aquela proteína na presença de muitas outras proteínas em cada estágio do procedimento. Frequentemente, a purificação deve prosseguir na ausência de qualquer informação sobre o tamanho ou propriedades físicas da proteína ou sobre a fração da massa proteica total que ela representa no extrato. Para proteínas que são enzimas, a quantidade de uma determinada solução ou extrato de tecido pode ser medida ou ensaiada, em termos do efeito catalítico que a enzima produz – isto é, o *aumento* na taxa em que seu substrato é convertido para produtos de reação quando a enzima está presente. Para esse propósito, o pesquisador deve conhecer (1) a equação geral da reação catalisada, (2) um procedimento analítico para determinar o desaparecimento do substrato ou o aparecimento de um produto de reação, (3) se a enzima necessita de cofatores, como íons metálicos ou coenzimas, (4) a dependência da atividade enzimática da concentração do substrato, (5) o pH ótimo e (6) uma zona de temperatura em que a enzima é estável e possui alta atividade. Enzimas são geralmente analisadas em seu pH ótimo em alguma temperatura conveniente na faixa de 25 a 38°C. Altas concentrações de substrato também são geralmente utilizadas, de modo que a velocidade de reação inicial, medida experimentalmente, é proporcional à concentração da enzima (Capítulo 6).

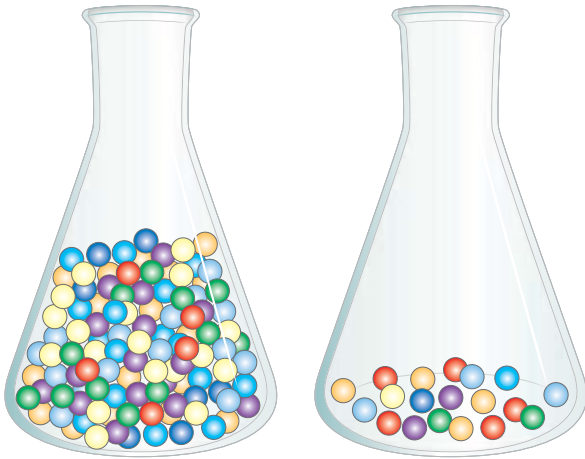
Por convenção internacional, a unidade 1,0 de atividade enzimática para a maior parte das enzimas é definida como a quantidade de enzima que leva à transformação de 1,0 µmol de substrato em produto, por minuto, a 25°C, sob condições ótimas de medição (para algumas enzimas, essa definição não é conveniente, e uma unidade pode ser definida diferentemente). O termo **atividade** se refere às unidades totais de enzima em uma solução. A **atividade específica** é o número de unidades de enzimas por miligra-



**FIGURA 3-21 Eletroforese bidimensional.** Em primeiro lugar, as proteínas são separadas por focalização isoeletrica em uma fita de gel fina. O gel é colocado então horizontalmente em um segundo gel em forma de placa, e as proteínas são separadas por eletroforese em gel de poliacrilamida-SDS. A separação horizontal reflete diferenças no pI; a separação vertical reflete diferenças na massa molecular. O complemento proteico original é, deste modo, espalhado em duas dimensões. Milhares de proteínas celulares podem ser resolvidas usando essa técnica. Manchas de proteínas individuais podem ser cortadas do gel e identificadas por espectrometria de massa (ver Figuras 3-30 e 3-31).

ma de proteína total (**Figura 3-22**). A atividade específica é uma medida de pureza enzimática: ela aumenta durante a purificação de uma enzima e se torna máxima e constante quando a enzima é pura (Tabela 3-5, p. 93).

Após cada etapa de purificação, a atividade da preparação (em unidades de atividade enzimática) é analisada, a quantidade total de proteína é determinada independente-



**FIGURA 3-22 Atividade versus atividade específica.** A diferença entre esses termos pode ser ilustrada considerando dois béqueres contendo esferas. Os béqueres contêm o mesmo número de esferas vermelhas, mas números diferentes de esferas de outras cores. Se as esferas representam proteínas, ambos os béqueres contêm a mesma *atividade* da proteína representada pelas esferas vermelhas. O segundo béquer, no entanto, apresenta a *atividade específica* maior porque as esferas vermelhas representam uma fração mais alta do total.

mente e a razão das duas fornece a atividade específica. A atividade e a proteína total geralmente diminuem em cada etapa. A atividade diminui porque há sempre alguma perda em consequência da inativação ou interações não ideais com materiais cromatográficos ou outras moléculas na solução. A proteína total diminui porque o objetivo é remover o máximo possível de proteína inespecífica e indesejada. Em uma etapa bem-sucedida, a perda de proteína inespecífica é muito maior que a perda de atividade; portanto, a atividade específica aumenta mesmo que a atividade total decaia. Os dados estão reunidos em uma tabela de purificação semelhante à Tabela 3-5. Em geral, uma proteína é considerada pura quando etapas de purificação adicionais não conseguem aumentar a atividade específica e quando apenas uma única espécie de proteína pode ser detectada (p. ex., por eletroforese).

Para proteínas não enzimas, outros métodos de quantificação são necessários. Proteínas de transporte podem ser

analisadas pela sua ligação à molécula que elas transportam e hormônios e toxinas pelo efeito biológico que produzem; por exemplo, hormônios de crescimento irão estimular o crescimento de certas células em cultura. Algumas proteínas estruturais representam uma grande fração de uma massa tecidual a ponto de ela ser rapidamente extraída e purificada sem um ensaio funcional. As abordagens são tão variadas quanto as próprias proteínas.

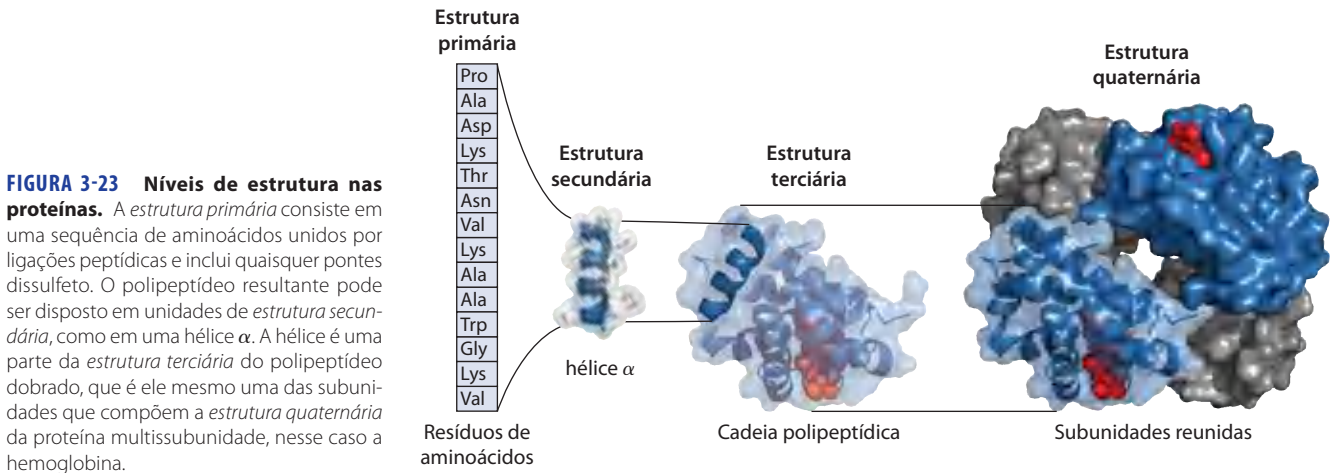
**RESUMO 3.3 Trabalhando com proteínas**

- ▶ Proteínas são separadas e purificadas com base nas diferenças de suas propriedades. Proteínas podem ser seletivamente precipitadas por mudanças no pH ou temperatura e, particularmente, pela adição de certos sais. Uma ampla gama de procedimentos cromatográficos faz uso de diferenças de tamanho, afinidades de ligação, carga e outras propriedades. Essas incluem a troca iônica, a exclusão por tamanho, a afinidade e a cromatografia líquida de alto desempenho.
- ▶ Eletroforese separa proteínas com base na massa ou carga. A eletroforese em gel SDS e a focalização isoeletrica podem ser utilizadas separadamente ou em combinação para uma resolução mais alta.
- ▶ Todos os procedimentos de purificação exigem um método para quantificação ou análise da proteína de interesse na presença de outras proteínas. A purificação pode ser monitorada por análise da atividade específica.

**3.4 A estrutura de proteínas: estrutura primária**

A purificação de uma proteína é geralmente apenas um prelúdio para uma dissecação bioquímica detalhada de sua estrutura e função. O que torna uma proteína uma enzima, outra um hormônio, outra uma proteína estrutural e ainda outra um anticorpo? Como elas diferem quimicamente? As distinções mais óbvias são estruturais, e agora será abordada a estrutura das proteínas.

A estrutura de grandes moléculas, tais como proteínas, pode ser descrita em vários níveis de complexidade, arranjada em um tipo de hierarquia conceitual. Quatro níveis de estrutura proteica são comumente definidos (**Figura 3-23**). Uma descrição de todas as ligações covalentes



(principalmente ligações peptídicas e ligações dissulfeto) ligando resíduos de aminoácidos em uma cadeia polipeptídica é a sua **estrutura primária**. O elemento mais importante da estrutura primária é a *sequência* de resíduos de aminoácidos. A **estrutura secundária** se refere a arranjos particularmente estáveis de resíduos de aminoácidos dando origem a padrões estruturais recorrentes. A **estrutura terciária** descreve todos os aspectos do enovelamento tridimensional de um polipeptídeo. Quando uma proteína tem duas ou mais subunidades polipeptídicas, seus arranjos no espaço são chamados de **estrutura quaternária**. Nossa exploração de proteínas, por fim, inclui máquinas proteicas complexas que consistem em dezenas de milhares de subunidades. A estrutura primária é o foco do restante deste capítulo; os níveis mais elevados de estrutura são discutidos no Capítulo 4.

As diferenças na estrutura primária podem ser especialmente informativas. Cada proteína tem um número e uma sequência de resíduos de aminoácidos distintos. Como será visto no Capítulo 4, a estrutura primária de uma proteína determina como ela se dobra em sua estrutura tridimensional única, e isso, por sua vez, determina a função da proteína. Em primeiro lugar, serão considerados os indícios empíricos de que a sequência de aminoácidos e a função da proteína são intimamente ligadas; em seguida, será descrito como a sequência de aminoácidos é determinada; e, finalmente, serão destacados os múltiplos usos dessas informações.

### A função de uma proteína depende de sua sequência de aminoácidos

A bactéria *Escherichia coli* produz mais de 3.000 proteínas diferentes; um ser humano tem ~25.000 genes que codificam um número muito maior de proteínas (por meio de processos genéticos discutidos na Parte III deste livro). Em ambos os casos, cada tipo de proteína possui uma sequência de aminoácidos única que confere uma determinada estrutura tridimensional. Essa estrutura, por sua vez, confere uma função específica.

Algumas observações simples ilustram a importância da estrutura primária ou a sequência de aminoácidos de uma proteína. Em primeiro lugar, como já observado, as proteínas com funções diferentes sempre possuem sequências de aminoácidos diferentes. Em segundo lugar, milhares de doenças genéticas humanas foram rastreadas para a produção de proteínas defeituosas. O defeito pode variar de uma simples troca na sequência de aminoácidos (como na anemia falciforme, descrita no Capítulo 5) à deleção de uma porção maior da cadeia polipeptídica (como na maior parte dos casos da distrofia muscular de Duchenne: uma grande deleção no gene que codifica a proteína distrofina leva à produção de uma proteína encurtada e inativa). Finalmente, comparando proteínas funcionalmente semelhantes de diferentes espécies, foi descoberto que essas proteínas frequentemente têm sequências de aminoácidos semelhantes. Portanto, uma ligação íntima entre a estrutura primária da proteína e sua função é evidente.

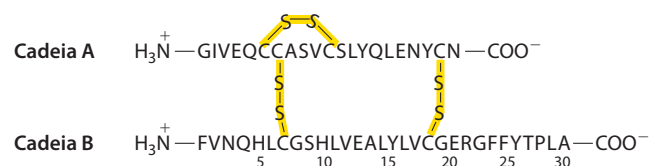
A sequência de aminoácidos é totalmente fixa ou invariável para uma determinada proteína? Não; alguma flexi-

bilidade é possível. Estima-se que 20 a 30% das proteínas humanas sejam **polimórficas**, possuindo variações nas sequências de aminoácidos na população humana. Muitas dessas variações na sequência têm pouco ou nenhum efeito na função da proteína. Além disso, proteínas que desempenham funções muito semelhantes em espécies distantemente relacionadas podem ser muito diferentes no tamanho geral e na sequência de aminoácidos.

Embora a sequência de aminoácidos em algumas regiões da estrutura primária possa variar consideravelmente sem afetar a função biológica, a maior parte das proteínas contém regiões cruciais que são essenciais para suas funções e cuja sequência é, portanto, conservada. A fração da sequência geral que é crítica varia de proteína para proteína, complicando a tarefa de relacionar a sequência à estrutura tridimensional, e a estrutura à função. Antes de se considerar esse problema com mais detalhe, é preciso, entretanto, examinar como a informação da sequência é obtida.

### As sequências de aminoácidos de milhões de proteínas foram determinadas

Duas grandes descobertas de 1953 tiveram fundamental importância na história da bioquímica. Nesse ano, James D. Watson e Francis Crick deduziram a estrutura em dupla-hélice do DNA e propuseram uma base estrutural para sua replicação precisa (Capítulo 8). Sua proposta iluminou a realidade molecular por trás da ideia de um gene. No mesmo ano, Frederick Sanger descobriu a sequência de resíduos de aminoácidos nas cadeias polipeptídicas do hormônio insulina (**Figura 3-24**), surpreendendo muitos pesquisadores que pensavam há muito tempo que a determinação da sequência de aminoácidos de um polipeptídeo seria uma tarefa irremediavelmente difícil. Rapidamente se tornou evidente que a sequência de nucleotídeos no DNA e a sequência de aminoácidos em proteínas estavam de algum modo relacionadas. Quase uma década após estas descobertas, o código genético foi elucidado, relacionando a sequência de nucleotídeos do DNA à sequência de aminoácidos em moléculas de proteínas (Capítulo 27). As sequências de aminoácidos de proteínas são agora mais frequentemente derivadas indiretamente a partir das sequências de DNA em bancos de dados genômicos. Entretanto, uma série de técnicas derivadas de métodos tradicionais de sequenciamento de polipeptídeos ainda ocupa um lugar importante na química de proteínas. Na sequência, é resumido o método tradicional e mencionadas algumas das técnicas derivadas dele.

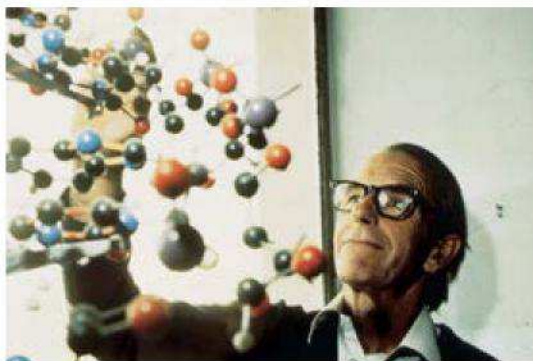


**FIGURA 3-24** Sequência de aminoácidos da insulina bovina. As duas cadeias de polipeptídeos estão unidas por ligações cruzadas de dissulfeto (amarelo). A cadeia A da insulina é idêntica em humanos, porcos, cães, coelhos e cachalotes. As cadeias B de vacas, porcos, cães, bodes e cavalos são idênticas.



## A química de proteínas é enriquecida por métodos derivados do clássico sequenciamento de polipeptídeos

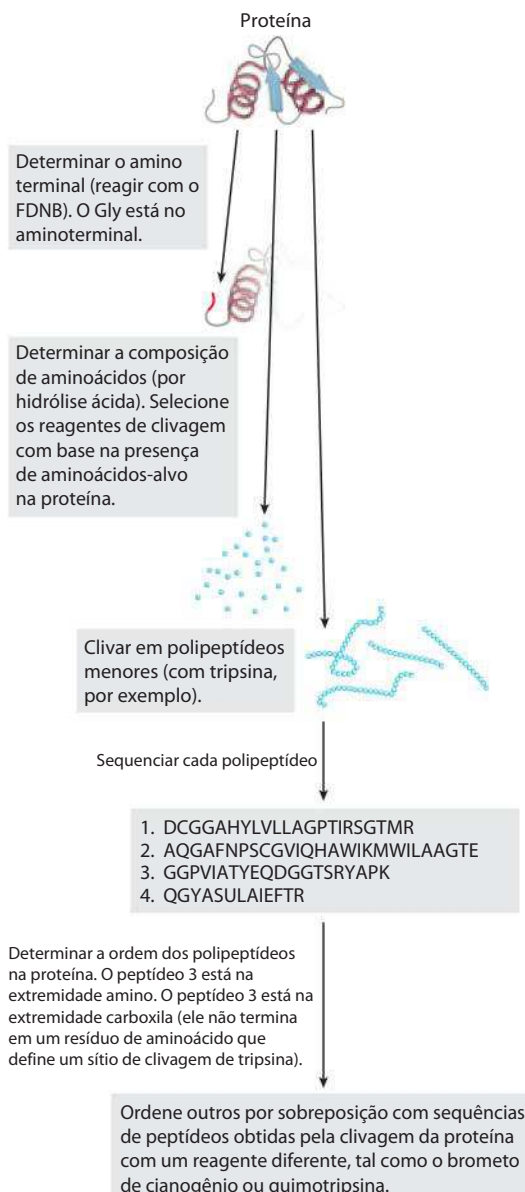
Os métodos utilizados na década de 1950 por Fred Sanger para determinar a sequência da proteína insulina são resumidos, em sua forma moderna, na **Figura 3-25**. Poucas proteínas são sequenciadas desse modo, atualmente, pelo menos em sua totalidade. Entretanto, esses protocolos de sequenciamento tradicionais têm proporcionado uma rica variedade de ferramentas para os bioquímicos, e quase todas as etapas na Figura 3-25 fazem uso de métodos que são amplamente utilizados, algumas vezes em contextos bastante diferentes.



Frederick Sanger, 1918-2013

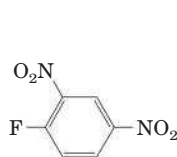
No esquema tradicional para sequenciamento de proteínas grandes, o resíduo aminoterminal do aminoácido foi inicialmente marcado e sua identidade determinada. O grupo  $\alpha$ -aminoterminal pode ser marcado com 1-fluoro-2,4-dinitrobenzeno (FDNB), cloreto de dansila ou cloreto de dabsilo (**Figura 3-26**).

O processo de sequenciamento químico em si é baseado em um processo de duas etapas desenvolvido por Pehr Edman (**Figura 3-27**). O procedimento de **degradação de Edman** marca e remove apenas o resíduo aminoterminal de um peptídeo, deixando todas as outras ligações peptídicas intactas. O peptídeo reage com o fenilisotiocianato em condições levemente alcalinas, o que converte o aminoácido aminoterminal em um aduto de feniltiocarbamoil (PTC). A ligação peptídica próxima ao aduto de PTC é, então, clivada em uma etapa efetuada em ácido trifluoracético anídrico, com remoção do aminoácido aminoterminal como um derivado anilintiazolinona. O aminoácido aminoterminal derivado é extraído com solventes orgânicos, convertido em um derivado da feniltioindantoina mais estável por tratamento com ácido aquoso e, em seguida, identificado. A utilização de reações sequenciais levadas a cabo em condições primeiro básicas e depois em condições ácidas fornece um meio de controlar todo o processo. Cada reação com o amino-

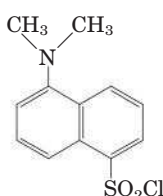


**FIGURA 3-25 Sequenciamento direto de proteínas.** Os procedimentos aqui apresentados foram aqueles desenvolvidos por Fred Sanger para sequenciar a insulina e têm sido usados posteriormente para várias proteínas adicionais. FDNB é 1-fluoro-2,4-dinitrobenzeno (ver texto e Figura 3-26).

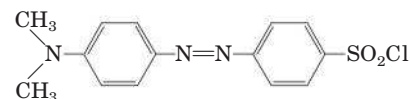
ácido aminoterminal pode chegar, essencialmente ao fim, sem afetar qualquer outra ligação peptídica no peptídeo. O processo se repete até que, tipicamente, 40 resíduos de aminoácidos sequenciais sejam identificados. As reações da degradação de Edman foram automatizadas.



FDNB



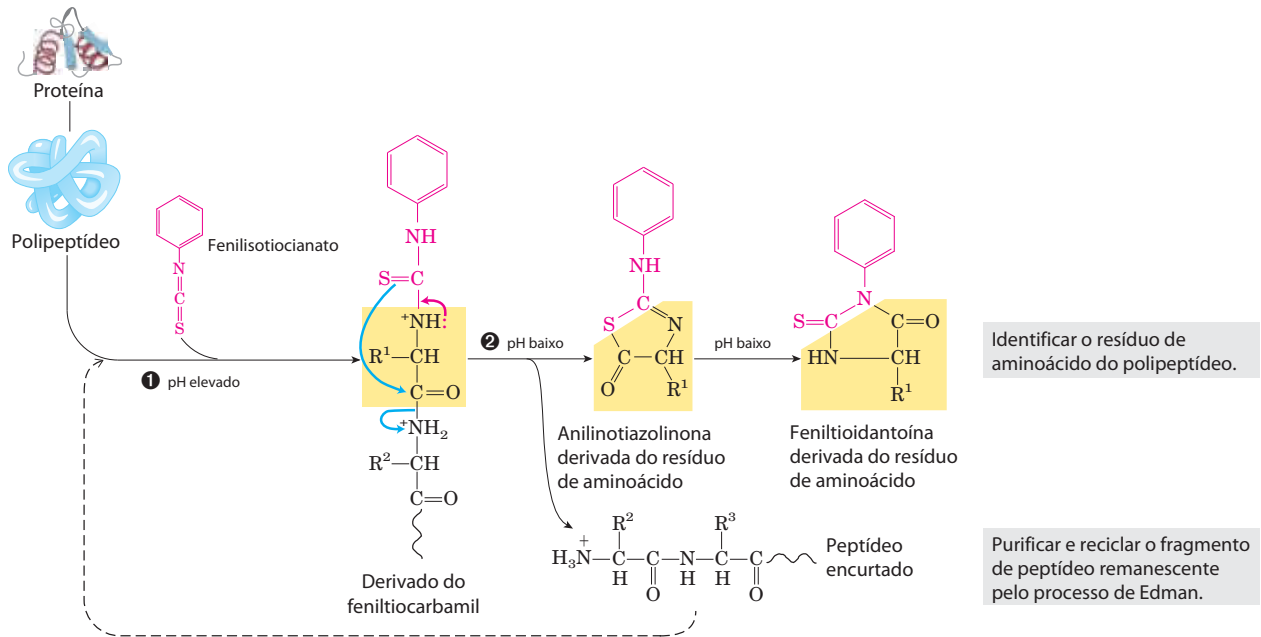
Cloreto de dansila



Cloreto de dabsilo

**FIGURA 3-26 Reagentes usados para modificar o grupo  $\alpha$ -amino do aminoácido terminal.**



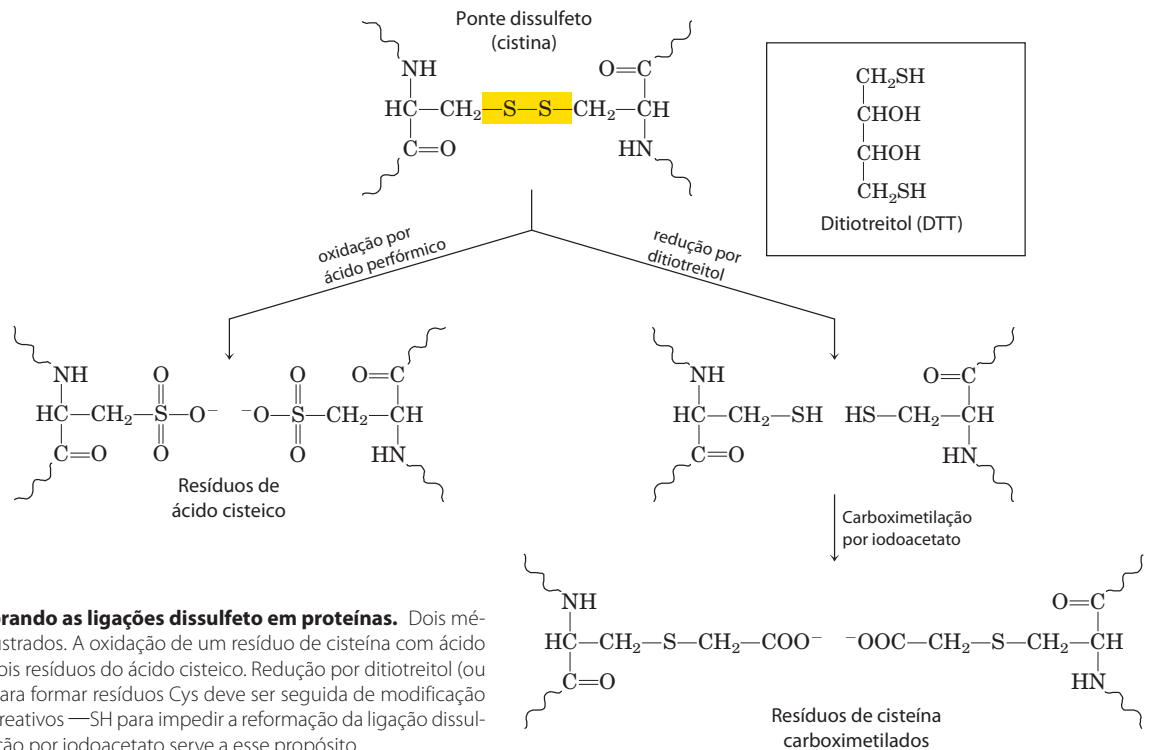


**FIGURA 3-27** A química do sequenciamento de proteínas desenvolvida por Pehr Edman. A ligação peptídica mais próxima do aminoterminal da proteína ou polipeptídeo é clivada em duas etapas. As duas etapas são

levadas a cabo sob condições de reação muito diferentes (condições básicas na etapa 1 e ácidas na etapa 2), permitindo que uma etapa prossiga até sua conclusão antes que a segunda se inicie.

Para determinar a sequência de proteínas grandes, os primeiros elaboradores de protocolos de sequenciamento tiveram que desenvolver métodos para eliminar as ligações dissulfeto e para clivar as proteínas com precisão em polipeptídeos menores. Duas abordagens para a degradação irreversível das ligações dissulfeto são destacadas na **Figura 3-28**. As enzimas chamadas **proteases** catalisam a cliva-

gem hidrolítica das ligações peptídicas. Algumas proteases clivam apenas a ligação peptídica adjacente a determinados resíduos de aminoácidos (Tabela 3-6) e, portanto, fragmentam uma cadeia polipeptídica de uma maneira previsível e reproduzível. Poucos reagentes químicos também clivam a ligação peptídica adjacente em resíduos específicos. Entre as proteases, a enzima digestiva tripsina catalisa a hidró-



**FIGURA 3-28** Quebrando as ligações dissulfeto em proteínas. Dois métodos comuns são ilustrados. A oxidação de um resíduo de cisteína com ácido perfórmico produz dois resíduos do ácido cisteico. Redução por ditioneitol (ou  $\beta$ -mercaptoetanol) para formar resíduos Cys deve ser seguida de modificação adicional dos grupos reativos —SH para impedir a reformação da ligação dissulfeto. A carboximetilação por iodoacetato serve a esse propósito.

**TABELA 3-6** Especificidade de alguns métodos comuns para fragmentação de cadeias polipeptídicas

Reagente (fonte biológica)*	Pontos de quebra†
Tripsina (pâncreas bovino)	Lys, Arg (C)
Protease submaxilar (glândula submaxilar de camundongos)	Arg (C)
Quimotripsina (pâncreas bovino)	Phe, Trp, Tyr (C)
Protease V8 de <i>Staphylococcus aureus</i> (bactéria <i>S. aureus</i> )	Asp, Glu (C)
Asp-N-protease (bactéria <i>Pseudomonas fragi</i> )	Asp, Glu (N)
Pepsina (estômago porcino)	Leu, Phe, Trp, Tyr (N)
Endopeptidase Lys C (bactéria <i>Lysobacter enzymogenes</i> )	Lys (C)
Brometo de cianogênio	Met (C)

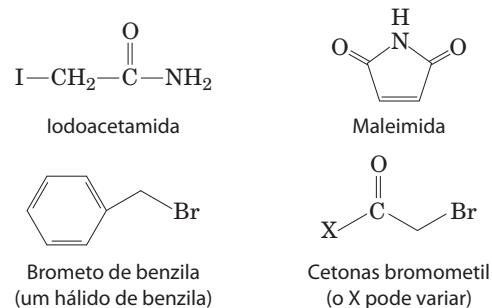
\*Todos os reagentes, com exceção do brometo de cianogênio, são proteases. Todos estão disponíveis a partir de fontes comerciais.

†Resíduos fornecendo o ponto principal de reconhecimento para a protease ou reagentes; a quebra da ligação peptídica ocorre ou no lado carbonílico (C) ou no lado amino (N) dos resíduos de aminoácidos indicados.

lise apenas daquelas ligações peptídicas em que o grupo carbonila é fornecido tanto por um resíduo de Lys quanto um de Arg, independente do comprimento ou da sequência de aminoácidos da cadeia. Um polipeptídeo com três resíduos Lys e/ou Arg irá normalmente gerar quatro peptídeos menores na clivagem com a tripsina. Além disso, todos, exceto um deles, terão um terminal carboxila Lys ou Arg. A escolha de um reagente para clivar a proteína em peptídeos menores pode ser auxiliada, primeiro, determinando-se o conteúdo de aminoácidos de toda a proteína, empregando ácido para reduzir a proteína a seus aminoácidos constituintes. A tripsina seria utilizada apenas em proteínas com um número adequado de resíduos de Lys ou Arg.

No sequenciamento clássico, uma proteína grande seria clivada em fragmentos duas vezes, utilizando uma protease ou um reagente de clivagem diferente a cada vez, de modo que as extremidades finais dos fragmentos fossem diferentes. Ambos os conjuntos de fragmentos seriam purificados e sequenciados. A ordem em que os fragmentos apareceram na proteína original poderia então ser determinada pela análise das sobreposições na sequência entre os dois conjuntos de fragmentos.

Mesmo não sendo mais utilizados para sequenciar proteínas inteiras, os métodos de sequenciamento tradicionais ainda são valiosos no laboratório. O sequenciamento de alguns aminoácidos a partir da terminação amino utilizando a química de Edman é frequentemente suficiente para confirmar a identidade de uma proteína conhecida que acabou de ser purificada ou para identificar uma proteína desconhecida purificada com base em sua atividade incomum. As técnicas empregadas nas etapas individuais do método de sequenciamento tradicional também são úteis para outros propósitos. Por exemplo, os métodos utilizados para quebrar as ligações dissulfeto podem também ser utilizados para desnaturar proteínas quando isto é necessário. Além disso, os esforços para marcar o resíduo de aminoácido aminoterminal levaram eventualmente ao desenvolvimento de uma série de reagentes que poderiam reagir com grupos específicos em uma proteína. Os mesmos reagentes utilizados para marcar o grupo  $\alpha$ -amino aminoterminal podem ser empregados para marcar as aminas primárias dos resíduos Lys (Figura 3-26). O grupo sulfidríla nos resíduos Cys pode ser



**FIGURA 3-29** Reagentes usados para modificar os grupos sulfidríla dos resíduos Cys. (Ver também Figura 3-28.)

modificado com iodoacetamidas, maleimididas, benzil halidas e bromometil cetonas (Figura 3-29). Outros resíduos de aminoácidos podem ser modificados por reagentes ligados a um corante ou outra molécula para auxiliar na detecção da proteína ou em estudos funcionais.

### A espectrometria de massa oferece um método alternativo para determinar sequências de aminoácidos

Adaptações modernas da **espectrometria de massa** fornecem uma importante alternativa aos métodos de sequenciamento descritos anteriormente. A espectrometria de massa pode fornecer uma medida altamente precisa da massa molecular de uma proteína, mas também pode fazer muito mais. Em especial, algumas variantes da espectrometria de massa podem fornecer muito rapidamente sequências de múltiplos pequenos segmentos de polipeptídeos (20 a 30 resíduos de aminoácidos) em uma amostra de proteína.

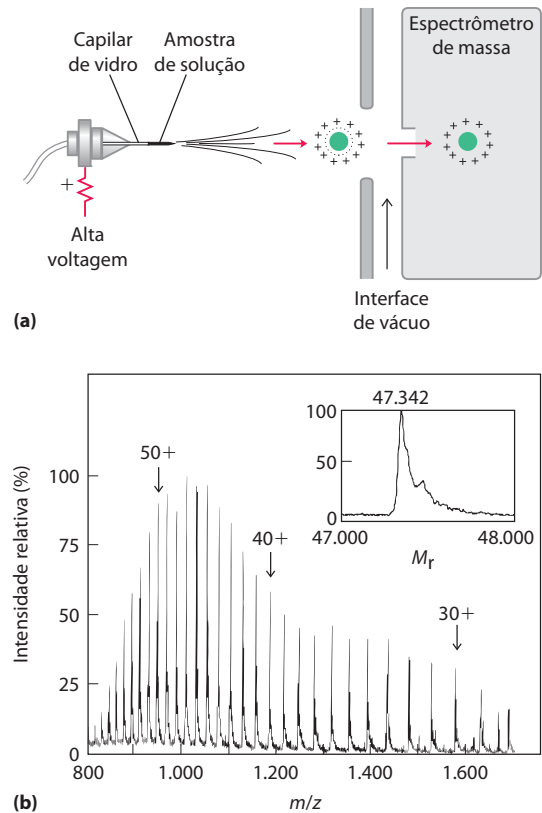
O espectrômetro de massa tem sido há muito tempo uma ferramenta indispensável na química. As moléculas a serem analisadas, chamadas de **analitos**, são inicialmente ionizadas no vácuo. Quando as moléculas recém-carregadas são introduzidas em um campo elétrico e/ou magnético, seus caminhos através do campo são uma função da razão de suas massas em relação às suas cargas,  $m/z$ . Essa propriedade medida de amostras ionizadas pode ser utilizada para deduzir a massa ( $m$ ) do analito com muita precisão.

Embora a espectrometria de massa esteja sendo utilizada há muitos anos, ela não poderia ser aplicada a macromoléculas, tais como proteínas e ácidos nucleicos. As medições de  $m/z$  são realizadas em moléculas na fase gasosa, e o aquecimento e outros tratamentos necessários para transferir uma macromolécula para a fase gasosa geralmente provocam sua rápida decomposição. Em 1988, duas técnicas diferentes foram desenvolvidas para superar esse problema. Em uma, as proteínas são colocadas em uma matriz absorvedora de luz. Com um pulso curto de luz a laser, as proteínas são ionizadas e, em seguida, desorvidas da matriz no sistema de vácuo. Esse processo, conhecido como **espectrometria de massas de desorção/ionização a laser assistida por matriz** ou **MALDI MS**, tem sido utilizada com sucesso para medir a massa de uma ampla variedade de macromoléculas. Em um segundo e bem-sucedido método, as macromoléculas em solução são forçadas diretamente da fase líquida para a gasosa. Uma solução de analitos é passada através de uma agulha carregada que é mantida em um alto potencial elétrico, dispersando a solução em uma névoa fina de microgotas carregadas. Os solventes que circundam as macromoléculas evaporam rapidamente, deixando íons de macromoléculas carregadas multiplamente na fase gasosa. Essa técnica é chamada de **espectrometria de massa com ionização por eletroaspersão**, ou **ESI MS**. Os prótons adicionados durante a passagem através da agulha fornecem carga adicional à macromolécula. O  $m/z$  da molécula pode ser analisado na câmara de vácuo.

A espectrometria de massa fornece uma riqueza de informações para a pesquisa proteômica, a enzimologia e a química de proteínas em geral. As técnicas exigem apenas quantidades minúsculas de amostra, de modo que podem ser rapidamente aplicadas a pequenas quantidades de proteína que podem ser extraídas de uma eletroforese em gel bidimensional. A medição precisa da massa molecular de uma proteína é crítica para sua identificação. Uma vez que a massa de uma proteína seja conhecida precisamente, a espectrometria de massa é um método conveniente e preciso para detectar alterações na massa devido à presença de cofatores ligados, íons metálicos ligados, modificações covalentes e assim por diante.

O processo para a determinação da massa molecular de uma proteína com ESI MS é ilustrado na **Figura 3-30**. À medida que é injetada na fase gasosa, uma proteína adquire um número variável de prótons e, portanto, cargas positivas, a partir do solvente. A adição variável dessas cargas cria um espectro de espécies com diferentes razões massa/carga. Cada pico sucessivo corresponde a uma espécie que difere do seu pico vizinho por uma diferença de carga de 1 e uma diferença de massa de 1 (1 próton). A massa da proteína pode ser determinada a partir de dois picos consecutivos.

A espectrometria de massa também pode ser utilizada para sequenciar trechos curtos de polipeptídeos, uma aplicação que surgiu como uma ferramenta inestimável para identificação rápida de proteínas desconhecidas. A informação da sequência é extraída utilizando-se uma técnica chamada de **tandem MS** ou **MS/MS**. Uma solução contendo a proteína investigada é inicialmente tratada com uma protease ou reagente químico para hidrolisá-la a uma mistura de peptídeos menores. A mistura, em seguida, é injetada em um equipamento que é essencialmente formado por

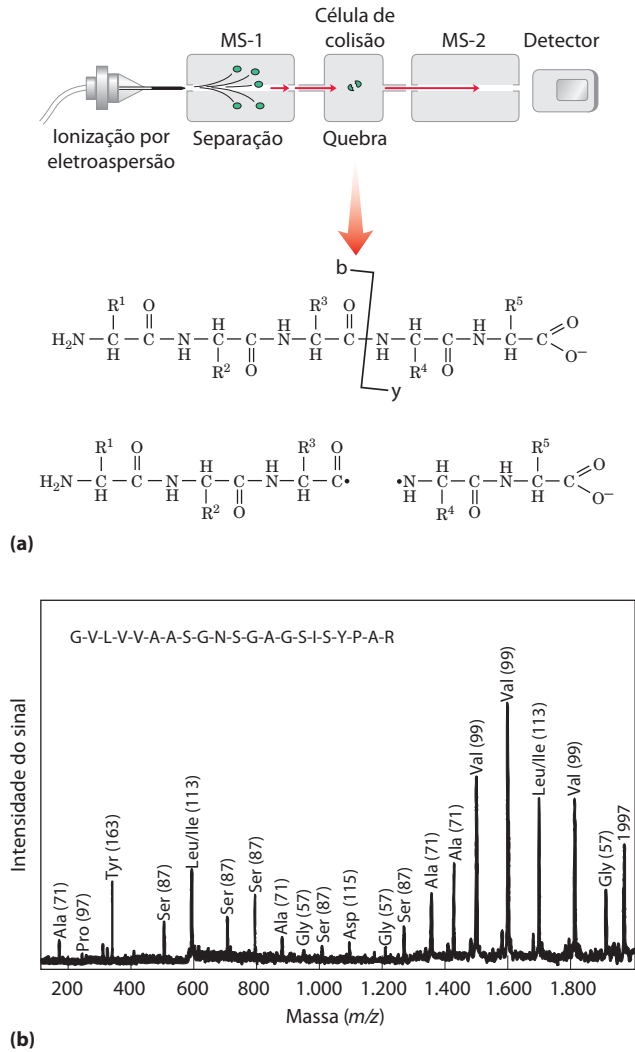


**FIGURA 3-30** Espectrometria de massa com ionização por eletroaspersão de uma proteína. (a) Uma solução de proteína é dispersa em gotículas altamente carregadas pela passagem através de uma agulha sob a influência de um campo elétrico de alta voltagem. As gotículas evaporam e os íons (com prótons adicionados, nesse caso) entram no espectrômetro de massa para medição de  $m/z$ . O espectro gerado (b) é uma família de picos, com cada pico sucessivo (da direita para a esquerda) correspondendo a uma espécie carregada com massa e carga aumentados em 1. A inserção mostra uma transformação desse espectro gerada por computador.

dois espectrômetros de massa em tandem (**Figura 3-31a**, em cima). No primeiro, a mistura de peptídeos é disposta de modo que apenas um dos vários tipos de peptídeos produzidos pela clivagem surge na outra extremidade.

A amostra do peptídeo selecionado, cada molécula do qual possui uma carga em algum ponto ao longo de seu comprimento, se desloca então através de uma câmara de vácuo entre os dois espectrômetros de massa. Nesse compartimento de colisão, o peptídeo é fragmentado adicionalmente por impacto de alta energia com um “gás de colisão”, tal como o hélio ou o argônio, que é colocado na câmara de vácuo. Cada peptídeo individual é quebrado em apenas um local, em média. Embora as quebras não sejam hidrolíticas, a maior parte ocorre nas ligações peptídicas.

O segundo espectrômetro de massa mede, em seguida, as razões  $m/z$  de todos os fragmentos carregados. Esse processo gera um ou mais conjuntos de picos. Um determinado conjunto de picos (**Figura 3-31b**) consiste em todos os fragmentos carregados que foram gerados pela quebra do mesmo tipo de ligação (mas em diferentes pontos no peptídeo). Um conjunto de picos inclui apenas os fragmentos nos quais a carga foi retida no lado aminoterminal das ligações quebradas; outro inclui apenas os fragmentos nos



**FIGURA 3-31 Obtendo informação da sequência proteica com MS em tandem.** (a) Após a hidrólise proteolítica, uma solução proteica é injetada em um espectrômetro de massa (MS-1). Os diferentes peptídeos são dispostos de modo que apenas um tipo é selecionado para análise adicional. O peptídeo selecionado é fragmentado em uma câmara entre dois espectrômetros de massa, e a  $m/z$  para cada fragmento é medida no segundo espectrômetro de massa (MS-2). Muitos dos íons gerados nessa segunda fragmentação resultam da quebra da ligação peptídica, como mostrado. Eles são chamados de íons tipo b ou íons tipo y, dependendo se a carga é retida no lado aminoterminal ou carboxiterminal, respectivamente. (b) Espectro típico com picos representando os fragmentos de peptídeos gerados a partir de uma amostra de um peptídeo pequeno (21 resíduos). Os picos marcados são íons tipo y e derivados de resíduos de aminoácidos. O número entre parênteses acima de cada pico é a massa molecular do íon do aminoácido. Os picos sucessivos diferem pela massa de um aminoácido particular no peptídeo original. A sequência deduzida é mostrada no topo.

quais a carga foi retida no lado carboxiterminal das ligações quebradas. Cada pico sucessivo em um determinado conjunto tem um aminoácido a menos que o pico anterior. A diferença na massa de pico para pico identifica o aminoácido que foi perdido em cada caso, revelando, portanto, a sequência do peptídeo. As únicas ambiguidades envolvem a leucina e a isoleucina, que têm a mesma massa. Embora múltiplos conjuntos de picos sejam normalmente gerados, os dois conjuntos mais proeminentes geralmente consistem

em fragmentos carregados derivados da quebra das ligações peptídicas. A sequência de aminoácidos derivada de um conjunto pode ser confirmada pela outra, melhorando a confiança na informação da sequência obtida.

Os vários métodos para obtenção da informação de sequências proteicas se complementam. O procedimento da degradação de Edman é algumas vezes conveniente para obter a informação da sequência unicamente a partir do terminal amino de uma proteína ou peptídeo. Entretanto, ele é relativamente lento e requer uma amostra maior do que a espectrometria de massa. A espectrometria de massa pode ser utilizada para pequenas quantidades de amostras e para amostras misturadas. Ela fornece a informação da sequência, mas os processos de fragmentação podem deixar lacunas imprevisíveis na sequência. Embora a maior parte das sequências de proteínas seja atualmente extraída de sequências do DNA genômico (Capítulo 9), empregando-se a compreensão do código genético (Capítulo 27), o sequenciamento direto de proteínas é, com frequência, necessário para identificar amostras de proteínas desconhecidas. Ambos os métodos de sequenciamento de proteínas permitem a identificação não ambígua de proteínas recém-purificadas. A espectrometria de massa é o método de escolha para identificar proteínas que estão presentes em pequenas quantidades. Por exemplo, a técnica é sensível o suficiente para analisar algumas centenas de nanogramas de proteínas que podem ser extraídos de uma única banda de proteína em um gel de poliácridamida. O sequenciamento direto por espectrometria de massa também pode revelar a adição de grupos fosforil ou outras modificações (Capítulo 6). O sequenciamento por qualquer um dos métodos pode revelar mudanças na sequência de proteínas que resultam da edição do RNA mensageiro em eucariontes (Capítulo 26). Portanto, todos esses métodos são parte de uma caixa de ferramentas robusta utilizada para investigar as proteínas e suas funções.

### Pequenos peptídeos e proteínas podem ser sintetizados quimicamente

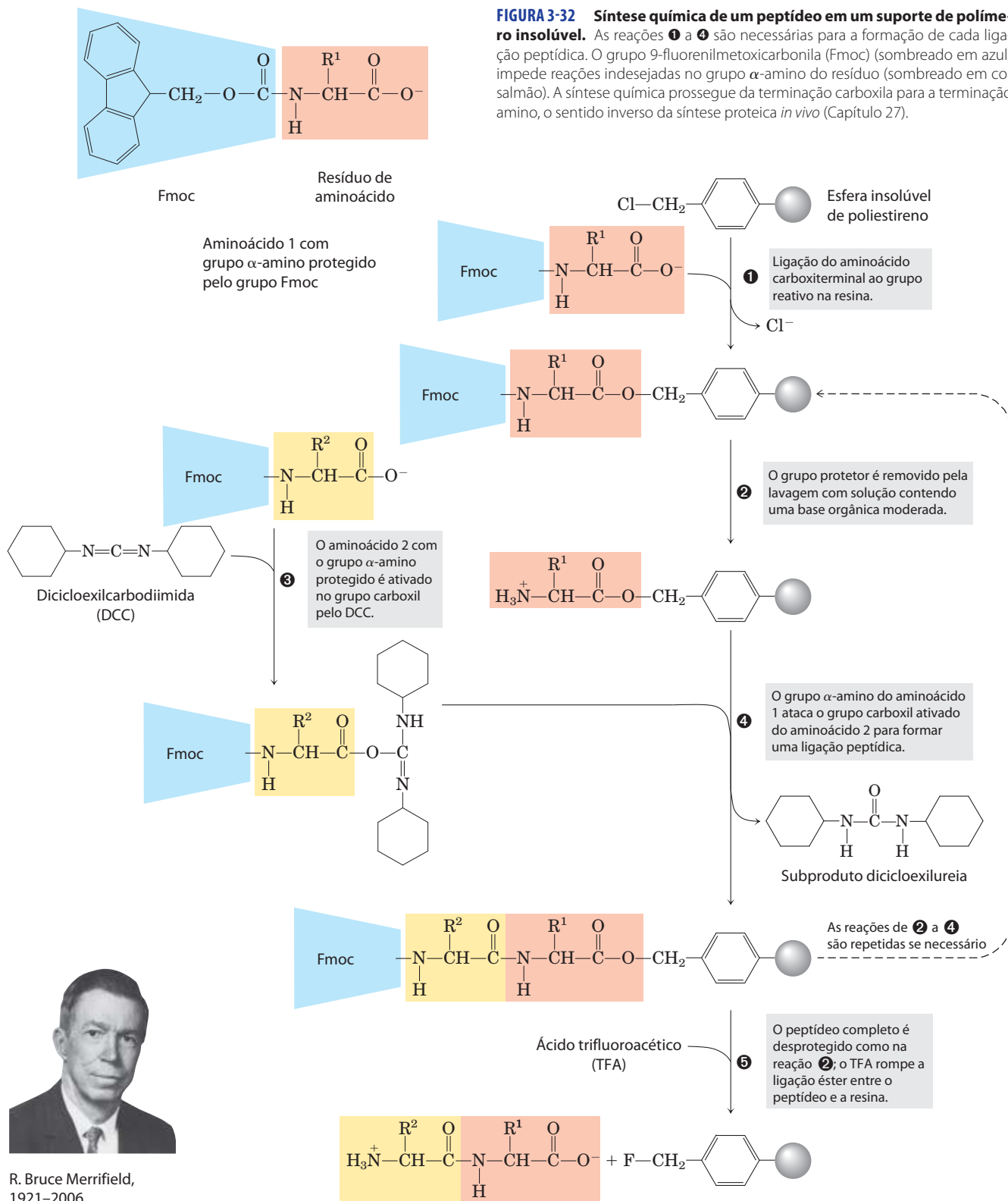
Muitos peptídeos são potencialmente úteis como agentes farmacológicos e sua produção é de considerável importância comercial. Há três modos de se obter um peptídeo: (1) purificação a partir de tecidos, tarefa frequentemente de difícil realização em consequência das concentrações infinitamente baixas de alguns peptídeos; (2) engenharia genética (Capítulo 9); ou (3) síntese química direta. Atualmente, técnicas poderosas tornam a síntese química direta uma opção atrativa em muitos casos. Além das aplicações comerciais, a síntese de porções peptídicas específicas de proteínas maiores é uma ferramenta cada vez mais importante no estudo da estrutura e função das proteínas.

A complexidade das proteínas torna as abordagens sintéticas tradicionais da química orgânica impraticáveis para peptídeos com mais de quatro ou cinco resíduos de aminoácidos. Um problema é a dificuldade de purificação do produto após cada etapa.

O principal avanço nessa tecnologia foi fornecido por R. Bruce Merrifield em 1962. Sua inovação foi sintetizar um peptídeo enquanto o mantinha ligado a uma extremidade de um suporte sólido. O suporte é um polímero insolúvel (resina) contido no interior de uma coluna, semelhante ao utilizado

em procedimentos cromatográficos. O peptídeo é construído sobre esse suporte, um aminoácido de cada vez, por meio de um conjunto padrão de reações em um ciclo repetitivo (**Figura 3-32**). Em cada etapa sucessiva no ciclo, grupos químicos de proteção bloqueiam as reações indesejadas.

A tecnologia para a síntese de peptídeos químicos é atualmente automatizada. Uma limitação importante do processo (limitação compartilhada com o processo de sequenciamento da degradação de Edman) é a eficiência de cada ciclo químico, como pode ser observado calculando a





**TABELA 3-7** Efeito do rendimento de cada etapa no rendimento global da síntese de peptídeos

Número de resíduos no polipeptídeo final	Rendimento geral do peptídeo final (%) quando o rendimento de cada etapa é:	
	96,0%	99,8%
11	66	98
21	44	96
31	29	94
51	13	90
100	1,8	82

produção total de peptídeos de vários comprimentos, quando o rendimento por adição de cada novo aminoácido é de 96,0% *versus* 99,8% (Tabela 3-7).

A reação incompleta em uma fase pode levar à formação de uma impureza (na forma de um peptídeo mais curto) na próxima. A química foi otimizada para permitir a síntese de proteínas de 100 resíduos de aminoácidos em poucos dias com rendimento razoável. Uma abordagem muito semelhante é utilizada para a síntese de ácidos nucleicos (ver Figura 8-35). É interessante notar que essa tecnologia, impressionante como é, ainda é pequena quando comparada aos processos biológicos. A mesma proteína de 100 resíduos poderia ser sintetizada com extraordinária fidelidade em cerca de 5 segundos em uma célula bacteriana.

Vários métodos novos para a ligação (união) eficiente de peptídeos tornaram possível a reunião de peptídeos sintéticos em polipeptídeos maiores e proteínas. Com esses métodos, novas formas de proteínas podem ser criadas com grupos químicos posicionados precisamente, incluindo aquelas que normalmente não podem ser encontradas em uma proteína celular. Essas novas formas fornecem novos caminhos para testar teorias de catálise enzimática, para criar proteínas com novas propriedades químicas e para desenhar sequências de proteínas que irão se dobrar em estruturas particulares. Esta última aplicação fornece o último teste de nossa capacidade crescente de relacionar a estrutura primária de um peptídeo com a estrutura tridimensional que ele assume na solução.

## As sequências de aminoácidos fornecem importantes informações bioquímicas

O conhecimento da sequência de aminoácidos em uma proteína pode oferecer ideias sobre sua estrutura tridimensional e função, localização celular e evolução. A maior parte destas ideias é derivada da procura de semelhanças entre uma proteína de interesse e as proteínas previamente estudadas. Milhares de sequências são conhecidas e estão disponíveis em bancos de dados acessíveis pela internet. Uma comparação de uma sequência recentemente obtida com este grande banco de sequências armazenadas frequentemente revela relações tanto surpreendentes quanto esclarecedoras.

Não é compreendido em detalhes como uma sequência de aminoácidos determina uma estrutura tridimensional e

tampouco é possível sempre prever a função a partir da sequência. Entretanto, famílias de proteínas com algumas características estruturais ou funcionais compartilhadas podem ser prontamente identificadas com base nas semelhanças nas suas sequências de aminoácidos. Proteínas individuais são associadas a famílias com base no grau de semelhança nas sequências de aminoácidos. Membros de uma família são geralmente idênticos em 25% ou mais de suas sequências e as proteínas nessas famílias geralmente compartilham pelo menos algumas características estruturais e funcionais. Algumas famílias são definidas, no entanto, pelas identidades envolvendo somente alguns poucos resíduos de aminoácidos que são críticos para uma determinada função. Diversas subestruturas semelhantes, ou “domínios” (a serem definidos com mais detalhes no Capítulo 4), ocorrem em muitas proteínas funcionalmente independentes. Estes domínios com frequência se dobram em configurações estruturais que possuem um grau incomum de estabilidade ou que são especializadas para um ambiente específico. Relações evolutivas também podem ser inferidas a partir das semelhanças estruturais e funcionais entre famílias de proteínas.

Certas sequências de aminoácidos funcionam como sinais que determinam a localização celular, a modificação química e a meia-vida de uma proteína. Sequências de sinalização específicas, frequentemente na porção aminoterminal, são utilizadas para direcionar certas proteínas para a exportação a partir da célula; outras proteínas são direcionadas para a distribuição para o núcleo, para a superfície celular, o citosol ou outras localizações celulares. Outras sequências atuam como sítios de ligação para grupos prostéticos, tais como os grupos de açúcares em glicoproteínas e lipídeos em lipoproteínas. Alguns desses sinais são bem caracterizados e são facilmente reconhecíveis na sequência de uma proteína recentemente caracterizada (Capítulo 27).

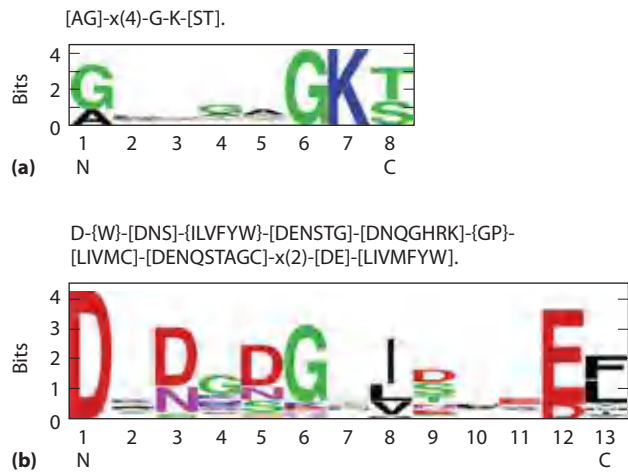
**CONVENÇÃO-CHAVE:** Muito da informação funcional encapsulada nas sequências de proteínas surge na forma de **sequências consenso**. Este termo é aplicado a sequências de DNA, RNA ou proteínas. Quando uma série de sequências de ácidos nucleicos ou de proteínas relacionadas é comparada, uma sequência consenso é aquela que reflete a base ou o aminoácido mais frequente em cada posição. Partes da sequência que apresentam concordância particularmente boa frequentemente representam domínios funcionais conservados evolutivamente. Diversas ferramentas matemáticas disponíveis na Internet podem ser utilizadas para gerar sequências consenso ou identificá-las nos bancos de dados de sequências. O Quadro 3-2 ilustra convenções comuns para a apresentação de sequências consenso. ■

## Sequências de proteínas podem elucidar a história da vida na Terra

A cadeia simples de letras que denota a sequência de aminoácidos de uma proteína tem uma riqueza surpreendente de informações. À medida que mais sequências de proteínas se tornaram disponíveis, o desenvolvimento de métodos mais poderosos para extrair informações a partir delas se tornou um importante empreendimento bioquímico. A análise das informações disponíveis nos muitos e sempre crescentes bancos de dados biológicos, incluindo as sequências

**QUADRO 3-2 Sequências consenso e logos de sequências**

Sequências consenso podem ser representadas de várias maneiras. Para ilustrar dois tipos de convenções, utilizamos dois exemplos de sequências consenso, mostrados na Figura 1: (a) estrutura de ligação ao ATP denominada alça P (ver Quadro 12-2) e (b) estrutura de ligação ao  $\text{Ca}^{2+}$  denominada mão EF (ver Figura 12-11). As regras descritas aqui são adaptadas daquelas utilizadas pela comparação de sequências do *website* PROSITE (expasy.org/prosite); elas utilizam os códigos padrão de uma letra para cada aminoácido.



**FIGURA Q-1 Representações de duas sequências de consenso.** (a) Alça P, estrutura ligadora de ATP; (b) mão EF, estrutura ligadora de  $\text{Ca}^{2+}$ .

Em um tipo de designação de sequência consenso (mostrado na parte superior de (a) e (b)), cada posição é separada de seu vizinho por um hífen. Uma posição em que qualquer aminoácido é permitido é designada x. As ambiguidades são indicadas listando os aminoácidos aceitáveis para uma determinada posição entre colchetes. Por exemplo, em (a) [AG] significa Ala ou Gly. Se todos, exceto alguns, aminoácidos são permitidos em uma posição, os aminoácidos *não* permitidos são listados entre chaves.

Por exemplo, em (b) {W} significa qualquer aminoácido, exceto Trp. A repetição de um elemento do padrão é indicada seguindo esse elemento com um número ou uma série de números entre parênteses. Em (a), por exemplo, x(4) significa x-x-x-x; x(2,4) significaria x-x ou x-x-x ou x-x-x-x. Quando um padrão é restrito ou ao grupo amino ou ao grupo carboxila terminal de uma sequência, esse padrão começa com < ou termina com >, respectivamente (não é o caso dos dois exemplos citados). Um ponto termina o padrão. Aplicando essas regras à sequência consenso em (a), tanto A como G podem ser encontrados na primeira posição. Qualquer aminoácido pode ocupar as quatro próximas posições, seguidos por um G e um K invariáveis. A última posição pode ser um S ou T.

Logos de sequência fornecem uma representação mais informativa e gráfica do alinhamento de sequência múltipla de um aminoácido (ou ácido nucleico). Cada logo consiste em uma pilha de símbolos para cada posição na sequência. A altura total da pilha (em bits) indica o grau de conservação da sequência naquela posição, enquanto a altura de cada símbolo na pilha indica a frequência relativa daquele aminoácido ou (nucleotídeo). Para sequências de aminoácidos, as cores representam as características do aminoácido: polar (G, S, T, Y, C, Q, N) verde; básico (K, R, H) azul; ácido (D, E) vermelho; e hidrofóbico (A, V, L, I, P, W, F, M) preto. Neste esquema, a classificação de aminoácidos é um pouco diferente daquela na Tabela 3-1 e na Figura 3-5. Os aminoácidos com cadeias laterais aromáticas são agrupados às classificações apolares (F, W) e polares (Y). A glicina, sempre difícil de agrupar, é colocada no grupo polar. Observe que quando múltiplos aminoácidos são aceitáveis em uma posição específica, eles raramente ocorrem com igual probabilidade. Um ou poucos em geral predominam. A representação logo torna o domínio claro e uma sequência conservada de uma proteína torna-se óbvia. Entretanto, o logo obscurece alguns resíduos de aminoácidos que podem ser permitidos em uma posição, tal como o Cys que ocorre ocasionalmente na posição 8 da mão EF em (b).

de genes e proteínas e as estruturas de macromoléculas, deram origem ao novo campo da **bioinformática**. Um dos resultados dessa disciplina é um conjunto crescente de programas de computador, muito rapidamente disponíveis na internet, que podem ser utilizados por qualquer cientista, estudante ou leigo interessado no assunto. A função de cada proteína depende de sua estrutura tridimensional, que, por sua vez, é determinada em grande parte por sua estrutura primária. Portanto, a informação transmitida por uma sequência de proteínas é limitada apenas por nossa própria compreensão dos princípios estruturais e funcionais. As ferramentas de bioinformática em constante evolução tornam possível identificar os segmentos funcionais em novas proteínas e ajudam a estabelecer tanto suas sequências quanto suas relações estruturais com proteínas já encontradas nos bancos de dados. Em um nível diferente

de investigação, as sequências de proteínas estão começando a demonstrar como as proteínas evoluíram e, em última instância, como a vida evoluiu neste planeta.

O campo da evolução molecular é frequentemente relacionado a Emile Zuckerkandl e Linus Pauling, cujos trabalhos em meados de 1960 introduziram o uso de sequências de nucleotídeos e proteínas para investigar a evolução. A premissa não é tão simples quanto aparenta. Se dois organismos são proximamente relacionados, as sequências de seus genes e proteínas devem ser semelhantes. As sequências divergem crescentemente à medida que a distância evolutiva entre dois organismos aumenta. A promessa dessa abordagem começou a ser compreendida na década de 1970, quando Carl Woese utilizou sequências de RNA ribossomal para definir as arqueias como um grupo de organismos vivos distinto de bactérias e eucariotos (ver Figura 1-4). As sequên-

cias de proteínas oferecem uma oportunidade para refinar significativamente a informação disponível. Com o advento de projetos de genoma investigando organismos de bactérias a seres humanos, o número de sequências disponíveis está crescendo a uma velocidade enorme. Essa informação pode ser utilizada para traçar a história biológica. O desafio está em aprender a ler os hieróglifos genéticos.

A evolução não tomou um caminho linear simples. As complexidades são abundantes em qualquer tentativa de extrair a informação evolutiva armazenada em sequências de proteínas. Para determinada proteína, os resíduos de aminoácidos essenciais para a atividade da proteína são conservados ao longo do tempo evolutivo. Os resíduos menos importantes para o funcionamento podem variar ao longo do tempo – isto é, um aminoácido pode ser substituído por outro – e esses resíduos variáveis podem fornecer a informação para traçar a evolução. Entretanto, as substituições de aminoácidos não são sempre aleatórias. Em algumas posições na estrutura primária, a necessidade de manter a função proteica pode significar que apenas determinadas substituições de aminoácidos podem ser toleradas. Algumas proteínas têm resíduos de aminoácidos mais variáveis que outras. Por essas e outras razões, proteínas diferentes podem evoluir em velocidades diferentes.

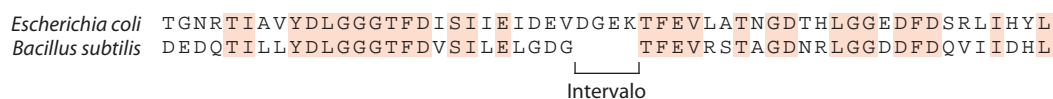
Outro fator complicador em traçar a história evolutiva é a rara transferência de um gene ou grupo de genes de um organismo para outro, um processo denominado **transferência gênica horizontal**. Os genes transferidos podem ser muito semelhantes aos genes dos quais eles foram derivados no organismo original, enquanto a maior parte dos outros genes nos mesmos dois organismos pode estar relacionada de modo muito distante. Um exemplo de transferência gênica horizontal é a recente rápida dispersão de genes de resistência a antibióticos em populações bacterianas. As proteínas derivadas desses genes transferidos não seriam bons candidatos para o estudo da evolução bacteriana, pois compartilham apenas uma história evolutiva muito limitada com seus organismos “hospedeiros”.

O estudo da evolução molecular geralmente se concentra em famílias de proteínas intimamente relacionadas. Na maior parte dos casos, as famílias escolhidas para análise têm funções essenciais no metabolismo celular que deviam ter estado presentes nas primeiras células viáveis, reduzindo, portanto, enormemente a chance de que tenham sido introduzidas há relativamente pouco tempo por transferência gênica horizontal. Por exemplo, uma proteína chamada de EF-1 $\alpha$  (fator de alongamento 1 $\alpha$ ) está envolvida na síntese de proteínas em todos os eucariontes. Uma proteína semelhante, EF-Tu, com a mesma função, é encontrada em bactérias. As semelhanças na sequência e na função indicam que a EF-1 $\alpha$  e a EF-Tu são membros de uma família de proteínas que compartilham um ancestral comum. Os mem-

bro de famílias de proteínas são denominados **proteínas homólogas** ou **homólogos**. O conceito de um homólogo pode ser mais aperfeiçoado. Se duas proteínas em uma família (isto é, dois homólogos) estão presentes nas mesmas espécies, elas são chamadas de **parálogos**. Homólogos de espécies diferentes são denominados **ortólogos**. O processo de rastrear a evolução envolve, primeiramente, a identificação de famílias adequadas de proteínas homólogas e, então, sua utilização para reconstruir as vias evolutivas.

Os homólogos são identificados pelo uso de programas de computador cada vez mais potentes que comparam diretamente duas ou mais sequências de proteínas escolhidas ou pesquisam vastos bancos de dados para descobrir os parentes evolutivos de uma sequência proteica selecionada. O processo de busca eletrônica pode ser entendido como o deslizamento de uma sequência sobre outra até que seja encontrada uma seção com boa correspondência. Nesse alinhamento de sequências, uma pontuação positiva é atribuída para cada posição onde os resíduos de aminoácidos nas duas sequências sejam idênticos – o valor da pontuação varia de um programa para o outro – para fornecer uma medida da qualidade do alinhamento. O processo tem certas complicações. Algumas vezes as proteínas comparadas apresentam correspondência, por exemplo, em dois segmentos de sequência, e esses segmentos estão conectados por sequências menos relacionadas de comprimentos diferentes. Assim, os dois segmentos correspondentes não podem ser alinhados ao mesmo tempo. Para contornar isso, o programa de computador introduz “lacunas” em uma das sequências para registrar os segmentos correspondentes (**Figura 3-33**). É claro que, se for introduzido um número suficiente de lacunas, quaisquer duas sequências poderiam ser colocadas em algum tipo de alinhamento. Para evitar alinhamentos sem informações significativas, os programas incluem penalidades para cada lacuna introduzida, reduzindo, portanto, a pontuação global do alinhamento. Com um método de tentativa e erro eletrônico, o programa seleciona o alinhamento com a pontuação ideal que maximiza os resíduos de aminoácidos idênticos enquanto minimiza a introdução de lacunas.

Com frequência, encontrar aminoácidos idênticos é inadequado para identificar proteínas relacionadas ou, principalmente, para determinar o quão proximamente relacionadas são as proteínas em uma escala de tempo evolutiva. Uma análise mais útil também leva em conta as propriedades químicas dos aminoácidos substituídos. Muitas das diferenças de aminoácidos no interior de uma família de proteínas podem ser conservativas – isto é, um resíduo de aminoácido é substituído por um resíduo com propriedades químicas semelhantes. Por exemplo, um resíduo Glu pode substituir em um membro da família o resíduo Asp encontrado em outro; ambos os aminoácidos são carregados negativamente. Tal substituição conservativa deveria logicamente receber uma



**FIGURA 3-33 Alinhando sequências de proteínas com o uso de intervalos.** Aqui é mostrada a sequência de alinhamento de uma curta seção das proteínas Hsp70 (classe muito difundida de chaperonas dobradoras de proteínas) de duas espécies de bactérias muito bem estudadas, *E. coli* e *Ba-*

*cillus subtilis*. A introdução de um intervalo na sequência de *B. subtilis* permite um melhor alinhamento dos resíduos de aminoácidos de cada lado do intervalo. Resíduos de aminoácidos idênticos estão sombreados.

pontuação maior em um alinhamento de sequências do que uma substituição não conservativa, tal como a substituição de um resíduo Asp por um resíduo hidrofóbico Phe.

Para a maioria dos esforços em encontrar homologias e explorar relações evolutivas, as sequências de proteínas (derivadas tanto diretamente do sequenciamento proteico quanto do sequenciamento do DNA que codifica as proteínas) são superiores às sequências de ácidos nucleicos não gênicas (aquelas que não codificam uma proteína ou um RNA funcional). Para um ácido nucleico, com seus quatro tipos diferentes de resíduos, o alinhamento aleatório de sequências não homólogas irá em geral produzir correspondências para, no mínimo, 25% das posições. A introdução de algumas poucas lacunas pode, com frequência, aumentar a fração de resíduos correspondentes para 40% ou mais, e a probabilidade de alinhamentos aleatórios de sequências não relacionadas torna-se bastante elevada. Os 20 resíduos de aminoácidos diferentes nas proteínas reduzem muito a probabilidade de alinhamentos aleatórios não informativos desse tipo.

Os programas utilizados para gerar um alinhamento de sequências são complementados por métodos que testam a confiabilidade dos alinhamentos. Um teste computadorizado comum consiste em embaralhar a sequência de aminoácidos de uma das proteínas que estiver sendo comparada, para produzir uma sequência aleatória e então instruir o programa a alinhar a sequência embaralhada com a outra, não embaralhada. Pontuações são designadas ao novo alinhamento, e o processo de embaralhar e alinhar é repetido muitas vezes. O alinhamento original, antes de embaralhar, deve ter uma pontuação significativamente maior do que qualquer uma daquelas pontuações geradas pelos alinhamentos aleatórios; isso aumenta a confiança de que o alinhamento de sequências identificou um par de homólogos. Observe que a ausência de pontuação de alinhamento significativo não necessariamente significa que não exista relação evolutiva entre as duas proteínas. Como será visto no Capítulo 4, as semelhanças de estruturas tridimensionais revelam algumas vezes relações evolutivas nas quais a homologia de sequências foi apagada pelo tempo.

Para utilizar uma família de proteínas para explorar a evolução, os pesquisadores identificam membros da família com funções moleculares semelhantes na faixa mais ampla possível de organismos. A informação da família pode então ser utilizada para rastrear a evolução desses organismos. Ao analisar a divergência nas sequências de famílias de proteínas selecionadas, os investigadores podem separar os organismos em classes com base em suas relações evolutivas. Esta informação deve ser conciliada com exames mais clássicos da fisiologia e da bioquímica dos organismos.

Certos segmentos de uma sequência de proteínas podem ser encontrados em organismos de um grupo taxonômico, mas não em outros grupos; esses segmentos podem ser utilizados como **sequências-assinatura** para o grupo no qual elas foram encontradas. Um exemplo de uma sequência-assinatura é a inserção de 12 aminoácidos próximos à terminação amino das proteínas EF-1 $\alpha$ /EF-Tu em todas as arqueobactérias e eucariontes, mas não em bactérias (Figura 3-34). Essa assinatura particular é um dos muitos indícios bioquímicos que podem ajudar a estabelecer o relacionamento evolutivo de eucariontes e arqueobactérias. Sequências-assinatura têm sido utilizadas para estabelecer relações evolutivas entre grupos de organismos em muitos níveis taxonômicos diferentes.

Ao considerar a sequência completa de uma proteína, os pesquisadores podem atualmente construir árvores evolutivas mais elaboradas com muitas espécies em cada grupo taxonômico. A Figura 3-35 apresenta uma dessas árvores para bactérias, com base na divergência de sequências na proteína GroEL (proteína presente em todas as bactérias que auxilia no enovelamento adequado de proteínas). A árvore pode ser aperfeiçoada utilizando as sequências de múltiplas proteínas e a complementação da informação de sequência com dados das propriedades bioquímicas e fisiológicas exclusivas de cada espécie. Há muitos métodos para gerar árvores, cada método com suas próprias vantagens e desvantagens, e diversas formas de representar as relações evolutivas resultantes. Na Figura 3-35, as extremidades livres das linhas são chamadas de “nós externos”; cada uma representa uma espécie atual, que é marcada assim. Os pontos onde duas linhas se unem, os “nós internos”, representam espécies ancestrais extintas. Na maior parte das representações (incluindo a Figura 3-35), os comprimentos das linhas que conectam os nós são proporcionais ao número de substituições de aminoácidos que separam uma espécie da outra. Ao rastrear duas espécies conservadas a um nó interno comum (representando o ancestral comum das duas espécies), o comprimento do ramo que conecta cada nó externo ao nó interno representa o número de substituições de aminoácidos que separam uma espécie atual de seu ancestral. A soma dos comprimentos de todos os segmentos de linhas que conectam uma espécie conservada a outras espécies conservadas com ancestral comum reflete o número de substituições que separam as duas espécies conservadas. Para determinar quanto tempo foi necessário para as várias espécies divergirem, a árvore precisa ser calibrada para compará-la com informações do registro fóssil e outras fontes.

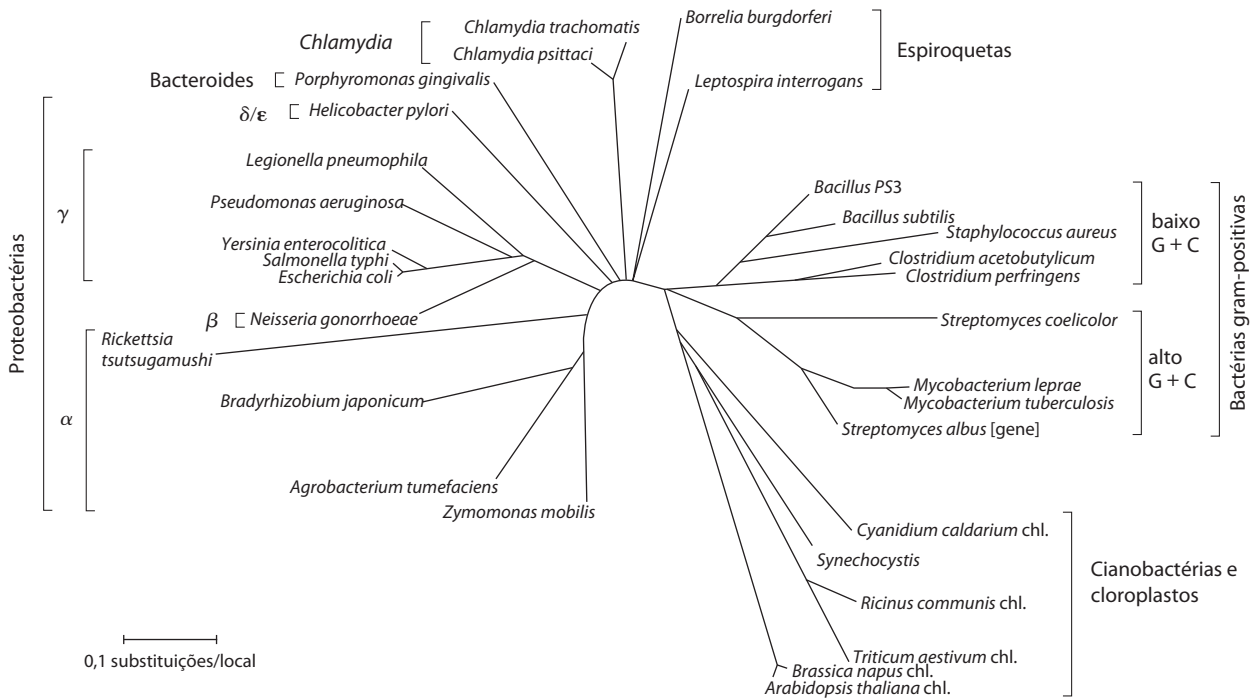
À medida que mais informação de sequência torna-se disponível nos bancos de dados, é possível gerar árvores

		Sequência-assinatura
Arqueobactérias	{ <i>Halobacterium halobium</i> <i>Sulfolobus solfataricus</i>	IGHVDHGKSTMVGR <span style="border: 1px solid black; padding: 2px;">LLYETGSVPEHV</span> IEQH
		IGHVDHGKSTLVGR <span style="border: 1px solid black; padding: 2px;">LLMDRGFIDEKT</span> VKEA
Eucariontes	{ <i>Saccharomyces cerevisiae</i> <i>Homo sapiens</i>	IGHVDSGKSTTTGHL <span style="border: 1px solid black; padding: 2px;">LIYKCGGIDKRT</span> IEKF
		IGHVDSGKSTTTGHL <span style="border: 1px solid black; padding: 2px;">LIYKCGGIDKRT</span> IEKF
Bactérias gram-positivas	<i>Bacillus subtilis</i>	IGHVDHGKSTMVGR ITTV
Bactérias gram-negativas	<i>Escherichia coli</i>	IGHVDHGKSTLTAA ITTV

**FIGURA 3-34 Uma sequência-assinatura na família de proteínas EF-1 $\alpha$ /EF-Tu.** A sequência-assinatura (no retângulo) é uma inserção de 12 resíduos próxima do terminal amino da sequência. Os resíduos que alinham em todas as espécies estão sombreados. Tanto as arqueias quanto os euca-

riontes apresentam a assinatura, embora as sequências de inserções sejam bem distintas para os dois grupos. A variação na sequência-assinatura reflete a divergência evolutiva significativa que ocorreu nesse ponto desde que ela apareceu primeiro em um ancestral comum de ambos os grupos.



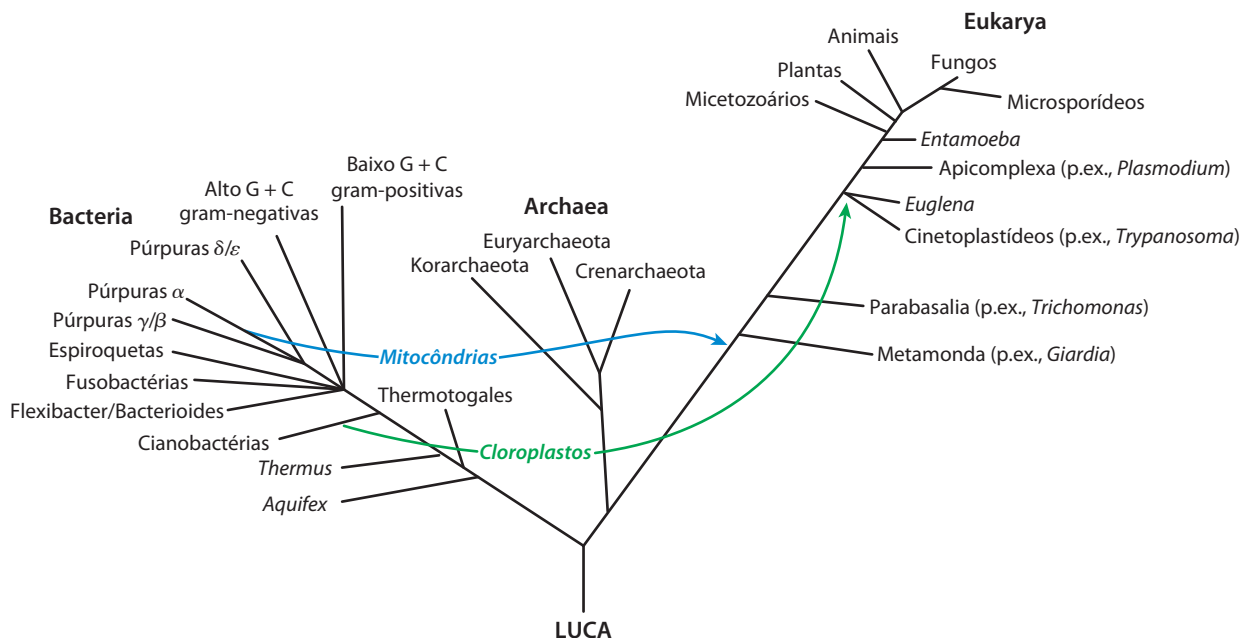


**FIGURA 3-35** **Árvore evolutiva derivada de comparações entre as sequências de aminoácidos.** Árvore evolutiva bacteriana, com base na divergência de sequências observada na família de proteínas GroEL. Tam-

bém estão incluídos nessa árvore (parte inferior direita) os cloroplastos de algumas espécies não bacterianas.

evolutivas com base em múltiplas proteínas, bem como aperfeiçoar essas árvores à medida que informação genômica adicional emerge de métodos de análise cada vez mais sofisticados. Todo esse trabalho tem o objetivo de criar uma

árvore detalhada da vida que descreva a evolução e o parentesco de cada organismo na Terra. A história é um trabalho contínuo, é claro (**Figura 3-36**). As questões levantadas e respondidas são fundamentais para definir como os huma-



**FIGURA 3-36** **Árvore de consenso da vida.** A árvore mostrada aqui baseia-se em análises de muitas sequências de proteínas e características genômicas adicionais. A árvore apresenta apenas uma fração da informação disponível, bem como apenas uma fração dos temas que ainda precisam ser solucionados. Cada grupo existente mostrado é uma história evolutiva

complexa em si mesma. LUCA é o último ancestral comum universal do qual todas as outras formas de vida evoluíram. As setas azul e verde indicam a assimilação endossimbiótica de tipos específicos de bactérias por células eucariontes para se tornarem mitocôndrias e cloroplastos, respectivamente (ver Figura 1–38).



nos veem a si mesmos e ao mundo ao seu redor. O campo da evolução molecular promete estar entre as mais vibrantes fronteiras científicas do século XXI.

**RESUMO 3.4 A estrutura de proteínas: estrutura primária**

- ▶ Diferenças na função de proteínas resultam de diferenças na composição e na sequência de aminoácidos. Algumas variações na sequência podem ocorrer em uma proteína particular, com pouco ou nenhum efeito em sua função.
- ▶ As sequências de aminoácidos são deduzidas pela fragmentação de polipeptídeos em peptídeos menores com reagentes conhecidos para clivar ligações peptídicas específicas; pela determinação das sequências de aminoácidos de cada fragmento pelo procedimento automatizado de degradação de Edman; então, pela ordenação dos fragmentos peptídicos pelo encontro de sobreposições de sequências entre os fragmentos gerados por diferentes reagentes. A sequência de uma proteína também pode ser deduzida a partir da sequência de nucleotídeos de seu gene correspondente no DNA ou por espectrometria de massa.
- ▶ Peptídeos e proteínas pequenas (até cerca de 100 resíduos) podem ser sintetizados quimicamente. O peptídeo é construído, um resíduo de aminoácido por vez, enquanto unido a um suporte sólido.

▶ Sequências proteicas são uma fonte rica de informação sobre a estrutura e a função da proteína, bem como sobre a evolução da vida na Terra. Métodos sofisticados estão sendo desenvolvidos para rastrear a evolução, analisando as lentas mudanças resultantes nas sequências de aminoácidos de proteínas homólogas.

**Término leitura complementar aula 4**

**Palavras-chave**

As palavras-chave em negrito são definidas no glossário.

<b>aminoácidos</b> 76	<b>diálise</b> 90
<b>resíduo</b> 76	cromatografia em coluna 90
<b>grupo R</b> 76	<b>cromatografia de troca iônica</b> 90
<b>centro quiral</b> 76	<b>cromatografia de exclusão por tamanho</b> 92
<b>enantiômeros</b> 76	cromatografia de afinidade 92
<b>configuração absoluta</b> 78	<b>cromatografia líquida de alto desempenho (HPLC)</b> 92
sistema D, L 78	<b>eletroforese</b> 92
<b>polaridade</b> 78	dodecil sulfato de sódio (SDS) 94
absorbância, A 80	<b>focalização isoelétrica</b> 94
<b>zwitterion</b> 81	<b>atividade específica</b> 95
<b>pH isoelétrico (ponto isoelétrico, pI)</b> 84	<b>estrutura primária</b> 97
<b>peptídeo</b> 85	<b>estrutura secundária</b> 97
<b>proteína</b> 85	<b>estrutura terciária</b> 97
<b>ligação peptídica</b> 85	<b>estrutura quaternária</b> 97
<b>oligopeptídeo</b> 86	degradação de Edman 98
<b>polipeptídeo</b> 86	
<b>proteína oligomérica</b> 88	
<b>protômero</b> 88	
<b>proteína conjugada</b> 89	
<b>grupo prostético</b> 89	
extrato bruto 89	
<b>fração</b> 89	
<b>fracionamento</b> 89	

<b>proteases</b> 99	horizontal 106
MALDI MS 101	<b>proteínas homólogas</b> 106
ESI MS 101	<b>homólogos</b> 106
<b>sequência consenso</b> 104	<b>parálogos</b> 106
<b>bioinformática</b> 104	<b>ortólogos</b> 106
transferência gênica	sequência-assinatura 107

**Leituras adicionais**

**Aminoácidos**

**Dougherty, D.A.** (2000) Unnatural amino acids as probes of protein structure and function. *Curr. Opin. Chem. Biol.* **4**, 645–652.

**Kreil, G.** (1997) D-Amino acids in animal peptides. *Annu. Rev. Biochem.* **66**, 337–345.

Detalhes da ocorrência destes estereoisômeros incomuns de aminoácidos.

**Meister, A.** (1965) *Biochemistry of the Amino Acids*, 2nd edn, Vols 1 and 2, Academic Press, Inc., New York.

Tratamento enciclopédico das propriedades, da ocorrência e do metabolismo de aminoácidos.

**Peptídeos e proteínas**

**Creighton, T.E.** (1992) *Proteins: Structures and Molecular Properties*, 2nd edn, W. H. Freeman and Company, New York.

Fonte geral muito útil.

**Trabalhando com proteínas**

**Dunn, M.J. & Corbett, J.M.** (1996) Two-dimensional polyacrylamide gel electrophoresis. *Methods Enzymol.* **271**, 177–203.

Descrição detalhada da tecnologia.

**Kornberg, A.** (1990) Why purify enzymes? *Methods Enzymol.* **182**, 1–5.

O papel crucial dos métodos bioquímicos clássicos em uma nova era.

**Scopes, R.K.** (1994) *Protein Purification: Principles and Practice*, 3rd edn, Springer-Verlag, New York.

Boa fonte para descrições mais completas sobre os princípios subjacentes à cromatografia e outros métodos.

**Estrutura primária de proteínas e evolução**

**Andersson, L., Blomberg, L., Flegel, M., Lepsa, L., Nilsson, B., & Verlander, M.** (2000) Large-scale synthesis of peptides. *Biopolymers* **55**, 227–250.

Discussão sobre as abordagens para a fabricação de peptídeos como produtos farmacêuticos.

**Dell, A. & Morris, H.R.** (2001) Glycoprotein structure determination by mass spectrometry. *Science* **291**, 2351–2356.

As glicoproteínas podem ser complexas; a espectrometria de massa é um método preferido de classificação.

**Delsuc, F., Brinkmann, H., & Philippe, H.** (2005) Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375.

**Gogarten, J.P. & Townsend, J.P.** (2005) Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3**, 679–687.

**Gygi, S.P. & Aebersold, R.** (2000) Mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.* **4**, 489–494.

Usos da espectrometria de massa para identificar e estudar as proteínas celulares.

**Koonin, E.V., Tatusov, R.L., & Galperin, M.Y.** (1998) Beyond complete genomes: from sequence to structure and function. *Curr. Opin. Struct. Biol.* **8**, 355–363.

Boa discussão sobre os possíveis usos do aumento da quantidade de informações sobre sequências de proteínas.

**Li, W.-H. & Graur, D.** (2000) *Fundamentals of Molecular Evolution*, 2nd edn, Sinauer Associates, Inc., Sunderland, MA.

Texto de fácil leitura descrevendo os métodos utilizados para a análise de seqüências de proteínas e de ácidos nucleicos. O Capítulo 5 fornece uma das melhores descrições disponíveis de como árvores evolutivas são construídas a partir de dados de seqüências.

**Mayo, K.H.** (2000) Recent advances in the design and construction of synthetic peptides: for the love of basics or just for the technology of it. *Trends Biotechnol.* **18**, 212–217.

**Miranda, L.P. & Alewood, P.F.** (2000) Challenges for protein chemical synthesis in the 21st century: bridging genomics and proteomics. *Biopolymers* **55**, 217–226.

Este é o artigo de Mayo (acima) descrevem como produzir peptídeos e uni-los em conjuntos para tratar uma ampla faixa de problemas na bioquímica de proteínas.

**Ramisetty, S.R. & Washburn, M.P.** (2011) Unraveling the dynamics of protein interactions with quantitative mass spectrometry. *Crit. Rev. Biochem. Mol. Biol.* **46**, 216–228.

**Rokas, A., Williams, B.L., King, N., & Carroll, S.B.** (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804.

Como as comparações de seqüência de múltiplas proteínas podem produzir informação evolutiva precisa.

**Sanger, F.** (1988) Sequences, sequences, sequences. *Annu. Rev. Biochem.* **57**, 1–28.

Bom relato histórico do desenvolvimento de métodos de seqüenciamento.

**Snel, B., Huynen, M.A., & Dutilh, B.E.** (2005) Genome trees and the nature of genome evolution. *Annu. Rev. Microbiol.* **59**, 191–209.

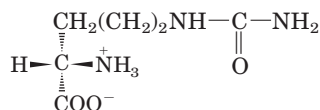
**Steen, H. & Mann, M.** (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5**, 699–711.

**Zuckerandl, E. & Pauling, L.** (1965) Molecules as documents of evolutionary history. *J. Theor. Biol.* **8**, 357–366.

Muitos consideram este artigo como o fundador no campo da evolução molecular.

## Problemas

**1. Configuração absoluta da citrulina.** A citrulina isolada de melancias tem a estrutura apresentada a seguir. Ela é um aminoácido D ou L? Explique.



**2. Relação entre a curva de titulação e as propriedades acidobásicas da glicina.** Uma solução de 100 mL de glicina a 0,1 M em pH 1,72 foi titulada com uma solução de 2 M de NaOH. O pH foi monitorado e os resultados foram plotados como mostrado no gráfico. Os pontos-chave na titulação são designados de I a V. Para cada uma das afirmações de (a) a (o), *identifique* o ponto-chave adequado na titulação e *justifique* sua escolha.

(a) A glicina está presente predominantemente como a espécie  $^+\text{H}_3\text{N}-\text{CH}_2-\text{COOH}$ .

(b) A carga final *média* da glicina é  $+\frac{1}{2}$ .

(c) Metade dos grupos amino está ionizado.

(d) O pH é igual ao  $\text{pK}_a$  do grupo carboxila.

(e) O pH é igual ao  $\text{pK}_a$  do grupo amino protonado.

(f) A glicina possui sua capacidade de tamponamento máxima.

(g) A carga final *média* da glicina é zero.

(h) O grupo carboxila foi completamente titulado (primeiro ponto de equivalência).

(i) A glicina está completamente titulada (segundo ponto de equivalência).

(j) A espécie predominante é  $^+\text{H}_3\text{N}-\text{CH}_2-\text{COO}^-$ .

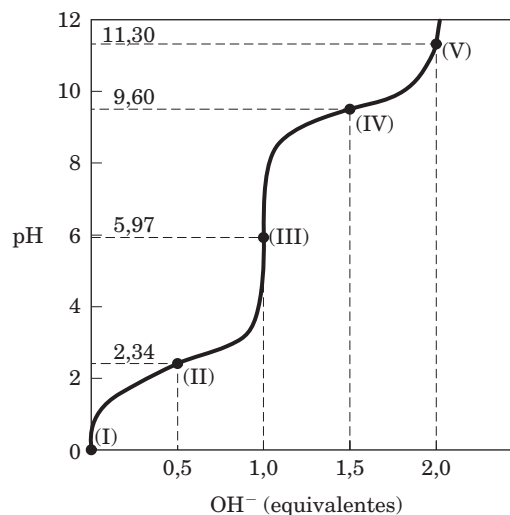
(k) A carga final *média* da glicina é  $-1$ .

(l) A glicina está presente predominantemente como uma mistura 50:50 de  $^+\text{H}_3\text{N}-\text{CH}_2-\text{COOH}$  e  $^+\text{H}_3\text{N}-\text{CH}_2-\text{COO}^-$ .

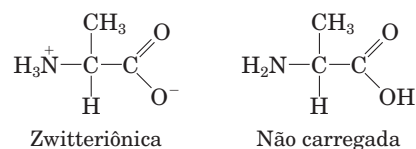
(m) Este é o ponto isoelétrico.

(n) Este é o final da titulação.

(o) Estas são as *piores* regiões de pH para poder de tamponamento.



**3. Quanta alanina está presente na forma da espécie completamente sem carga?** Em um pH igual ao ponto isoelétrico da alanina, a sua carga *final* é zero. Duas estruturas podem ser desenhadas que apresentam carga final igual a zero, mas a forma predominante de alanina em seu pI é zwitteriônica.



(a) Por que a alanina é predominantemente zwitteriônica em vez de completamente não carregada em seu pI?

(b) Que fração de alanina está na forma completamente não carregada em seu pI? Justifique suas suposições.

**4. Estado de ionização da histidina.** Cada grupo ionizável de um aminoácido pode existir em um de dois estados, carregado ou neutro. A carga elétrica no grupo funcional é determinada pela relação entre seu  $\text{pK}_a$  e o pH da solução. Essa relação é descrita pela equação de Henderson-Hasselbalch.

(a) A histidina tem três grupos funcionais ionizáveis. Escreva as equações de equilíbrio para suas três ionizações e assinale o  $\text{pK}_a$  adequado para cada ionização. Desenhe a estrutura da histidina em cada estado de ionização. Qual é a carga final na molécula de histidina em cada estado de ionização?

(b) Desenhe as estruturas do estado de ionização predominante da histidina em pH 1, 4, 8 e 12. Observe que o estado de

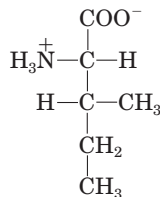
ionização pode ser aproximado tratando-se cada grupo ionizável independentemente.

(c) Qual é a carga final da histidina em pH 1, 4, 8 e 12? Para cada pH, a histidina irá migrar em direção ao ânodo (+) ou ao cátodo (-) quando colocada em um campo elétrico?

**5. Separação de aminoácidos por cromatografia de troca iônica.** Misturas de aminoácidos podem ser analisadas primeiramente pela separação da mistura em seus componentes por uma cromatografia de troca iônica. Os aminoácidos inseridos em uma resina permutadora de cátions (ver Figura 3-17a) contendo grupos sulfonados ( $-\text{SO}_3^-$ ) fluem pela resina em velocidades diferentes em consequência de dois fatores que influenciam seu movimento: (1) atração iônica entre os resíduos sulfonados na coluna e os grupos funcionais carregados positivamente nos aminoácidos e (2) interações hidrofóbicas entre as cadeias laterais de aminoácidos e o esqueleto fortemente hidrofóbico da resina de poliestireno. Para cada par de aminoácidos listados, determine qual será eluído primeiro em uma coluna permutadora de cátions por um tampão de pH 7,0.

- (a) Asp e Lys
- (b) Arg e Met
- (c) Glu e Val
- (d) Gly e Leu
- (e) Ser e Ala

**6. Nomeando os estereoisômeros de isoleucina.** A estrutura do aminoácido isoleucina é



- (a) Quantos centros quirais ela tem?
- (b) Quantos isômeros ópticos?
- (c) Desenhe fórmulas em perspectiva para todos os isômeros ópticos da isoleucina.

**7. Comparação dos valores de  $pK_a$  de alanina e polialanina.** A curva de titulação da alanina mostra a ionização de dois grupos funcionais com valores de  $pK_a$  de 2,34 e 9,69, correspondendo à ionização do grupo carboxila e dos grupos amino protonados, respectivamente. A titulação de di, tri e oligopeptídeos maiores de alanina também mostra a ionização de somente dois grupos funcionais, embora os valores experimentais de  $pK_a$  sejam diferentes. A tendência nos valores de  $pK_a$  está resumida na tabela.

Aminoácido ou peptídeo	$pK_1$	$pK_2$
Ala	2,34	9,69
Ala-Ala	3,12	8,30
Ala-Ala-Ala	3,39	8,03
Ala-(Ala) $_n$ -Ala, $n \geq 4$	3,42	7,94

- (a) Desenhe a estrutura de Ala-Ala-Ala. Identifique os grupos funcionais associados a  $pK_1$  e  $pK_2$ .
- (b) Por que o valor de  $pK_1$  *augmenta* com cada resíduo Ala adicional no oligopeptídeo?
- (c) Por que o valor de  $pK_2$  *diminui* com cada resíduo Ala adicional no oligopeptídeo?

**8. O tamanho das proteínas.** Qual é a massa molecular aproximada de uma proteína com 682 resíduos de aminoácidos em uma única cadeia polipeptídica?

**9. O número de resíduos de triptofano na albumina sérica bovina.** Uma análise quantitativa de aminoácidos revela que a albumina sérica bovina (BSA) contém 0,58% de triptofano ( $M_r$  204) por peso.

(a) Calcule a massa molecular *mínima* da BSA (i.e, presumindo-se que haja apenas um resíduo de Trp por molécula de proteína).

(b) A cromatografia de exclusão por tamanho da BSA fornece uma massa molecular estimada de BSA de 70.000. Quantos resíduos de Trp estão presentes em uma molécula de albumina sérica?

**10. Composição de subunidades de uma proteína.** Uma proteína tem uma massa molecular de 400 kDa quando medida por cromatografia de exclusão por tamanho. Quando submetida a uma eletroforese em gel na presença de dodecil sulfato de sódio (SDS), a proteína fornece três bandas com massas moleculares de 180, 160 e 60 kDa. Quando a eletroforese é realizada na presença de SDS e ditioneitol, três bandas são novamente formadas, desta vez com massas moleculares de 160, 90 e 60 kDa. Determine a composição das subunidades da proteína.

**11. Carga elétrica final de peptídeos.** Um peptídeo tem a sequência



- (a) Qual é a carga final da molécula em pH 3, 8 e 11? (Utilize os valores de  $pK_a$  para cadeias laterais e grupos amino e carboxila terminais como fornecidos na Tabela 3-1.)
- (b) Estime o pI para este peptídeo.

**12. Ponto isoelétrico da pepsina.** Pepsina é o nome dado a uma mistura de diversas enzimas digestivas secretadas (como proteínas precursoras maiores) por glândulas no estômago. Essas glândulas também secretam ácido clorídrico, que dissolve o material particulado no alimento, permitindo à pepsina clivar de modo enzimático moléculas de proteínas individuais. A mistura resultante de alimento, HCl e enzimas digestivas é conhecida como quimo e apresenta pH próximo a 1,5. Qual pI você poderia prever para as proteínas da pepsina? Que grupos funcionais devem estar presentes para conferir esse pI à pepsina? Quais aminoácidos nas proteínas iriam contribuir com tais grupos?

**13. Ponto isoelétrico de histonas.** As histonas são proteínas encontradas no núcleo de células eucarióticas, fortemente ligadas ao DNA, com muitos grupos fosfato. O pI das histonas é muito alto, cerca de 10,8. Que resíduos de aminoácidos devem estar presentes em quantidades relativamente elevadas nas histonas? De que forma esses resíduos contribuem para a forte ligação das histonas ao DNA?

**14. Solubilidade de polipeptídeos.** Um método para separar polipeptídeos faz uso de suas diferentes solubilidades. A solubilidade de polipeptídeos grandes em água depende da polaridade relativa de seus grupos R, particularmente do número de grupos ionizáveis: quanto mais grupos ionizáveis existirem, mais solúvel será o polipeptídeo. Qual, de cada par de polipeptídeos a seguir, é mais solúvel no pH indicado?

- (a) (Gly) $_{20}$  ou (Glu) $_{20}$  em pH 7,0
- (b) (Lys-Ala) $_3$  ou (Phe-Met) $_3$  em pH 7,0
- (c) (Ala-Ser-Gly) $_5$  ou (Asn-Ser-His) $_5$  em pH 6,0
- (d) (Ala-Asp-Gly) $_5$  ou (Asn-Ser-His) $_5$  em pH 3,0

**15. Purificação de uma enzima.** Um bioquímico descobre e purifica uma nova enzima, gerando a tabela de purificação a seguir.

Procedimento	Proteína total (mg)	Atividade (unidades)
1. Extrato bruto	20.000	4.000.000
2. Precipitação (sal)	5.000	3.000.000
3. Precipitação (pH)	4.000	1.000.000
4. Cromatografia de troca iônica	200	800.000
5. Cromatografia de afinidade	50	750.000
6. Cromatografia de exclusão por tamanho	45	675.000

(a) A partir da informação contida na tabela, calcule a atividade específica da enzima após cada procedimento de purificação.

(b) Qual dos procedimentos de purificação utilizados para essa enzima é mais eficaz (i.e. fornece o maior aumento relativo em pureza)?

(c) Qual dos procedimentos de purificação é menos efetivo?

(d) Há alguma indicação, com base nos resultados apresentados na tabela, de que a enzima está pura após a etapa 6? O que mais poderia ser feito para estimar a pureza da preparação da enzima?

**16. Diálise.** Uma proteína purificada está em um tampão Hepes (*N*-(2-hidróxi-etil)piperazina-*N'*-(ácido 2-etanossulfônico)) em pH 7 com 500 mM de NaCl. Uma amostra (1 mL) da solução de proteína é inserida em um tubo feito de membrana de diálise e dialisado contra 1 L do mesmo tampão Hepes com 0 mM de NaCl. Moléculas pequenas e íons (como Na<sup>+</sup>, Cl<sup>-</sup> e Hepes) podem se difundir através da membrana de diálise, mas a proteína não.

(a) Uma vez que a diálise alcança o equilíbrio, qual é a concentração de NaCl na amostra de proteína? Assuma que nenhuma mudança de volume ocorra na amostra durante a diálise.

(b) Se a amostra de 1 mL original fosse dialisada duas vezes, sucessivamente, contra 100 mL do mesmo tampão Hepes com 0 mM NaCl, qual seria a concentração final de NaCl na amostra?

**17. Purificação de peptídeos.** Em pH 7,0, em que ordem os três peptídeos a seguir seriam eluídos em uma coluna preenchida com um polímero permutador de cátions? Suas composições em aminoácidos são:

Peptídeo A: Ala 10%, Glu 5%, Ser 5%, Leu 10%, Arg 10%, His 5%, Ile 10%, Phe 5%, Tyr 5%, Lys 10%, Gly 10%, Pro 5% e Trp 10%.

Peptídeo B: Ala 5%, Val 5%, Gly 10%, Asp 5%, Leu 5%, Arg 5%, Ile 5%, Phe 5%, Tyr 5%, Lys 5%, Trp 5%, Ser 5%, Thr 5%, Glu 5%, Asn 5%, Pro 10%, Met 5% e Cys 5%.

Peptídeo C: Ala 10%, Glu 10%, Gly 5%, Leu 5%, Asp 10%, Arg 5%, Met 5%, Cys 5%, Tyr 5%, Phe 5%, His 5%, Val 5%, Pro 5%, Thr 5%, Ser 5%, Asn 5% e Gln 5%.

**18. Determinação da sequência do peptídeo cerebral leucina encefalina.** Um grupo de peptídeos que influencia a transmissão nervosa em certas partes do cérebro foi isolado de tecido cerebral normal. Esses peptídeos são conhecidos como opioides porque se ligam a receptores específicos que também se ligam a fármacos opiáceos, como a morfina e a naloxona. Os opioides, portanto, mimetizam algumas propriedades dos fár-

macos opiáceos. Alguns pesquisadores consideram que esses peptídeos sejam os analgésicos próprios do cérebro. Utilizando as informações a seguir, determine a sequência de aminoácidos do opioide leucina encefalina. Explique como sua estrutura é consistente com cada uma das informações fornecidas.

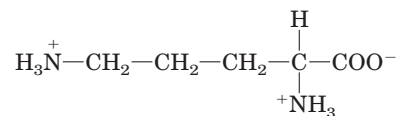
(a) A hidrólise completa por 6 M de HCl a 110°C, seguida pela análise de aminoácidos, indicou a presença de Gly, Leu, Phe e Tyr em uma razão molar de 2:1:1:1.

(b) O tratamento do peptídeo com 1-fluoro-2,4-dinitrobenzeno seguido pela hidrólise completa e cromatografia indicou a presença de um derivado 2,4-dinitrofenila da tirosina. Nenhuma tirosina livre foi encontrada.

(c) A digestão completa do peptídeo com quimotripsina seguida por cromatografia forneceu tirosina e leucina livres mais um tripeptídeo contendo Phe e Gly em uma razão de 1:2.

**19. Estrutura do peptídeo antibiótico de *Bacillus brevis*.** Extratos da bactéria *Bacillus brevis* contêm um peptídeo com propriedades antibióticas. Esse peptídeo forma complexos com íons metálicos e parece interromper o transporte iônico através de membranas celulares de outras espécies bacterianas, matando-as. A estrutura do peptídeo foi determinada a partir das seguintes observações:

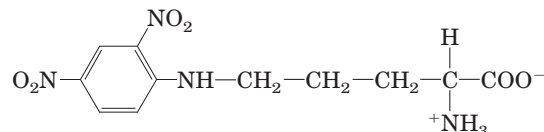
(a) A hidrólise ácida completa do peptídeo seguida de análise de aminoácidos produziu quantidades equimolares de Leu, Orn, Phe, Pro e Val. Orn é ornitina, um aminoácido que não está presente em proteínas, mas aparece em alguns peptídeos. Ela tem a seguinte estrutura:



(b) A massa molecular do peptídeo foi estimada em aproximadamente 1.200.

(c) O peptídeo não sofreu hidrólise quando tratado com a enzima carboxipeptidase. Essa enzima catalisa a hidrólise do resíduo carboxiterminal de um polipeptídeo a menos que o resíduo seja Pro ou, por alguma razão, não contenha um grupo carboxila livre.

(d) O tratamento do peptídeo intacto com 1-fluoro-2,4-dinitrobenzeno, seguido por hidrólise completa e cromatografia, produziu apenas aminoácidos livres e o seguinte derivado:



Dica: o derivado de 2,4-dinitrofenila envolve o grupo amino de uma cadeia lateral em vez de um grupo  $\alpha$ -amino.

(e) A hidrólise parcial do peptídeo seguida por separação cromatográfica e análise de sequência produziu os seguintes di e tripeptídeos (o aminoácido aminoterminal está sempre à esquerda):

Leu-Phe Phe-Pro Orn-Leu Val-Orn

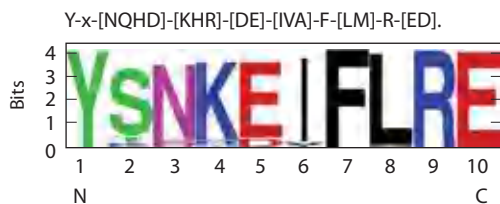
Val-Orn-Leu Phe-Pro-Val Pro-Val-Orn

A partir das informações fornecidas acima, deduza a sequência de aminoácidos do peptídeo antibiótico. Mostre seu raciocínio. Quando você tiver chegado a uma estrutura, demonstre que ela é consistente com *cada* observação experimental.



**20. Eficiência no sequenciamento de peptídeos** Um peptídeo com a estrutura primária Lys-Arg-Pro-Leu-Ile-Asp-Gly-Ala é sequenciado pelo procedimento de Edman. Se cada ciclo de Edman tiver uma eficiência de 96%, que porcentagem dos aminoácidos liberados no quarto ciclo será leucina? Faça o cálculo uma segunda vez, mas presuma uma eficiência de 99% para cada ciclo.

**21. Comparação de sequências** Proteínas denominadas chaperonas moleculares (descritas no Capítulo 4) auxiliam no processo de enovelamento proteico. Uma classe de chaperonas encontrada em organismos desde bactérias a mamíferos é a proteína de choque térmico 90 (Hsp90). Todas as chaperonas Hsp90 contêm uma “sequência-assinatura” de 10 aminoácidos que permite uma identificação rápida dessas proteínas em bancos de dados de sequências. Duas representações dessa sequência-assinatura são apresentadas a seguir.



(a) Nesta sequência, que resíduos de aminoácidos não variam (conservados ao longo de todas as espécies)?

(b) A qual(is) posição(ões) estão limitados aqueles aminoácidos com cadeias laterais carregadas positivamente? Para cada posição, qual aminoácido é mais comumente encontrado?

(c) Em quais posições as substituições estão restritas a aminoácidos com cadeias laterais carregadas negativamente? Para cada posição, qual aminoácido predomina?

(d) Há uma posição que pode ser qualquer aminoácido, embora um aminoácido apareça com muito mais frequência do que qualquer outro. Que posição é esta, e qual aminoácido aparece com mais frequência?

**22. Métodos cromatográficos** Três aminoácidos, cujas sequências são apresentadas a seguir, utilizando o código de uma letra para seus aminoácidos, estão presentes em uma mistura:

1. ATKNRASCLVPHKALMFWRHKQLVSDPILQKRQHILVCRNAAG
2. GPYFGDEPLDVHDEPEEG
3. PHLLSAWKGMGEGVSKSQAALIVILA

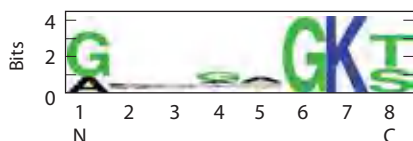
Qual deles migraria mais lentamente durante a cromatografia através de:

(a) uma resina de troca iônica; grânulos revestidos com grupos carregados positivamente?

(b) uma resina de troca iônica; grânulos revestidos com grupos carregados negativamente?

(c) uma coluna de exclusão por tamanho (filtração em gel) projetada para separar peptídeos pequenos como esses?

(d) Quais os peptídeos que contêm os motivos de ligação ao ATP mostrados na sequência logo a seguir?

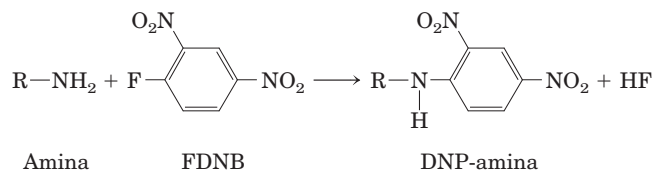


**Problemas de análise de dados**

**23. Determinação da sequência de aminoácidos da insulina** A Figura 3-24 mostra a sequência de aminoácidos da insulina bovina. Essa estrutura foi determinada por Frederick Sanger e colaboradores. A maior parte desse trabalho está descrita em uma série de artigos publicados no *Biochemical Journal* de 1945 a 1955.

Quando Sanger e seus colaboradores iniciaram seu trabalho em 1945, sabia-se que a insulina era uma proteína pequena consistindo em duas ou quatro cadeias polipeptídicas ligadas por ligações dissulfeto. Sanger e seus colaboradores desenvolveram alguns poucos métodos simples para o estudo de sequências de proteínas.

*Tratamento com FDNB.* O FDNB (1-fluoro-2,4-dinitrobenzeno) reage com grupos amino livres (exceto amida ou guanidina) em proteínas para produzir derivados dinitrofenil de aminoácidos:



*Hidrólise ácida.* Ferver uma proteína na presença de HCl a 10% por várias horas hidrolisa todas as suas ligações peptídicas e amídicas. Tratamentos curtos produzem polipeptídeos curtos; quanto mais longo o tratamento, mais completa é a quebra da proteína em seus aminoácidos.

*Oxidação de cisteínas.* O tratamento de uma proteína com ácido perbórico clivou todas as ligações dissulfeto e converteu todos os resíduos Cys a resíduos de ácido cisteico (ver Figura 3-28).

*Cromatografia em papel.* Esta versão mais primitiva da cromatografia em camada delgada (ver Figura 10-25) separava compostos com base em suas propriedades químicas, permitindo a identificação de aminoácidos isolados e, em alguns casos, dipeptídeos. A cromatografia em camada delgada também separa peptídeos maiores.

Como relatado em seu primeiro artigo (1945), Sanger promoveu a reação da insulina com o FDNB e hidrolisou a proteína resultante. Ele encontrou muitos aminoácidos livres, mas apenas três aminoácidos-DNP: α-DNP-glicina (o grupo DNP ligado ao grupo α-amino); α-DNP-fenilalanina; e ε-DNP-lisina (DNP ligado ao grupo ε-amino). Sanger interpretou esses resultados indicando que a insulina tinha duas cadeias proteicas: uma com Gly em sua extremidade aminoterminal e outra com Phe em sua extremidade aminoterminal. Uma das duas cadeias também continha um resíduo Lys, mas não na extremidade aminoterminal. Ele nomeou a cadeia iniciada com o resíduo Gly de “A” e a cadeia iniciada com Phe de “B”.

(a) Explique como os resultados de Sanger apoiam suas conclusões.

(b) Esses resultados são consistentes com a estrutura conhecida da insulina bovina (ver Figura 3-24)?

Em um artigo posterior (1949), Sanger descreveu como ele utilizou essas técnicas para determinar os primeiros poucos aminoácidos (extremidade aminoterminal) de cada cadeia de insulina. Para analisar a cadeia B, por exemplo, ele seguiu as seguintes etapas:



- Oxidou a insulina para separar as cadeias A e B.
- Preparou uma amostra de cadeia B pura por cromatografia em papel.
- Reagiu a cadeia B com FDNB.
- Submeteu a proteína à hidrólise ácida branda, de modo a produzir peptídeos pequenos.
- Separou os peptídeos-DNP dos peptídeos que não continham grupos DNP.
- Isolou quatro dos peptídeos-DNP, os quais foram nomeados B1 a B4.
- Submeteu à hidrólise intensa cada peptídeo-DNP para obter os aminoácidos livres.
- Identificou os aminoácidos em cada peptídeo por cromatografia em papel.

Os resultados foram os seguintes:

B1: apenas  $\alpha$ -DNP-fenilalanina

B2:  $\alpha$ -DNP-fenilalanina; valina

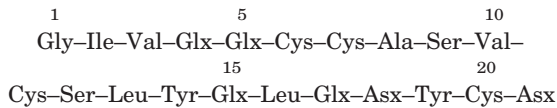
B3: ácido aspártico;  $\alpha$ -DNP-fenilalanina; valina

B4: ácido aspártico; ácido glutâmico;  $\alpha$ -DNP-fenilalanina; valina

(c) Com base nesses dados, quais são os primeiros quatro aminoácidos (aminoterminais) da cadeia B? Explique seu raciocínio.

(d) Esse resultado coincide com a sequência conhecida da insulina bovina (ver Figura 3-24)? Explique quaisquer discrepâncias.

Sanger e colaboradores utilizaram esses e outros métodos relacionados para determinar a sequência completa das cadeias A e B. Suas sequências para a cadeia A foram as seguintes (aminoterminal à esquerda):



Como a hidrólise ácida converteu todo Asn a Asp e todo Gln a Glu, esses resíduos tiveram de ser denominados Asx e Glx, respectivamente (a identidade exata no peptídeo desconhecida). Sanger resolveu esse problema utilizando enzimas proteases que clivam ligações peptídicas, mas não as ligações amídicas nos resíduos Asn e Gln, para preparar peptídeos curtos. Ele então determinou o número de grupos amida presentes em cada peptídeo medindo a liberação de  $\text{NH}_4^+$  quando o peptídeo era hidrolisado em ácido. Alguns dos resultados obtidos para a cadeia A são mostrados a seguir. Os peptídeos podem não ter sido completamente puros, de modo que os números foram aproximados – mas bons o bastante para os propósitos de Sanger.

Nome do peptídeo	Sequência peptídica	Número de grupos amida no peptídeo
Ac1	Cys-Asx	0,7
Ap15	Tyr-Glx-Leu	0,98
Ap14	Tyr-Glx-Leu-Glx	1,06
Ap3	Asx-Tyr-Cys-Asx	2,10
Ap1	Glx-Asx-Tyr-Cys-Asx	1,94
Ap5pa1	Gly-Ile-Val-Glx	0,15
Ap5	Gly-Ile-Val-Glx-Glx-Cys-Cys-Ala-Ser-Val-Cys-Ser-Leu	1,16

(e) Com base nesses dados, determine a sequência de aminoácidos da cadeia A. Explique como você obteve sua resposta e a compare com a Figura 3-24.

## Referências

**Sanger, F.** (1945) The free amino groups of insulin. *Biochem. J.* **39**, 507–515.

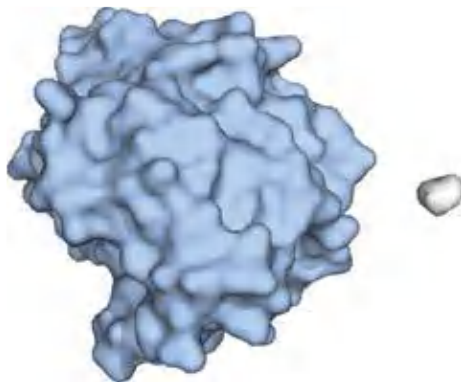
**Sanger, F.** (1949) The terminal peptides of insulin. *Biochem. J.* **45**, 563–574.

# Estrutura Tridimensional de Proteínas

Início leituras complementar e avançada

- 4.1 Visão geral sobre a estrutura das proteínas 115
- 4.2 Estrutura secundária das proteínas 119
- 4.3 Estruturas terciária e quaternária das proteínas 125
- 4.4 Desnaturação e enovelamento das proteínas 143

As proteínas são moléculas grandes. O esqueleto covalente de uma proteína clássica é formado por centenas de ligações simples. Como é possível a livre rotação entre várias dessas ligações, a proteína consegue, em princípio, assumir um número de conformações praticamente incontáveis. Entretanto, cada proteína tem uma função química e uma estrutura específica, sugerindo que cada uma delas tenha uma estrutura tridimensional única (**Figura 4-1**). Quão estável é essa estrutura, quais fatores guiam sua formação e o que a mantém unida? No final de 1920, várias proteínas foram cristalizadas, incluindo a hemoglobina ( $M_r$  64.500) e a enzima urease ( $M_r$  483.000). Como, geralmente, o arranjo ordenado das moléculas em um cristal pode ocorrer somente se as unidades moleculares forem idênticas, a descoberta de que várias proteínas poderiam ser cristalizadas era a evidência de que até mesmo proteínas muito grandes são entidades químicas separadas com estruturas únicas. Essa conclusão revolucionou o pensamento sobre as proteínas e suas funções, mas o conhecimento que ela gerou foi incompleto. A estrutura proteica é sempre ma-



**FIGURA 4-1** Estrutura da enzima quimotripsina, uma proteína globular. A molécula de glicina (em cinza) é representada para comparação de tamanho. As estruturas tridimensionais conhecidas das proteínas estão arquivadas no *Protein Data Bank* – PDB (ver Quadro 4-4). A imagem mostrada aqui foi elaborada utilizando os dados do PDB ID 6GCH.

leável, algumas vezes de forma surpreendente. Mudanças na estrutura podem ser tão importantes para a função da proteína quanto a estrutura por si só.

Neste capítulo, será examinada a estrutura das proteínas. São enfatizados seis temas. Primeiro, a estrutura tridimensional, ou seja, as estruturas de uma proteína são determinadas por sua sequência de aminoácidos. Segundo, a função de uma proteína típica depende de sua estrutura. Terceiro, a maior parte das proteínas isoladas existem em um ou em um pequeno número de formas estruturalmente estáveis. Quarto, as forças mais importantes de estabilização das estruturas específicas de uma dada proteína são as interações não covalentes. Quinto, dentre desse enorme número de estruturas únicas das proteínas, podem ser reconhecidos alguns padrões estruturais comuns, que ajudam a organizar o entendimento sobre a arquitetura das proteínas. Sexto, as estruturas proteicas não são estáticas. Todas as proteínas passam por mudanças na conformação variando desde sutis até bastante significativas. Partes de muitas proteínas possuem estruturas não discerníveis. Para algumas proteínas, a ausência de estrutura definida é fundamental para sua função.

## 4.1 Visão geral sobre a estrutura das proteínas

O arranjo espacial dos átomos em uma proteína ou qualquer parte da proteína é chamado de **conformação**. As conformações possíveis de uma proteína ou de qualquer segmento proteico incluem qualquer estado estrutural que ela possa assumir sem a quebra de suas ligações covalentes. Uma mudança conformacional pode ocorrer, por exemplo, pela rotação sobre as ligações simples. Das várias conformações teoricamente possíveis para uma proteína com centenas de ligações simples, uma ou (mais comumente) poucas predominam em condições biológicas. A necessidade de múltiplas conformações estáveis reflete as mudanças que devem ocorrer na proteína quando ela se liga a outras moléculas ou catalisa reações. As conformações que existem em determinadas condições são, normalmente, aquelas termodinamicamente mais estáveis – isto é, aquelas com energia livre de Gibbs ( $G$ ) menores. Proteínas dobradas, em qualquer uma de suas conformações funcionais, são chamadas de proteínas **nativas**.

Para a grande maioria das proteínas, uma estrutura em particular ou um pequeno grupo de estruturas é crucial para a função. No entanto, em muitos casos, partes das proteínas carecem de estruturas perceptíveis. Esses seg-

mentos proteicos são intrinsecamente desordenados. Em alguns casos, proteínas inteiras são intrinsecamente desordenadas, e ainda assim funcionais.

Quais os princípios que determinam as conformações mais estáveis de uma proteína típica? Uma compreensão da conformação de proteínas pode ser construída, passo a passo, a partir da discussão sobre estrutura primária no Capítulo 3, passando pela consideração das estruturas secundárias, terciárias e quaternárias. A essa abordagem clássica, é preciso acrescentar a ênfase mais recente dada aos padrões comuns e classificáveis de enovelamento, variavelmente chamados de estruturas supersecundárias, enovelamentos ou motivos, que estabelecem um importante contexto organizacional para esse esforço complexo. A título de introdução, serão apresentados alguns princípios básicos.

### A conformação de uma proteína é estabilizada por interações fracas

No contexto da estrutura de proteínas, o termo **estabilidade** pode ser definido como a tendência em manter a conformação nativa. Proteínas nativas são apenas marginalmente estáveis: o  $\Delta G$  que separa os estados dobrados e não dobrados em proteínas comuns, sob condições fisiológicas, está na faixa de apenas 20 a 65 kJ/mol. Uma dada cadeia polipeptídica pode, teoricamente, assumir inúmeras conformações e, como resultado, o estado não dobrado de uma proteína é caracterizado por um alto grau de entropia conformacional. Essa entropia, junto com as interações de ligações de hidrogênio dos diversos grupos da cadeia polipeptídica com o solvente (água), tendem a manter o estado não dobrado. As interações químicas que contrabalançam esses efeitos e estabilizam a conformação nativa incluem ligações dissulfeto (covalentes) e interações fracas (não covalentes), descritas no Capítulo 2: ligações de hidrogênio e interações hidrofóbicas e iônicas.

Várias proteínas não têm ligações dissulfeto. O ambiente dentro da maioria das células é altamente redutor devido à alta concentração de agentes redutores, como a glutatona, e a maior parte das sulfidrilas permanece então no estado reduzido. Fora da célula, o ambiente é frequentemente mais oxidante e a formação de dissulfeto é mais provável de ocorrer. Em eucariotos, as ligações dissulfeto são encontradas, principalmente, em proteínas secretadas, extracelulares (p. ex., o hormônio insulina). As ligações dissulfeto também são incomuns em proteínas de bactérias. Entretanto, bactérias termofílicas, assim como arqueobactérias, geralmente apresentam várias proteínas com ligações dissulfeto, que as estabilizam. Presumivelmente, isso é uma adaptação para a vida a altas temperaturas.

Para todas as proteínas de todos os organismos, as interações fracas são especialmente importantes para o enovelamento das cadeias polipeptídicas em suas estruturas secundárias e terciárias. A associação de múltiplos polipeptídeos para formar estruturas quaternárias também tem como base estas interações fracas.

Aproximadamente 200 a 460 kJ/mol são necessários para quebrar uma ligação covalente simples, enquanto interações fracas podem ser rompidas com apenas 0,4 a 30 kJ/mol. Individualmente, uma ligação covalente, como as

ligações dissulfeto conectando regiões distintas de uma única cadeia polipeptídica, é claramente muito mais forte que uma interação fraca. Entretanto, por serem muito numerosas, são as interações fracas que predominam como forças estabilizadoras da estrutura proteica. Em geral, a conformação proteica de energia livre mais baixa (i.e., de conformação mais estável) é aquela com o número máximo de interações fracas.

A estabilidade de uma proteína não é simplesmente o somatório das energias livres de formação das diversas interações fracas internas. Para cada ligação de hidrogênio formada em uma proteína durante seu enovelamento, uma ligação de hidrogênio (de força equivalente) entre o mesmo grupo e a água é quebrada. A estabilidade resultante da contribuição de uma dada ligação de hidrogênio, ou a *diferença* de energia livre entre os estados dobrado e não dobrado, deve ser próxima de zero. Interações iônicas podem ser tanto estabilizadoras, quanto desestabilizadoras. Portanto, é preciso olhar em outros lugares para entender por que uma determinada conformação nativa é favorável.

A partir de um exame cuidadoso da contribuição das interações fracas na estabilidade das proteínas, fica evidente que as **interações hidrofóbicas** geralmente predominam. A água pura contém moléculas de  $H_2O$  formando uma rede de ligações de hidrogênio. Nenhuma outra molécula tem o potencial de ligação de hidrogênio da água, e a presença de outras moléculas na solução aquosa rompe estas ligações de hidrogênio. Quando a água envolve uma molécula hidrofóbica, o arranjo ótimo de ligação de hidrogênio resulta em uma camada altamente estruturada, ou **camada de solvatação**, de água em torno da molécula (ver Figura 2-7). O aumento da ordem das moléculas de água na camada de solvatação está correlacionado com uma redução desfavorável na entropia da água. Entretanto, quando grupos apolares se agrupam, o tamanho da camada de solvatação diminui, porque cada grupo não mais expõe toda sua superfície à solução. O resultado é um aumento favorável de entropia. Como descrito no Capítulo 2, esse aumento de entropia é a principal força termodinâmica que rege a associação de grupos hidrofóbicos em solução aquosa. Cadeias laterais de aminoácidos hidrofóbicos tendem a se agrupar no interior das proteínas, longe da água (pense em uma gota de óleo na água). A sequência de aminoácidos da maioria das proteínas, assim, apresenta um conteúdo significativo de cadeias laterais de aminoácidos hidrofóbicos (especialmente Leu, Ile, Val, Phe e Trp). Posicionam-se de forma a se aglomerar quando a proteína é dobrada, formando um núcleo hidrofóbico da proteína.

Sob condições fisiológicas, a formação de ligações de hidrogênio em uma proteína é em grande parte dirigida pelo mesmo efeito entrópico. Grupos polares normalmente podem formar ligações de hidrogênio com a água, e, por isso, são solúveis em água. Entretanto, o número de ligações de hidrogênio por unidade de massa normalmente é maior para a água pura do que para qualquer outro líquido ou solução, e há limites de solubilidade até para as moléculas mais polares, pois sua presença causa uma diminuição no número total de ligações de hidrogênio por unidade de massa. Portanto, uma camada de solvatação, até certo ponto, também se forma em torno de moléculas polares. Apesar

de a energia de formação de uma ligação de hidrogênio intramolecular entre dois grupos polares em uma macromolécula ser em grande parte anulada pela eliminação de tais interações entre esses grupos polares e a água, a liberação da água estruturada, na forma de interações intramoleculares, garante a força entrópica que leva ao enovelamento. A maior parte da variação de energia livre na forma de interações fracas dentro da proteína é assim originária do aumento de entropia na solução aquosa circundante, resultante do confinamento das superfícies hidrofóbicas. Isso mais do que contrabalança a grande perda de entropia conformacional, pois o polipeptídeo é limitado à sua conformação dobrada.

As interações hidrofóbicas são importantes na estabilização da conformação: o interior de uma proteína geralmente é um núcleo altamente empacotado de cadeias laterais de aminoácidos hidrofóbicos. Também é importante que cada grupo polar ou carregado no interior da proteína tenha um par adequado para fazer ligação de hidrogênio ou interação iônica. Uma ligação de hidrogênio parece contribuir pouco para a estabilidade de uma estrutura nativa, mas a presença de grupos que fazem ligações de hidrogênio sem par no núcleo hidrofóbico de uma proteína pode ser tão *desestabilizadora* que conformações contendo esse grupo são termodinamicamente insustentáveis. A variação favorável de energia livre resultante da combinação de vários desses grupos com parceiros na solução que os circunda pode ser maior do que a diferença de energia livre entre os estados dobrados e não dobrados. Além disso, ligações de hidrogênio entre grupos em uma proteína se formam cooperativamente (a formação de uma torna mais provável a formação da próxima) em estruturas secundárias repetidas que otimizam as ligações de hidrogênio, como descrito a seguir. Dessa forma, as ligações de hidrogênio normalmente têm um importante papel na condução do processo de enovelamento de proteínas.

A interação entre grupos carregados com cargas opostas, que formam um par iônico ou uma ponte salina, pode exercer tanto um efeito estabilizante quanto desestabilizante na estrutura da proteína. Como no caso das ligações de hidrogênio, cadeias laterais de aminoácidos carregados interagem com a água e com sais quando a proteína não está dobrada, e a perda dessas interações deve ser considerada quando se avalia o efeito da ponte salina na estabilidade geral de uma proteína dobrada. Entretanto, a força de uma ponte salina aumenta à medida que se desloca para um ambiente com constante dielétrica mais baixa,  $\epsilon$  (p. 50): do solvente aquoso polar ( $\epsilon$  próximo a 80) para o interior apolar da proteína ( $\epsilon$  próximo a 4). Pontes salinas, especialmente aquelas parcial ou totalmente internas na proteína, podem, assim, proporcionar uma estabilização significativa da estrutura de uma proteína. Essa tendência explica o aumento da ocorrência de pontes salinas internas nas proteínas de organismos termofílicos. Interações iônicas também limitam a flexibilidade estrutural e conferem uma singularidade a uma determinada estrutura proteica que as interações hidrofóbicas não específicas não conseguem proporcionar.

No ambiente atômico altamente empacotado de uma proteína, mais um tipo de interação fraca pode ter um efeito significativo – as interações de van der Waals (p. 54).

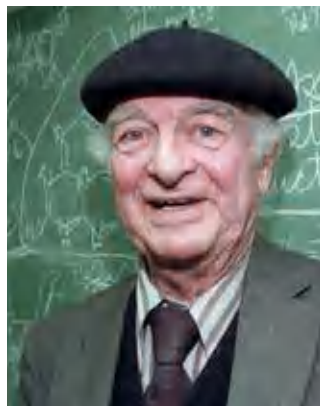
As interações de van der Waals são interações dipolo-dipolo envolvendo os dipolos elétricos permanentes de grupos, tal como as carbonilas, dipolos transitórios derivados das flutuações das nuvens de elétrons em torno de qualquer átomo, e dipolos induzidos pela interação de um átomo com outro que contém um dipolo permanente ou transitório. À medida que os átomos interagem um com o outro, essas interações dipolo-dipolo fornecem uma força intermolecular atrativa que opera apenas sobre uma distância intermolecular limite (0,3 a 0,6 nm). As interações de van der Waals são fracas e individualmente contribuem pouco para a estabilidade da proteína em geral. No entanto, em uma proteína bem empacotada ou na interação de uma proteína com outra proteína ou com outra molécula em uma superfície complementar, o número de tais interações pode ser substancial.

A maioria dos padrões estruturais resumidos neste capítulo reflete duas regras simples: (1) resíduos hidrofóbicos estão basicamente escondidos no interior da proteína, longe da água, e (2) o número de ligações de hidrogênio dentro da proteína é maximizado, reduzindo assim o número de grupos capazes de fazer ligações de hidrogênio e grupos iônicos que não estão adequadamente pareados. As proteínas de membrana (examinadas no Capítulo 11) e proteínas intrinsecamente desordenadas ou que têm segmentos intrinsecamente desordenados seguem regras diferentes. Isso reflete suas funções ou ambientes específicos, mas as interações fracas ainda são elementos estruturais importantes. Por exemplo, proteínas solúveis, mas com segmentos intrinsecamente desordenados, são ricas em cadeias laterais de aminoácidos carregados (especialmente Arg, Lys e Glu) ou pequenos (Gly e Ala), gerando pouca ou nenhuma oportunidade para formação do núcleo hidrofóbico estável.

## A ligação peptídica é rígida e planar

**Arquitetura das proteínas – Estrutura primária** As ligações covalentes também impõem importantes restrições na conformação de um polipeptídeo. No final de 1930, Linus Pauling e Robert Corey iniciaram uma série de estudos que lançaram os fundamentos do entendimento atual sobre estrutura de proteínas. Eles começaram com uma cuidadosa análise da ligação peptídica.

Os carbonos  $\alpha$  de resíduos adjacentes de aminoácidos são separados por três ligações covalentes, arranjados na



Linus Pauling, 1901-1994



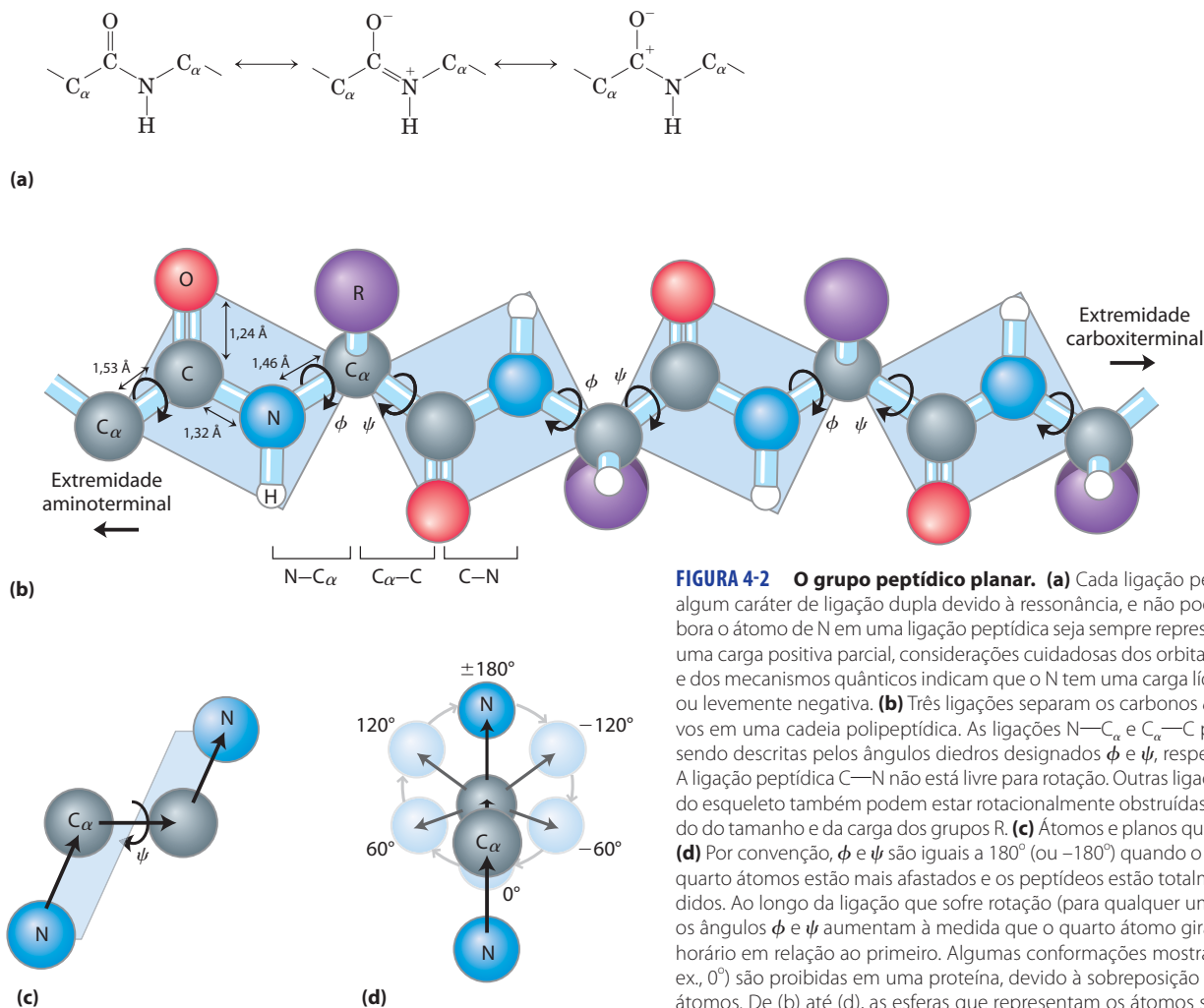
Robert Corey, 1897-1971



forma  $C_\alpha-C-N-C_\alpha$ . Estudos de difração de raios X de cristais de aminoácidos e de dipeptídeos e tripeptídeos simples mostraram que a ligação peptídica  $C-N$  é de alguma forma mais curta que a ligação  $C-N$  de uma amina simples, e que os átomos associados à ligação peptídica são planares. Isso indicava a ressonância ou o compartilhamento parcial de dois pares de elétrons entre o oxigênio carbonílico e o nitrogênio da amida (**Figura 4-2a**). O oxigênio tem uma carga parcial negativa e o hidrogênio ligado ao nitrogênio tem uma carga líquida parcial positiva, formando um pequeno dipolo elétrico. Os seis átomos do **grupo peptídico** estão em um único plano, com o átomo de oxigênio do grupo carbonílico *trans* ao átomo de hidrogênio do nitrogênio da amida. A partir destas observações, Pauling e Corey concluíram que as ligações peptídicas  $C-N$  não podem girar livremente, devido ao seu caráter parcial de ligação dupla. A rotação é permitida ao redor das ligações  $N-C_\alpha$  e  $C_\alpha-C$ . O esqueleto de uma cadeia polipeptídica pode, então, ser descrito como uma série de planos rígidos, com planos consecutivos compartilhando um ponto comum de rotação no  $C_\alpha$  (Figura 4-2b). As ligações peptídicas rígidas limitam a variação de conformações possíveis para uma cadeia polipeptídica.

A conformação da ligação peptídica é definida por três ângulos diedros (também conhecidos como ângulos de torção), chamados de  $\phi$  (phi),  $\psi$  (psi) e  $\omega$  (ômega), que refletem a rotação sobre cada uma das três ligações que se repetem no esqueleto peptídico. Um ângulo diedro é o ângulo da intersecção de dois planos. No caso dos peptídeos, os planos são definidos pelos vetores das ligações do esqueleto peptídico. Dois vetores de ligações sucessivas descrevem um plano. Três vetores de ligações sucessivas descrevem dois planos (o vetor da ligação central é comum a ambos; Figura 4-2c), e o ângulo entre esses dois planos é medido para descrever a conformação da proteína.

**CONVENÇÃO-CHAVE:** Os ângulos diedros importantes para um peptídeo são definidos por três vetores das ligações que conectam quatro átomos consecutivos da cadeia principal (esqueleto peptídico) (Figura 4-2c):  $\phi$  envolve as ligações  $C-N-C_\alpha-C$  (com a rotação ocorrendo entre a ligação  $N-C_\alpha$ ), e  $\psi$  envolve as ligações  $N-C_\alpha-C-N$ . Ambos,  $\phi$  e  $\psi$ , são definidos como  $\pm 180^\circ$  quando o polipeptídeo está completamente estendido e todos os grupos peptídicos estão no mesmo plano (Figura 4-2d). Quando se observa ao longo do vetor da ligação central, na direção da flecha do

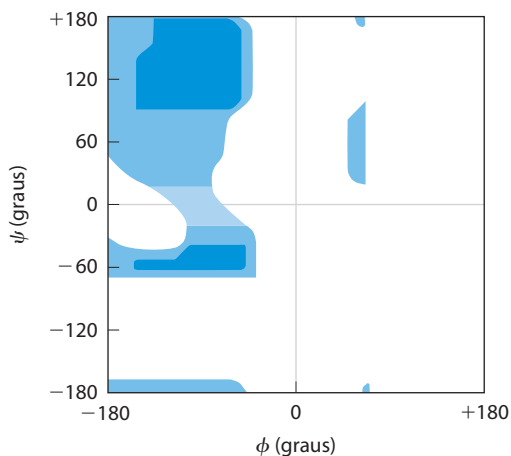


**FIGURA 4-2 O grupo peptídico planar.** (a) Cada ligação peptídica tem algum caráter de ligação dupla devido à ressonância, e não pode girar. Embora o átomo de N em uma ligação peptídica seja sempre representado com uma carga positiva parcial, considerações cuidadosas dos orbitais de ligação e dos mecanismos quânticos indicam que o N tem uma carga líquida neutra ou levemente negativa. (b) Três ligações separam os carbonos  $\alpha$  consecutivos em uma cadeia polipeptídica. As ligações  $N-C_\alpha$  e  $C_\alpha-C$  podem girar, sendo descritas pelos ângulos diedros designados  $\phi$  e  $\psi$ , respectivamente. A ligação peptídica  $C-N$  não está livre para rotação. Outras ligações simples do esqueleto também podem estar rotacionalmente obstruídas, dependendo do tamanho e da carga dos grupos R. (c) Átomos e planos que definem  $\psi$ . (d) Por convenção,  $\phi$  e  $\psi$  são iguais a  $180^\circ$  (ou  $-180^\circ$ ) quando o primeiro e o quarto átomos estão mais afastados e os peptídeos estão totalmente estendidos. Ao longo da ligação que sofre rotação (para qualquer um dos lados), os ângulos  $\phi$  e  $\psi$  aumentam à medida que o quarto átomo gira no sentido horário em relação ao primeiro. Algumas conformações mostradas aqui (p. ex.,  $0^\circ$ ) são proibidas em uma proteína, devido à sobreposição espacial dos átomos. De (b) até (d), as esferas que representam os átomos são menores do que os raios de van der Waals para esta escala.



vetor (como mostrado na Figura 4-2c para  $\psi$ ), os ângulos diedros aumentam à medida que o átomo distal (quarto átomo) gira no sentido horário (Figura 4-2d). A partir das posições  $\pm 180^\circ$ , o ângulo diedro aumenta de  $-180^\circ$  para  $0^\circ$ , o ponto no qual o primeiro e o quarto átomos estão eclipsados. A rotação pode continuar de  $0^\circ$  a  $+180^\circ$  (mesma posição que  $-180^\circ$ ) para retornar a estrutura ao ponto de partida. O terceiro ângulo diedro,  $\omega$ , nem sempre é considerado. Ele envolve as ligações  $C_\alpha-C-N-C_\alpha$ . A ligação central nesse caso é a ligação peptídica, cuja rotação é restrita. A ligação peptídica está, normalmente (99,6% do tempo), na conformação *trans*, restringindo o  $\omega$  a um valor de  $\pm 180^\circ$ . Em um caso raro de ligação peptídica *cis*,  $\omega = 0^\circ$ . ■

A princípio,  $\phi$  e  $\psi$  podem ter qualquer valor entre  $+180^\circ$  e  $-180^\circ$ , mas diversos valores são proibidos por impedimento estérico entre os átomos do esqueleto polipeptídico e as cadeias laterais dos aminoácidos. A conformação na qual ambos,  $\phi$  e  $\psi$ , são  $0^\circ$  (Figura 4-2d) é proibida por esta razão: essa conformação é apenas um ponto de referência para a descrição dos ângulos diedros. Valores permitidos de  $\phi$  e  $\psi$  tornam-se evidentes quando  $\psi$  é colocado em um gráfico *versus*  $\phi$ , no **diagrama de Ramachandran** (Figura 4-3), introduzido por G. N. Ramachandran. Os diagramas de Ramachandran são ferramentas muito úteis e de uso frequente para testar a qualidade de estruturas tridimensionais de proteínas depositadas em bancos de dados internacionais.



**FIGURA 4-3** Diagrama de Ramachandran para resíduos L-Ala. A conformação dos peptídeos é definida pelos valores de  $\phi$  e  $\psi$ . Conformações consideradas possíveis são aquelas que envolvem pouco ou nenhum impedimento estérico, com base nos cálculos dos raios de van der Waals conhecidos e dos ângulos diedros. As áreas coloridas em azul-escuro representam as conformações que não envolvem sobreposição estérica se os raios de van der Waals de cada átomo estão modelados como esferas rígidas, e, portanto, são totalmente permitidas; o azul médio indica as conformações permitidas, se for possível a aproximação dos átomos mais 0,1 nm, um leve choque; o azul-claro indica as conformações que são admitidas se for possível uma pequena flexibilidade (poucos graus) no ângulo diedro  $\omega$  que descreve a própria ligação peptídica (geralmente presa a  $180^\circ$ ). As regiões em branco são conformações não permitidas. A assimetria do diagrama é resultante da estereoquímica L dos resíduos de aminoácidos. Os diagramas para outros resíduos L com cadeias laterais não ramificadas são quase idênticos. Os limites permitidos para resíduos ramificados como Val, Ile e Thr são um pouco menores do que para Ala. O resíduo Gly, que é estericamente menos impedido, apresenta um limite bem mais amplo de conformações permitidas. O limite para os resíduos Pro é muito mais restrito porque seu  $\phi$  é limitado entre  $-35^\circ$  e  $-85^\circ$  pela cadeia lateral cíclica.

#### RESUMO 4.1 Visão geral sobre a estrutura das proteínas

- ▶ Uma proteína típica geralmente tem uma ou mais estruturas tridimensionais, ou conformações que refletem sua função. Algumas proteínas têm segmentos intrinsecamente desordenados.
- ▶ A estrutura da proteína é estabilizada em grande parte por múltiplas interações fracas. As interações hidrofóbicas, derivadas do aumento da entropia da água circundante quando moléculas ou grupos apolares estão agrupados, são os principais contribuintes para a estabilização da forma globular da maioria das proteínas solúveis; as interações de van der Waals também contribuem. As ligações de hidrogênio e interações iônicas são otimizadas nas estruturas termodinamicamente mais estáveis.
- ▶ Ligações covalentes não peptídicas, particularmente ligações dissulfeto, são importantes na estabilização da estrutura de algumas proteínas.
- ▶ A natureza das ligações covalentes no esqueleto polipeptídico estabelece restrições à estrutura. A ligação peptídica tem um caráter parcial de ligação dupla, que mantém todo o grupo peptídico de seis átomos em uma configuração planar rígida. As ligações  $N-C_\alpha$  e  $C_\alpha-C$  podem girar para definir os ângulos diedros  $\phi$  e  $\psi$ , respectivamente.
- ▶ O diagrama de Ramachandran é uma descrição visual das combinações dos ângulos diedros  $\phi$  e  $\psi$  permitidos em um esqueleto peptídico ou não permitidos devido a impedimentos estéricos.

## 4.2 Estrutura secundária das proteínas

O termo **estrutura secundária** se refere a qualquer segmento de uma cadeia polipeptídica e descreve o arranjo espacial de seus átomos na cadeia principal, sem considerar a posição de suas cadeias laterais ou sua relação com outros segmentos. Uma estrutura secundária *comum* ocorre quando cada ângulo diedro,  $\phi$  e  $\psi$ , permanece igual, ou quase igual, ao longo do segmento. Existem alguns tipos de estruturas secundárias que são particularmente estáveis e ocorrem extensamente em proteínas. As mais conhecidas são as hélices  $\alpha$  e as conformações  $\beta$ ; outro tipo comum é a volta  $\beta$ . Quando um padrão regular não é observado, a estrutura secundária algumas vezes é chamada de indefinida ou espiral aleatória. Esta última, entretanto, não descreve adequadamente a estrutura desses segmentos. O curso da maioria dos esqueletos polipeptídicos em uma proteína típica não é aleatório; ele é altamente específico e inalterado na estrutura e na função de uma proteína em particular. Aqui, a discussão se concentra nas estruturas mais comuns.

### A hélice $\alpha$ é uma estrutura secundária comum em proteínas

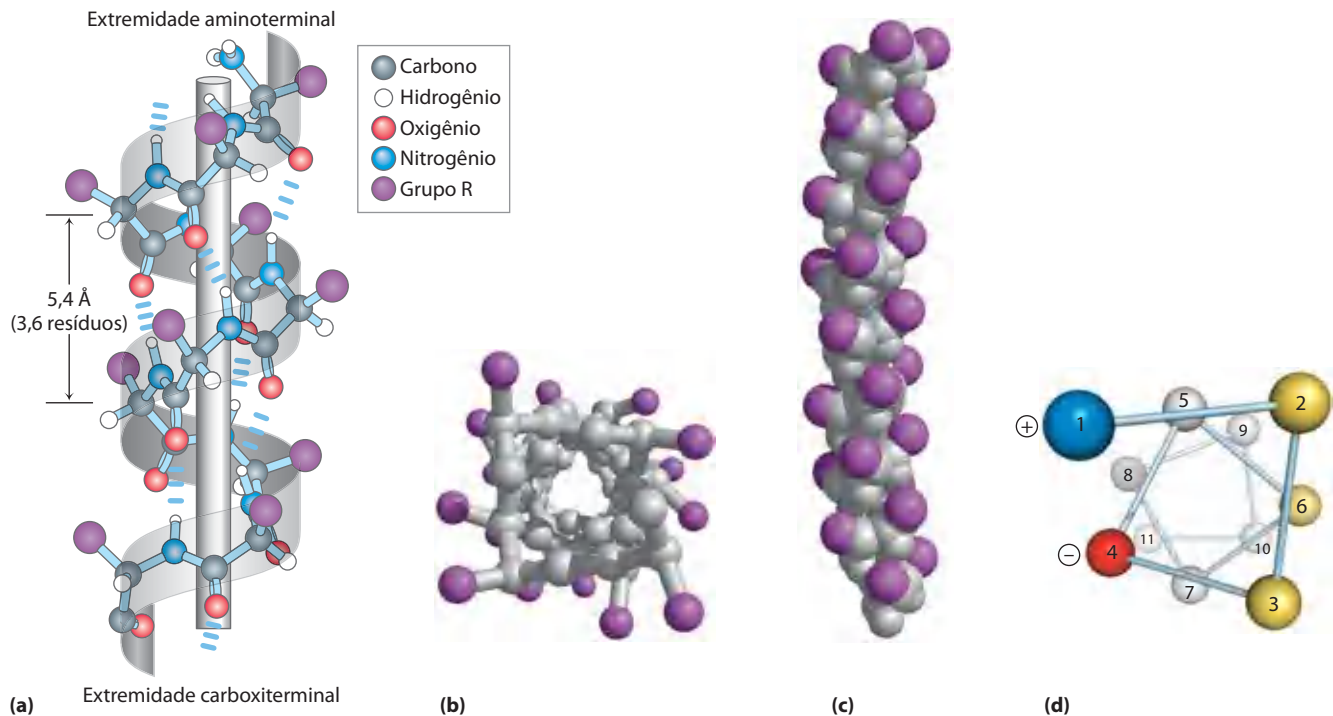
**Arquitetura proteica – hélice  $\alpha$**  Pauling e Corey sabiam da importância das ligações de hidrogênio na orientação de grupos polares como os grupos  $C=O$  e  $N-H$  das ligações peptídicas. Também conheciam os resultados experimentais de William Astbury, que, em 1930, conduziu estudos pio-

neiros de proteínas com raios X. Astbury demonstrou que proteínas que formam os cabelos e os espinhos do porco-espinho (a proteína fibrosa  $\alpha$ -queratina) têm uma estrutura regular que se repete a cada 5,15 a 5,2 Å. (O ångstrom, Å, em homenagem ao físico Anders J. Ångström, é igual a 0,1 nm. Apesar de não ser uma unidade do SI, ela é universalmente utilizada pelos biólogos estruturais para descrever as distâncias atômicas – é aproximadamente o tamanho de uma ligação C—H comum.) Com essa informação e seus dados sobre ligação peptídica, e com a ajuda de modelos construídos de forma precisa, Pauling e Corey iniciaram a determinação das conformações prováveis das moléculas de proteínas.

O primeiro avanço ocorreu em 1948. Pauling foi um professor visitante na Universidade de Oxford, ficou doente e se recolheu a seu apartamento por alguns dias para descansar. Entediado com a leitura disponível, Pauling pegou alguns papéis e lápis para trabalhar em uma estrutura estável plausível que poderia ser adotada por uma cadeia polipeptídica. O modelo que ele desenvolveu, confirmado mais tarde no trabalho com Corey e o colaborador Herman Branson, foi o arranjo mais simples que a cadeia polipeptídica pode assumir que maximiza o uso de ligações de hidrogênio internas. É uma estrutura helicoidal que Pauling e Corey chamaram de **hélice  $\alpha$**  (Figura 4-4). Nessa estrutura, o esqueleto polipeptídico é firmemente enrolado em torno de um eixo imaginário desenhado longitudinalmente no centro

da hélice, e os grupos R dos resíduos de aminoácidos se projetam para fora do esqueleto helicoidal. A unidade que se repete forma uma volta de hélice, que se estende por cerca de 5,4 Å ao longo do eixo, levemente maior do que a periodicidade observada por Astbury na análise por raios X da queratina do cabelo. Os átomos do esqueleto dos resíduos de aminoácidos em uma hélice  $\alpha$  típica têm um grupo característico de ângulos diedros que definem a conformação da hélice  $\alpha$  (Tabela 4-1), e cada volta de hélice é formada por 3,6 resíduos de aminoácidos. Os segmentos de hélice  $\alpha$  em proteínas normalmente se desviam um pouco desses ângulos diedros, podendo até variar dentro de um mesmo segmento, gerando curvaturas ou torções do eixo da hélice. Pauling e Corey consideraram as variantes da hélice  $\alpha$  voltadas tanto para direita quanto para esquerda. A elucidação da estrutura tridimensional subsequente da hélice  $\alpha$  da mioglobina e de outras proteínas mostrou que a hélice  $\alpha$  voltada para direita é a forma comum (Quadro 4-1). As hélices  $\alpha$  estendidas voltadas para esquerda são teoricamente menos estáveis e não foram observadas em proteínas. A hélice  $\alpha$  é a estrutura predominante nas  $\alpha$ -queratins. De forma geral, cerca de um quarto de todos os resíduos de aminoácidos das proteínas é encontrado em hélices  $\alpha$ . A fração exata varia muito de uma proteína para outra.

Por que as hélices  $\alpha$  se formam mais facilmente do que qualquer outra conformação possível? A resposta encontra-se, em parte, no uso otimizado das ligações de hidrogênio



**FIGURA 4-4 Modelos de hélice  $\alpha$ , mostrando os diferentes aspectos de sua estrutura.** (a) Modelo de esfera e bastão mostrando as ligações de hidrogênio internas da cadeia. A unidade que se repete forma uma volta da hélice: 3,6 resíduos. (b) Hélice  $\alpha$  vista de uma de suas extremidades, ao longo do eixo central (obtida a partir do PDB ID 4TNC). Observe as posições dos grupos R, representados pelas esferas roxas. Observe modelo de esfera e bastão, que ressalta o arranjo helicoidal, dá uma falsa impressão de que a hélice é oca, pois as esferas não mostram os raios de van der Waals de cada um dos áto-

mos. (c) Como este modelo de volume atômico mostra, os átomos no centro da hélice  $\alpha$  estão em contato estreito. (d) Projeção da rotação helicoidal de uma hélice  $\alpha$ . Esta representação é colorida para a identificação de superfícies com determinadas propriedades. Os resíduos em amarelo, por exemplo, podem ser hidrofóbicos e fazer parte de uma interface entre a hélice mostrada aqui e outra parte do mesmo ou de outro polipeptídeo. Os resíduos vermelhos (negativo) e azuis (positivo) ilustram o potencial de interação de cadeias laterais de cargas opostas, separadas por dois resíduos na hélice.

**TABELA 4-1** Ângulos  $\phi$  e  $\psi$  ideais para estruturas secundárias comuns em proteínas

Estrutura	$\phi$	$\psi$
Hélice $\alpha$	$-57^\circ$	$-47^\circ$
Conformação $\beta$		
Antiparalela	$-139^\circ$	$+135^\circ$
Paralela	$-119^\circ$	$+113^\circ$
Tripla hélice de colágeno	$-51^\circ$	$+153^\circ$
Volta $\beta$ tipo I		
$i + 1^*$	$-60^\circ$	$-30^\circ$
$i + 2^*$	$-90^\circ$	$0^\circ$
Volta $\beta$ tipo II		
$i + 1$	$-60^\circ$	$+120^\circ$
$i + 2$	$+80^\circ$	$0^\circ$

**Nota:** Nas proteínas reais, os ângulos diedros frequentemente são um pouco diferentes desses valores ideais.

\*Os ângulos  $i + 1$  e  $i + 2$  são aqueles para o segundo e terceiro resíduos de aminoácidos na volta  $\beta$ , respectivamente.

internas. A estrutura é estabilizada por uma ligação de hidrogênio entre o átomo de hidrogênio ligado ao átomo de nitrogênio eletronegativo de uma ligação peptídica e o átomo de oxigênio eletronegativo da carbonila do quarto aminoácido no lado aminoterminal da ligação peptídica (Figura 4-4a). Na hélice  $\alpha$ , cada ligação peptídica (exceto aquelas próximas às extremidades da hélice  $\alpha$ ) participa de tais ligações de hidrogênio. Cada volta sucessiva da hélice  $\alpha$  é mantida por voltas adjacentes por três ou quatro ligações de hidrogênio, conferindo uma significativa estabilidade à estrutura global. Nas extremidades de um segmento  $\alpha$ -helicoidal, sempre há três ou quatro grupos carbonila ou amino que não podem participar desse padrão helicoidal de ligações de hidrogênio. Esses podem estar expostos ao solvente circundante, onde suas ligações de hidrogênio com

a água ou com outras partes da proteína podem proteger a hélice e proporcionar os parceiros necessários para a ligação de hidrogênio.

Outros experimentos mostraram que uma hélice  $\alpha$  pode ser formada em polipeptídeos constituídos de L ou D-aminoácidos. Entretanto, todos os resíduos devem ser de um mesmo estereoisômero; um D-aminoácido irá romper a estrutura regular formada por L-aminoácidos, e vice-versa. A forma mais estável de uma hélice  $\alpha$  de D-aminoácidos é a voltada para a esquerda.

#### PROBLEMA RESOLVIDO 4-1 Estrutura secundária e dimensões de uma proteína

Qual é o comprimento de uma cadeia polipeptídica de 80 resíduos de aminoácidos em uma única hélice  $\alpha$ ?

**Solução:** Uma hélice  $\alpha$  ideal tem 3,6 resíduos por volta, e o avanço ao longo do eixo helicoidal é de 5,4 Å. Portanto, o avanço sobre o eixo, para cada resíduo de aminoácido, é de 1,5 Å. O comprimento do peptídeo é, portanto, 80 resíduos  $\times$  1,5 Å/resíduo = 120 Å.

#### A sequência de aminoácidos afeta a estabilidade da hélice $\alpha$

Nem todos os polipeptídeos podem formar uma hélice  $\alpha$  estável. Cada resíduo de aminoácido em um polipeptídeo tem uma propensão intrínseca de formar uma hélice  $\alpha$  (Tabela 4-2), consequência das propriedades de seu grupo R e como elas interferem na capacidade de seus átomos de conexão da cadeia principal em aceitar os ângulos  $\phi$  e  $\psi$  característicos. A alanina apresenta a melhor tendência a formar hélices  $\alpha$  na maioria dos sistemas-modelo experimentais.

A posição de um resíduo de aminoácido em relação a seus vizinhos também é importante. Interações entre cadeias laterais dos aminoácidos podem estabilizar ou desestabilizar a estrutura  $\alpha$ -helicoidal. Por exemplo, se uma cadeia polipeptídica possui uma longa sequência de resíduos

#### QUADRO 4-1 MÉTODOS Distinção entre o giro para a direita e o giro para a esquerda

Existe um método simples para determinar se uma estrutura helicoidal gira para a direita ou para a esquerda. Faça uma associação com suas duas mãos, com os polegares esticados, apontando para longe de você. Olhando para a sua mão direita, pense em uma hélice se enrolando ao longo de seu polegar direito, na direção dos seus outros dedos, dobrados conforme mostrado na figura (sentido horário). A hélice resultante gira para a direita. Sua mão esquerda representará uma hélice que gira para a esquerda no sentido anti-horário à medida que se enrola em seu polegar.



**TABELA 4-2** Tendência dos resíduos de aminoácidos em assumir a conformação de hélice  $\alpha$ 

Aminoácido	$\Delta\Delta G^\circ$ (kJ/mol)*	Aminoácido	$\Delta\Delta G^\circ$ (kJ/mol)*
Ala	0	Leu	0,79
Arg	0,3	Lys	0,63
Asn	3	Met	0,88
Asp	2,5	Phe	2,0
Cys	3	Pro	> 4
Gln	1,3	Ser	2,2
Glu	1,4	Thr	2,4
Gly	4,6	Tyr	2,0
His	2,6	Trp	2,0
Ile	1,4	Val	2,1

**Fontes:** Dados (exceto para a prolina) de Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X., O'Neil, K.T. e DeGrado, W.F. (1995) Protein design: a hierarchic approach. *Science* **270**, 935. Dados sobre a prolina de Myers, J.K., Pace, C.N. e Scholtz, J.M. (1997) Helix propensities are identical in proteins and peptides. *Biochemistry* **36**, 10, 926.

\* $\Delta\Delta G^\circ$  é a diferença de variação de energia livre, relativa àquela para a alanina, necessária para que os resíduos de aminoácidos assumam a conformação em hélice  $\alpha$ . Valores maiores refletem uma grande dificuldade para assumir a conformação em hélice. O conjunto de dados é derivado de múltiplos experimentos e sistemas experimentais.

Glu, esse segmento da cadeia não irá formar uma hélice  $\alpha$  em pH 7,0. Os grupos carboxílicos, carregados negativamente, dos resíduos Glu adjacentes repelem-se mutuamente de forma tão forte que impedem a formação da hélice  $\alpha$ . Pela mesma razão, se existem muitos resíduos Lys e/ou Arg, com grupos R carregados positivamente em pH 7,0, eles também se repelem, impedindo a formação da hélice  $\alpha$ . O volume e a forma dos resíduos Asn, Ser, Thr e Cys também podem desestabilizar uma hélice  $\alpha$  se estiverem muito próximos na cadeia.

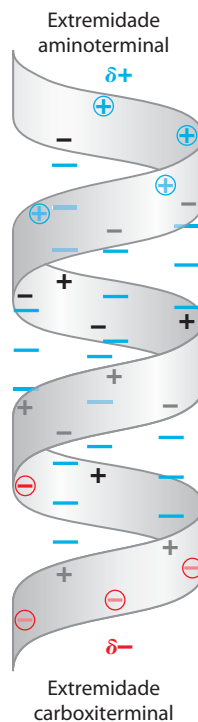
A torção de uma hélice garante que ocorram interações críticas entre a cadeia lateral de um aminoácido e a cadeia lateral do terceiro (às vezes quarto) resíduo adiante, para ambos os lados da hélice. Isso fica claro quando a hélice  $\alpha$  é representada como roda helicoidal (Figura 4-4d). Os aminoácidos carregados positivamente costumam ser encontrados a três resíduos de distância dos aminoácidos carregados negativamente, possibilitando a formação de pares iônicos. Dois aminoácidos aromáticos com frequência são espaçados de forma semelhante, resultando em uma interação hidrofóbica.

Uma restrição à formação da hélice  $\alpha$  é a presença de resíduos de Pro e Gly, que apresentam a menor propensão em formar hélices  $\alpha$ . Na prolina, o átomo de nitrogênio faz parte de um anel rígido (ver Figura 4-8), e a rotação sobre a ligação N—C $_{\alpha}$  não é possível. Dessa forma, um resíduo Pro gera uma torção, que desestabiliza a hélice. Além disso, o átomo de nitrogênio do resíduo Pro em uma ligação peptídica não tem hidrogênio para participar em ligações com outros resíduos. Por essas razões, a prolina raramente é encontrada em uma hélice  $\alpha$ . A glicina com frequência não ocorre em hélices por outro

motivo: ela apresenta maior flexibilidade conformacional do que os outros aminoácidos. Polímeros de glicina tendem a formar estruturas espiraladas bem diferentes de uma hélice  $\alpha$ .

Um último fator que afeta a estabilidade de uma hélice  $\alpha$  é a identidade dos resíduos de aminoácido próximos às extremidades do segmento  $\alpha$ -helicoidal do polipeptídeo. Existe um pequeno dipolo elétrico em cada ligação peptídica (Figura 4-2a). Esses dipolos estão alinhados através das ligações de hidrogênio da hélice, resultando em um dipolo livre ao longo do eixo helicoidal que aumenta com o comprimento da hélice (Figura 4-5). As cargas parcialmente positivas e negativas do dipolo da hélice ocorrem nos grupos amino e carbonil próximos às extremidades amino e carboxiterminal, respectivamente. Por isso, aminoácidos carregados negativamente costumam ser encontrados próximos à extremidade aminoterminal do segmento helicoidal, onde apresentam interações estabilizantes com a carga positiva do dipolo da hélice; um aminoácido positivamente carregado na extremidade aminoterminal desestabilizaria o sistema. O oposto é verdade para a extremidade carboxiterminal do segmento helicoidal.

Resumindo, cinco tipos de restrições influenciam a estabilidade de uma hélice  $\alpha$ : (1) a tendência intrínseca de um resíduo de aminoácido de formar uma hélice  $\alpha$ ; (2) as interações entre os grupos R, especialmente aqueles espaçados por três (ou quatro) aminoácidos; (3) os volumes de grupos R adjacentes; (4) a ocorrência de resíduos Pro e Gly; e (5) interações entre os resíduos de aminoácidos das extremidades do segmento helicoidal e o dipolo elétrico inerente da hélice  $\alpha$ . A tendência de um determinado segmento de uma cadeia polipeptídica de formar uma hélice  $\alpha$  depende, portanto, da identidade e da sequência de resíduos de aminoácidos do segmento.

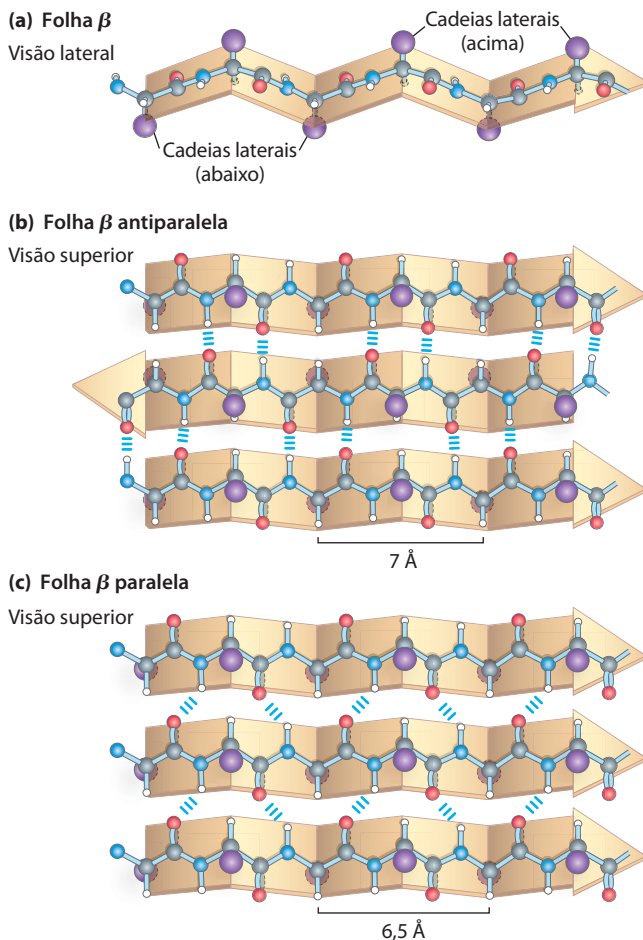


**FIGURA 4-5** Dipolo da hélice. O dipolo elétrico da ligação peptídica (ver Figura 4-2a) é transmitido ao longo do segmento  $\alpha$ -helicoidal pelas ligações de hidrogênio intracadeia, resultando em um dipolo da hélice. Nesta ilustração, os componentes amino e carbonil de cada ligação peptídica estão indicados com os símbolos + e -, respectivamente. Os componentes amino e carbonil das ligações peptídicas próximas às extremidades da região  $\alpha$ -helicoidal, que não participam de ligações de hidrogênio, estão circulares e mostrados em cores.



## As conformações $\beta$ organizam as cadeias polipeptídicas em forma de folha

**Arquitetura proteica – folha  $\beta$**  Em 1951, Pauling e Corey reconheceram um segundo tipo de estrutura recorrente, a **conformação  $\beta$** . Essa é uma conformação mais estendida das cadeias polipeptídicas, e sua estrutura é de novo definida pelos esqueletos dos átomos arranjados de acordo com grupo característico de ângulos diedro (Tabela 4-1). Na conformação  $\beta$ , o esqueleto da cadeia polipeptídica está estendido em forma de zigue-zague, em vez de em estrutura helicoidal (**Figura 4-6**). O arranjo de vários segmentos lado a lado, os quais estão na conformação  $\beta$ , é chamado de **folha  $\beta$** . A estrutura em zigue-zague dos segmentos polipeptídicos individuais dá origem a uma aparência pregueada da folha em geral. As ligações de hidrogênio são formadas entre segmentos adjacentes da cadeia polipeptídica, dentro da folha. Os segmentos que formam a folha  $\beta$  normalmente estão próximos na cadeia polipeptídica, mas também podem estar bem distantes uns dos outros na sequência linear



**FIGURA 4-6** A conformação  $\beta$  das cadeias polipeptídicas. Estas visões (a) lateral e (b, c) superior mostram os grupos R saindo do plano da folha  $\beta$  e enfatizam a forma pregueada formada pelos planos das ligações peptídicas. (Um nome alternativo para esta estrutura é folha  $\beta$  pregueada.) As ligações de hidrogênio entre as cadeias adjacentes também são mostradas. A orientação das cadeias adjacentes (setas), do aminoterminal para carboxiterminal, pode ser a mesma ou oposta, formando (b) folhas  $\beta$  antiparalelas ou (c) folhas  $\beta$  paralelas.

do polipeptídeo; eles podem até estar em cadeias polipeptídicas diferentes. Os grupos R dos aminoácidos adjacentes se projetam da estrutura em zigue-zague em direções opostas, criando um padrão alternado que pode ser observado na visão lateral da Figura 4-6.

As cadeias polipeptídicas adjacentes em uma folha  $\beta$  podem ser tanto paralelas quanto antiparalelas (apresentando uma orientação aminocarboxiterminal igual ou oposta, respectivamente). As estruturas são, de alguma forma, semelhantes, apesar de o período de repetição ser menor na conformação paralela (6,5 Å, *versus* 7 Å para a antiparalela) e o padrão das ligações de hidrogênio ser diferente. As ligações de hidrogênio intersegmentos são alinhadas (ver Figura 2-5) na folha  $\beta$  antiparalela, enquanto elas são distorcidas ou não alinhadas na variante paralela. As estruturas ideais exibem os ângulos de ligação dados na Tabela 4-1; esses valores variam um pouco nas proteínas verdadeiras, resultando em uma variação estrutural, conforme visto anteriormente para as hélices  $\alpha$ .

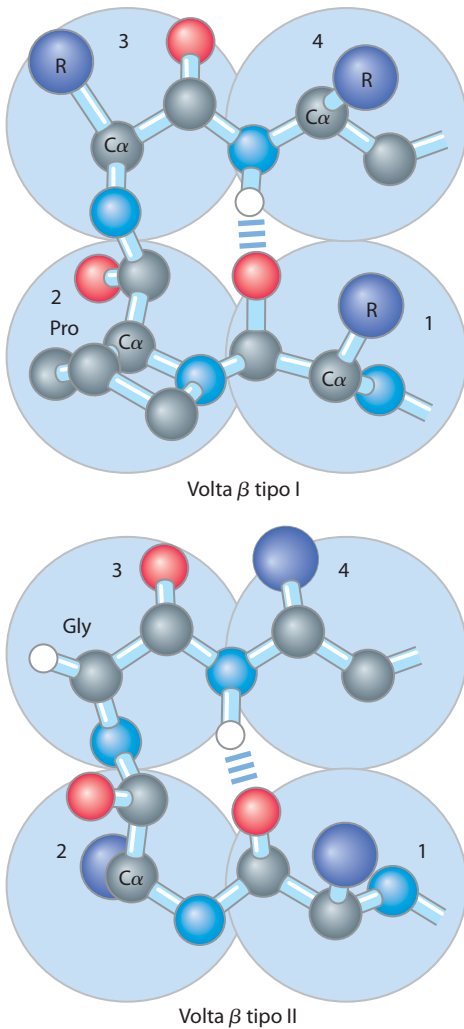
## Volts $\beta$ são comuns em proteínas

**Arquitetura proteica – volts  $\beta$**  Em proteínas globulares, que apresentam estrutura dobrada compacta, alguns resíduos de aminoácidos estão em volts ou alças onde a cadeia polipeptídica inverte sua direção (**Figura 4-7**). Esses são elementos conectores que ligam estruturas sucessivas de hélices  $\alpha$  e conformações  $\beta$ . As **volts  $\beta$**  são particularmente comuns e conectam as extremidades de dois segmentos adjacentes de uma folha  $\beta$  antiparalela. A estrutura é uma volta de 180° que envolve quatro resíduos de aminoácidos, com o oxigênio carbonílico do primeiro resíduo formando uma ligação de hidrogênio com o hidrogênio do grupo amino do quarto resíduo. Os grupos peptídicos dos dois resíduos centrais não participam de nenhuma ligação de hidrogênio inter-resíduos. Vários tipos de volts  $\beta$  têm sido descritas, cada uma definida pelos ângulos  $\phi$  e  $\psi$  das ligações que ligam os quatro resíduos de aminoácidos que formam a volta em particular (Tabela 4-1). Os resíduos Gly e Pro frequentemente ocorrem em volts  $\beta$ ; o primeiro porque é pequeno e flexível, e o último porque as ligações peptídicas envolvendo o nitrogênio imino da prolina facilmente assumem configuração *cis* (**Figura 4-8**), forma particularmente acessível em uma volta fechada. Os dois tipos de volts  $\beta$ , mostrados na Figura 4-7, são os mais comuns. As volts normalmente são encontradas próximas à superfície das proteínas, onde os grupos peptídicos dos dois resíduos de aminoácidos centrais da alça podem fazer ligação de hidrogênio com a água. Muito menos comum é a volta  $\gamma$ , uma volta com três resíduos e ligação de hidrogênio entre o primeiro e o terceiro resíduo.

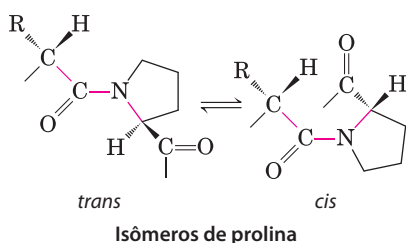
## Estruturas secundárias comuns têm ângulos diedros característicos

As hélices  $\alpha$  e as conformações  $\beta$  são as principais estruturas secundárias que se repetem em um grande número de proteínas, apesar de existirem outras estruturas que se repetem em algumas proteínas especializadas (um exemplo é o colágeno; ver Figura 4-13). Cada tipo de estrutura secundária pode ser completamente descrito pelos ângulos diedros  $\phi$  e  $\psi$ , associados a cada resíduo. Como mos-





**FIGURA 4-7 Estruturas de voltas  $\beta$ .** Voltas  $\beta$  dos tipos I e II são as mais comuns; distinguem-se pelos ângulos  $\phi$  e  $\psi$  adotados pelo esqueleto peptídico na volta (ver Tabela 4-1). As voltas do tipo I ocorrem duas vezes mais do que as voltas do tipo II. As voltas  $\beta$  do tipo II normalmente possuem uma Gly como terceiro resíduo. Observe a ligação de hidrogênio entre os grupos peptídicos do primeiro e do quarto resíduos da volta. (Cada um dos resíduos de aminoácidos está identificado por grandes círculos azuis. Nem todos os átomos de H estão mostrados.)

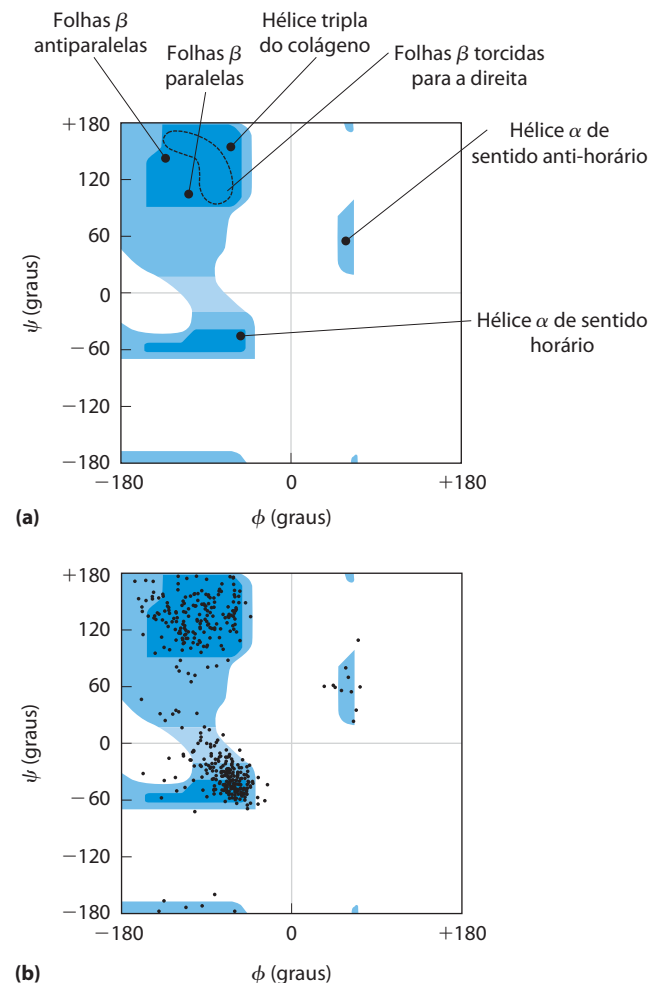


**FIGURA 4-8 Isômeros *trans* e *cis* de uma ligação peptídica envolvendo o nitrogênio imino da prolina.** Das ligações peptídicas entre resíduos de aminoácidos que não Pro, mais de 99,95% estão na configuração *trans*. Para as ligações peptídicas envolvendo o nitrogênio imino da prolina, no entanto, cerca de 6% estão na configuração *cis*; muitas delas ocorrem em voltas  $\beta$ .

trado no diagrama de Ramachandran, os ângulos diedros que definem a hélice  $\alpha$  e a conformação  $\beta$  se encontram em uma região restrita de estruturas estericamente permitidas (**Figura 4-9a**). A maioria dos valores de  $\phi$  e  $\psi$  obtidos de estruturas de proteínas conhecidas cai nas regiões esperadas, com alta concentração próximo aos valores preditos para as hélices  $\alpha$  e as conformações  $\beta$  (Figura 4-9b). O único resíduo de aminoácido normalmente encontrado fora dessas regiões é a glicina. Como sua cadeia lateral é pequena, o resíduo Gly pode assumir diversas conformações estericamente proibidas para os demais aminoácidos.

**As estruturas secundárias comuns podem ser identificadas por dicroísmo circular**

Qualquer forma de assimetria estrutural em uma molécula leva a diferenças de absorção da luz circularmente polari-

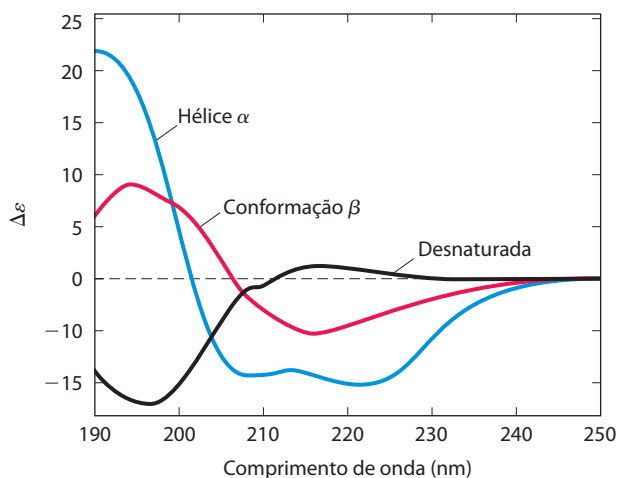


**FIGURA 4-9 Diagrama de Ramachandran mostrando uma variedade de estruturas.** (a) Os valores de  $\phi$  e  $\psi$  de várias estruturas secundárias permitidas estão sobrepostos no diagrama da Figura 4-3. Apesar de as hélices  $\alpha$  de sentido anti-horário com vários resíduos de aminoácidos serem teoricamente permitidas, elas não têm sido observadas em proteínas. (b) Os valores de  $\phi$  e  $\psi$  para todos os aminoácidos, exceto Gly, da enzima piruvato-cinase (isolada de coelho) estão sobrepostos no diagrama das conformações teoricamente permitidas (Figura 4-3). Os pequenos e flexíveis resíduos de Gly foram excluídos, pois em geral caem fora da região esperada (em azul).

zada para a esquerda em relação à direita. A medida desta diferença é chamada de **espectroscopia de dicroísmo circular (CD, de circular dichroism)**. Uma estrutura ordenada, como de uma proteína dobrada, resulta em um espectro de absorção que pode ter picos ou regiões com valores tanto positivos, quanto negativos. Para proteínas, os espectros são obtidos na região de UV distante (190 a 250 nm). A entidade que absorve luz, ou cromóforo, nessa região é a ligação peptídica; um sinal é obtido quando esta ligação peptídica está em um ambiente dobrado. A diferença em coeficiente de extinção molar (ver Quadro 3-1) para a luz circularmente polarizada para a esquerda e para a direita ( $\Delta\epsilon$ ) é colocada no gráfico em função do comprimento de onda. As hélices  $\alpha$  e as conformações  $\beta$  têm espectros de CD característicos (**Figura 4-10**). Com o espectro de CD, os bioquímicos podem determinar se as proteínas estão dobradas corretamente, estimar a fração da proteína que assume qualquer uma das duas estruturas secundárias comuns e monitorar as transições entre os estados dobrados e não dobrados.

#### RESUMO 4.2 Estrutura secundária das proteínas

- ▶ Estrutura secundária é o arranjo espacial dos átomos da cadeia principal em um determinado segmento da cadeia polipeptídica.
- ▶ As estruturas secundárias regulares mais comuns são as hélices  $\alpha$ , as conformações  $\beta$  e as voltas  $\beta$ .
- ▶ A estrutura secundária de um segmento polipeptídico pode ser completamente definida se seus ângulos  $\phi$  e  $\psi$  são conhecidos para todos os aminoácidos do segmento.
- ▶ A espectroscopia de dicroísmo circular é um método para a identificação das estruturas secundárias comuns e o monitoramento do enovelamento das proteínas.



**FIGURA 4-10 Espectroscopia de dicroísmo circular (CD).** Este espectro mostra a polilisina inteiramente como hélice  $\alpha$ , como conformação  $\beta$  ou desnaturada, espiral aleatória. A unidade do eixo y é uma simplificação das unidades comumente utilizadas nos experimentos de CD. Como as curvas são diferentes para as hélices  $\alpha$ , as conformações  $\beta$  e a desnaturada, o espectro de CD fornece a determinada proteína uma estimativa da fração da proteína, formada pelas duas estruturas secundárias mais comuns. O espectro de CD da proteína nativa pode servir como referência para o estado enovelado, e é útil no monitoramento da desnaturação, ou de mudanças conformacionais resultantes de alterações nas condições de solução.

### 4.3 Estruturas terciária e quaternária das proteínas

**Arquitetura proteica – Introdução da estrutura terciária** O arranjo tridimensional total de todos os átomos de uma proteína é chamado de **estrutura terciária**. Enquanto o termo “estrutura secundária” se refere ao arranjo espacial dos resíduos de aminoácidos adjacentes em um segmento polipeptídico, a estrutura terciária inclui aspectos de *alcance mais longo* da sequência de aminoácidos. Aminoácidos que estão bem distantes na sequência polipeptídica e em diferentes tipos de estruturas secundárias podem interagir na estrutura da proteína completamente dobrada. A localização das curvaturas (incluindo as voltas  $\beta$ ) nas cadeias polipeptídicas e sua direção e seu ângulo são determinados pelo número e pela localização de resíduos específicos que tendem a formá-las, como Pro, Thr, Ser e Gly. Segmentos da cadeia polipeptídica que interagem entre si são mantidos em suas posições terciárias características por diferentes tipos de interações fracas (e algumas vezes por ligações covalentes, como ligações dissulfeto) entre os segmentos.

Algumas proteínas contêm duas ou mais cadeias polipeptídicas distintas, ou subunidades, que podem ser idênticas ou diferentes. O arranjo destas subunidades proteicas em complexos tridimensionais constitui a **estrutura quaternária**.

Considerando esses níveis mais altos de estrutura, é conveniente designar dois grandes grupos nos quais muitas proteínas podem ser classificadas: **proteínas fibrosas**, com cadeias polipeptídicas arranjadas em longos filamentos ou folhas, e **proteínas globulares**, com cadeias polipeptídicas dobradas em forma esférica ou globular. Os dois grupos são estruturalmente distintos. As proteínas fibrosas em geral são formadas por um único tipo de estrutura secundária, e sua estrutura terciária é relativamente simples. As proteínas globulares normalmente contêm diversos tipos de estruturas secundárias. Os dois grupos também se diferenciam funcionalmente: as estruturas que garantem suporte, forma e proteção externa aos vertebrados são feitas de proteínas fibrosas, enquanto as enzimas e as proteínas reguladoras em sua maioria são proteínas globulares.

#### As proteínas fibrosas são adaptadas às funções estruturais

**Arquitetura proteica – Estrutura terciária da proteína fibrosa** A  $\alpha$ -queratina, o colágeno e a fibroína da seda ilustram bem a relação entre a estrutura da proteína e sua função biológica (Tabela 4-3). As proteínas fibrosas compartilham propriedades que dão força e/ou flexibilidade às estruturas nas quais ocorrem. Em cada caso, a unidade estrutural fundamental é um elemento simples de estrutura secundária que se repete. Todas as proteínas fibrosas são insolúveis em água, propriedade conferida pela alta concentração de resíduos de aminoácidos hidrofóbicos tanto no interior quanto na superfície da proteína. Essas superfícies hidrofóbicas estão em grande parte escondidas, visto que muitas cadeias polipeptídicas similares são reunidas formando um complexo supramolecular. A simplicidade estrutural das proteínas fibrosas as torna particularmente interessantes para ilus-

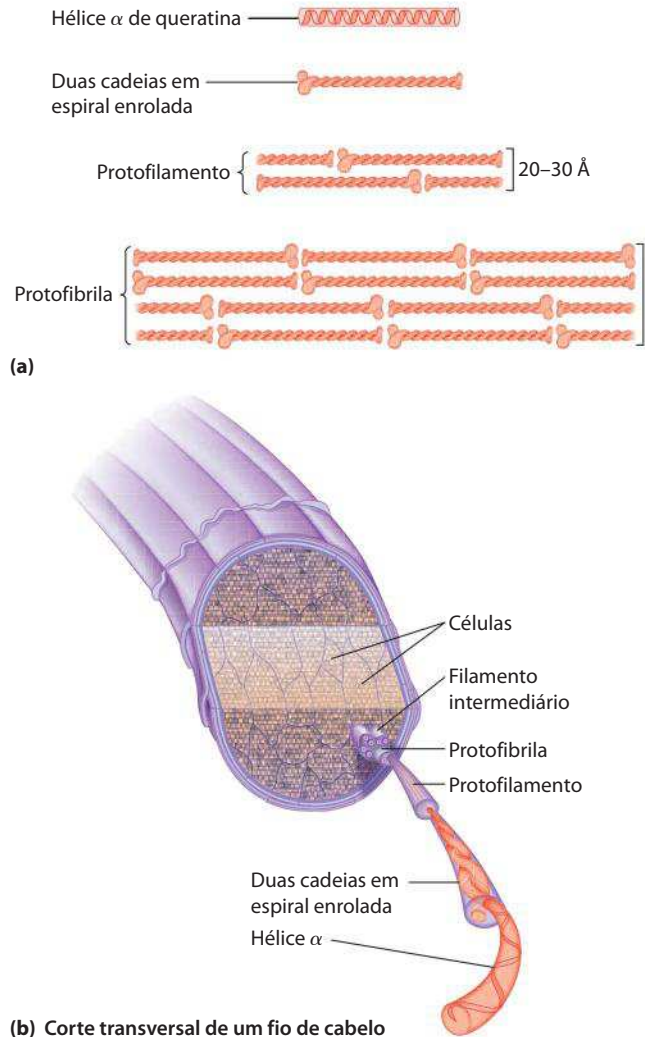
trar alguns dos princípios fundamentais da estrutura proteica discutidos anteriormente.

**α-Queratina** As α-queratinas foram desenvolvidas para força. Encontradas somente em mamíferos, essas proteínas constituem praticamente todo o peso seco de cabelos, pelos, unhas, garras, penas, chifres, cascos e grande parte da camada mais externa da pele. As α-queratinas fazem parte de uma família mais ampla de proteínas chamadas de proteínas de filamento intermediário (FI). Outras proteínas FI são encontradas no citoesqueleto de células animais. Todas as proteínas FI têm função estrutural e compartilham das características estruturais das α-queratinas.

A hélice da α-queratina é uma hélice α voltada para a direita, a mesma hélice encontrada em várias outras proteínas. Francis Crick e Linus Pauling, no início de 1950, sugeriram, independentemente, que as hélices α da queratina estavam arranjadas na forma de espiral enrolada. Duas fibras de α-queratina, orientadas em paralelo (com seus aminoterminais na mesma extremidade), são enroladas uma sobre a outra, formando uma espiral enrolada supertorcida. A supertorção amplifica a força da estrutura como um todo, assim como as fibras são trançadas para formar uma corda forte (Figura 4-11). A torção do eixo de uma hélice α para formar uma espiral enrolada explica a discrepância entre os 5,4 Å por volta preditos por Pauling e Corey, e as estruturas repetidas a cada 5,15 a 5,2 Å observadas na difração de raios X do cabelo (p. 120). O sentido da hélice das estruturas supertorcidas é anti-horário, oposto ao da hélice α. As superfícies onde as duas hélices α se tocam são formadas por resíduos de aminoácidos hidrofóbicos, e seus grupos R se entrelaçam em um padrão regular interconectado. Isso permite um arranjo próximo das cadeias polipeptídicas dentro da estrutura supertorcida no sentido anti-horário. Não é surpresa, portanto, que a α-queratina seja rica em resíduos hidrofóbicos Ala, Val, Leu, Ile, Met e Phe.

Um polipeptídeo na α-queratina com as cadeias em espiral enrolada tem uma estrutura terciária relativamente simples, dominada por uma estrutura secundária α-helicoidal, com seu eixo enrolado em uma super-hélice anti-horária. O entrelaçamento de dois peptídeos α-helicoidais é um exemplo de estrutura quaternária. Espirais enroladas desse tipo são elementos estruturais comuns em proteínas filamentosas e na proteína muscular miosina (ver Figura 5-27). A estrutura quaternária da α-queratina pode ser bem complexa. Várias espirais enroladas podem ser associadas em grandes complexos supramoleculares, como o arranjo da α-queratina para formar o filamento intermediário do cabelo (Figura 4-11b).

A resistência das proteínas fibrosas é aumentada pelas ligações covalentes entre as cadeias polipeptídicas nas “cordas” multi-helicoidais e entre cadeias adjacentes em



**FIGURA 4-11 Estrutura do cabelo.** (a) A α-queratina do cabelo é uma longa hélice α com elementos mais densos próximos às extremidades amino e carboxiterminais. Os pares destas hélices são enrolados em um sentido anti-horário para formar as duas cadeias em espiral enrolada. Estas, por sua vez, se combinam em estruturas mais complexas, chamadas de protofilamentos e protofibrilas. Cerca de quatro protofibrilas – 32 moléculas de α-queratina ao todo – se combinam para formar um filamento intermediário. As duas cadeias em espiral enrolada das várias subestruturas também parecem estar entrelaçadas, mas a orientação destas torções e outros detalhes estruturais não são conhecidos. (b) O cabelo é um conjunto de filamentos de α-queratina, formado pelas subestruturas mostradas em (a).

um arranjo supramolecular. Nas α-queratinas, as ligações entre as cadeias que estabilizam a estrutura quaternária são ligações dissulfeto (Quadro 4-2). Nas α-queratinas mais

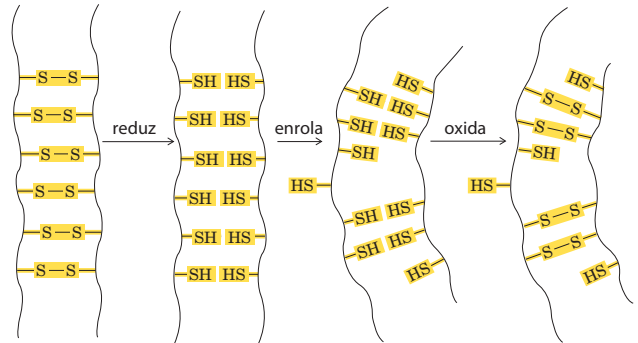
**TABELA 4-3** Estrutura secundária e propriedades de algumas proteínas fibrosas

Estrutura	Características	Exemplos de ocorrência
Hélices α, ligadas por ligações dissulfeto	Estruturas de proteção insolúveis e resistentes, com dureza e flexibilidade variáveis	α-Queratina de cabelos, penas, unhas
Conformação β	Filamentos macios e flexíveis	Fibroínas de seda
Hélice tripla de colágeno	Grande resistência à tração, sem elasticidade	Colágeno de tendões, matriz óssea

**QUADRO 4-2 Ondulação permanente é engenharia bioquímica**

Quando os cabelos são expostos ao calor úmido, eles podem ser esticados. No nível molecular, as hélices  $\alpha$  da  $\alpha$ -queratina do cabelo são esticadas até que cheguem a uma conformação  $\beta$  completamente estendida. Com o resfriamento espontâneo, elas retornam para a conformação  $\alpha$ -helicoidal. A “esticabilidade” característica das  $\alpha$ -queratinas e suas inúmeras ligações dissulfeto constituem a base dos processos da ondulação permanente. Os cabelos que serão cacheados ou encrespados são inicialmente enrolados ao redor de uma forma com um formato adequado. Uma solução de um agente redutor, normalmente um composto contendo um grupo tiol ou sulfidril ( $-\text{SH}$ ), é então aplicada com calor. O agente redutor rompe as ligações transversais pela redução de cada ligação dissulfeto em dois resíduos Cys. O calor úmido quebra as ligações de hidrogênio e provoca o desenrolamento da estrutura da cadeia polipeptídica. Depois de um tempo, a solução redutora é removida, e um agente oxidante é adicionado para fazer *novas* ligações dissulfeto entre pares de resíduos Cys de cadeias adjacentes, mas não dos mesmos pares de antes do tratamento. Depois do cabelo lavado e resfriado, as cadeias polipeptídicas voltam à sua conformação

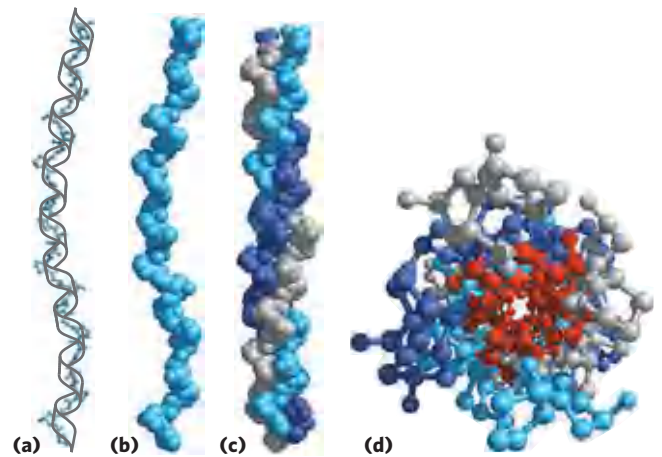
helicoidal. As fibras do cabelo agora enrolam da forma desejada porque as novas ligações dissulfeto transversais exercem uma torção ou rotação sobre os feixes de hélices das fibras. O mesmo processo pode ser usado para o alisamento dos cabelos naturalmente crespos. Um alisamento permanente de cabelo não é permanente de fato, pois o cabelo cresce; e no cabelo novo que substituirá o antigo, a  $\alpha$ -queratina tem o padrão natural de ligações dissulfeto.



duras e resistentes, como as dos chifres dos rinocerontes, até 18% dos resíduos são cisteínas envolvidas em ligações dissulfeto.

**Colágeno** Assim como as  $\alpha$ -queratinas, o colágeno evoluiu para garantir resistência. Ele é encontrado nos tecidos conectivos como os tendões, as cartilagens, a matriz orgânica dos ossos e na córnea dos olhos. A hélice de colágeno é uma estrutura secundária única, bem diferente da hélice  $\alpha$ . Ela gira para esquerda e tem três resíduos de aminoácidos por volta (Figura 4-12 e Tabela 4-1). O colágeno também é uma espiral enrolada, mas com estruturas terciária e quaternária distintas: três polipeptídeos separados, chamados de cadeias  $\alpha$  (não confundir com hélices  $\alpha$ ), são supertorcidos uns sobre os outros (Figura 4-12c). No colágeno, a torção super-helicoidal tem sentido horário, oposto ao da hélice anti-horária das cadeias  $\alpha$ .

Existem vários tipos de colágenos nos vertebrados. Normalmente, eles contêm em torno de 35% de Gly, 11% de Ala e 21% de Pro e 4-Hyp (4-hidroxiprolina, um aminoácido incomum; ver Figura 3-8a). A gelatina comestível é derivada do colágeno. Tem baixo poder nutricional como proteína, pois o colágeno tem uma quantidade muito baixa dos aminoácidos essenciais à dieta humana. O conteúdo de aminoácidos incomuns do colágeno está relacionado com as restrições estruturais únicas de sua hélice. A sequência de aminoácidos no colágeno geralmente é uma repetição de uma unidade tripeptídica, Gly-X-Y, onde X normalmente é uma Pro e Y em geral é uma 4-Hyp. Somente os resíduos Gly podem ser acomodados nas junções muito apertadas entre as cadeias  $\alpha$  individuais (Figura 4-12d). Os resíduos Pro e 4-Hyp permitem a torção acentuada da hélice do colágeno. A sequência de aminoácidos e a estrutura quaternária supertorcida do colágeno permitem uma compactação



**FIGURA 4-12 Estrutura do colágeno.** (Obtida do PDB ID 1CGD). (a) A cadeia  $\alpha$  do colágeno tem uma estrutura secundária repetitiva que é única desta proteína. A sequência tripeptídica que se repete Gly-X-Pro ou Gly-X-4-Hyp adota uma estrutura helicoidal anti-horária com três resíduos por volta. A sequência repetida para gerar este modelo é a Gly-Pro-4-Hyp. (b) Modelo de volume atômico da mesma cadeia  $\alpha$ . (c) Três destas hélices (mostradas aqui em cinza, azul e roxo) se enrolam uma sobre as outras, no sentido horário. (d) A super-hélice de colágeno formada por três cadeias, mostrada a partir de uma das extremidades, está representada em um modelo de esfera e bastão. Os resíduos de Gly são mostrados em vermelho. A glicina, por ser pequena, é necessária para uma junção firme na região onde as três cadeias estão em contato. As esferas nesta ilustração não representam os raios de van der Waals dos átomos individuais. O centro da super-hélice de três cadeias não é oco, como aparece aqui, mas sim firmemente compacto.

muito justa de seus três polipeptídeos. A 4-hidroxiprolina tem um papel importante na estrutura do colágeno – e na história do ser humano (Quadro 4-3).



## QUADRO 4-3



## MEDICINA

## Por que marinheiros, exploradores e universitários devem comer frutas e vegetais frescos

“... por um azar, junto com a insalubridade do país, onde nunca cai uma gota de chuva, fomos acometidos pela “doença”, que era tal que toda a carne de nossos braços murchou, e a pele de nossas pernas ficou com manchas escuras, com pedaços bolorentos, como uma bota velha; e uma carne esponjosa surgiu nas gengivas daqueles que pegaram a doença, e ninguém escapou dela, indo direto para as garras da morte. O sinal era o seguinte: quando o nariz começava a sangrar, então a morte estava próxima...”

— Memórias do Lorde de Joinville, 1300 d.C.

Essa passagem descreve a situação do exército de Luís IX no fim da Sétima Cruzada (1248-1254), quando o exército, enfraquecido pelo escorbuto, foi destruído pelos egípcios. Qual a natureza desse mal que acometeu esses soldados do décimo terceiro século?

O escorbuto é causado pela falta de vitamina C, ou ácido ascórbico (ascorbato). A vitamina C é necessária para, entre outras coisas, a hidroxilação da prolina e da lisina no colágeno. O escorbuto é uma doença caracterizada pela degeneração do tecido conectivo, sendo que sua manifestação em estágio avançado inclui inúmeras pequenas hemorragias causadas por vasos sanguíneos frágeis, perda dos dentes, difícil cicatrização de feridas e reabertura de feridas antigas, dor e degeneração dos ossos, e, no final, falência cardíaca. Casos mais brandos de deficiência de vitamina C são acompanhados de fadiga, irritabilidade e aumento da gravidade das infecções do trato respiratório. A maioria dos animais sintetiza grande quantidade de vitamina C pela conversão da glicose em ascorbato em quatro etapas enzimáticas. Contudo, no curso da evolução, os humanos e alguns outros animais – gorilas, cobaias e morcegos frugívoros – perderam a última enzima dessa rota, devendo obter o ascorbato da dieta. A vitamina C é encontrada em uma grande variedade de frutas e vegetais. Até 1800, entretanto, ela estava ausente nos alimentos desidratados e outros suprimentos alimentares estocados para o inverno ou para longas viagens.

O escorbuto foi registrado pelos egípcios em 1500 a.C., e descrito em escritos de Hipócrates, que datam do quinto século a.C. Entretanto, ele não tornou-se uma notícia pública até as viagens de descobrimento europeias de 1500 a 1800. A primeira navegação ao redor do mundo (1519-1522), liderada por Ferdinand Magellan, acabou com a perda de mais de 80% de sua tripulação para o escorbuto. Durante a segunda viagem de Jacques Cartier para explorar o rio St. Lawrence (1535-1536), a tripulação foi ameaçada por um completo desastre, até que um americano nativo ensinou os homens a fazer chá de cedro (contendo vitamina C) que curava e prevenia o escorbuto. As epidemias de escorbuto nos invernos da Europa foram gradativamente sendo eliminadas no século XIX, à medida que a cultura de batata, originária da América do Sul, foi disseminada.

Em 1747, James Lind, cirurgião escocês da Marinha Real, realizou o primeiro estudo clínico controlado registrado na história. Durante uma longa viagem no navio de guerra *HMS Salisbury*, Lind selecionou 12 marinheiros com escorbuto e os dividiu em grupos de dois. Todos os 12 receberam a mesma dieta, exceto que cada um dos grupos recebeu um remédio diferente, dentre os recomendados na época, para o escorbuto. Os marinheiros que receberam limões e laranjas se recuperaram e voltaram ao trabalho. Os marinheiros que receberam suco de maçã fervida tiveram uma pequena melhora. Os outros continuaram a piorar. O *Tratado sobre escorbuto* de Lind foi publicado em 1753, mas a inércia permaneceu na Marinha real por mais 40 anos. Em 1795, o Ministério da Marinha Britânica finalmente ordenou que fosse dada uma ração com suco concentrado de lima ou limão para todos os marinheiros britânicos. O escorbuto continuou a ser um problema em algumas partes do mundo até 1932, quando o cientista húngaro Albert Szent-Györgyi, e W. A. Waugh e C. G. King da Universidade de Pittsburgh isolaram e sintetizaram o ácido ascórbico.

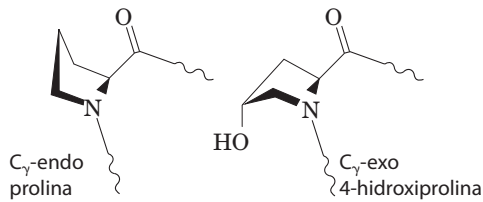
O ácido L-ascórbico (vitamina C) é um pó branco, inodoro e cristalino, totalmente solúvel em água e relativamente insolúvel em solventes orgânicos. O pó seco, longe da luz, é estável por um tempo considerável. O consumo diário apropriado dessa vitamina ainda é alvo de debates. O valor recomendado nos Estados Unidos é 90 mg (na Austrália e no Reino Unido, o recomendado é 60 mg; na Rússia, são recomendados 125 mg). Além das frutas cítricas e praticamente todas as outras frutas frescas, boas fontes de vitamina C incluem pimentas, tomates, batatas e brócolis. A vitamina C de frutas e vegetais é destruída por supercozimento ou longa estocagem.

Por que o ascorbato é tão necessário para uma boa saúde? De particular interesse é seu papel na formação de colágeno. Como dito no texto, o colágeno é formado por unidades repetidas de tripeptídeos Gly-X-Y, dos quais X e Y em geral são Pro e 4-Hyp – o derivado de prolina (4R)-L-hidroxi prolina, que tem um papel importante no entrelaçamento das fibras do colágeno e na manutenção de sua estrutura. O anel de prolina normalmente é encontrado em duas conformações, chamadas de C<sub>γ</sub>-endo e C<sub>γ</sub>-exo (Figura Q-1). A estrutura da hélice de colágeno necessita de resíduos de Pro na posição Y para estar na conformação C<sub>γ</sub>-exo, sendo essa conformação favorecida pela substituição do hidroxil no C-4 da 4-Hyp. A estrutura de colágeno também requer que o resíduo Pro na posição X tenha uma conformação C<sub>γ</sub>-endo, e a introdução



James Lind, 1716–1794; cirurgião da marinha, 1739–1748





**FIGURA Q-1** Conformação  $C_{\gamma}$ -endo da prolina e conformação  $C_{\gamma}$ -exo da 4-hidroxiprolina.

da 4-Hyp nessa posição pode desestabilizar a hélice. Na ausência de vitamina C, as células não conseguem hidroxilar a Pro da posição Y. Isso leva a uma instabilidade do colágeno e aos problemas no tecido conectivo observados no escorbuto.

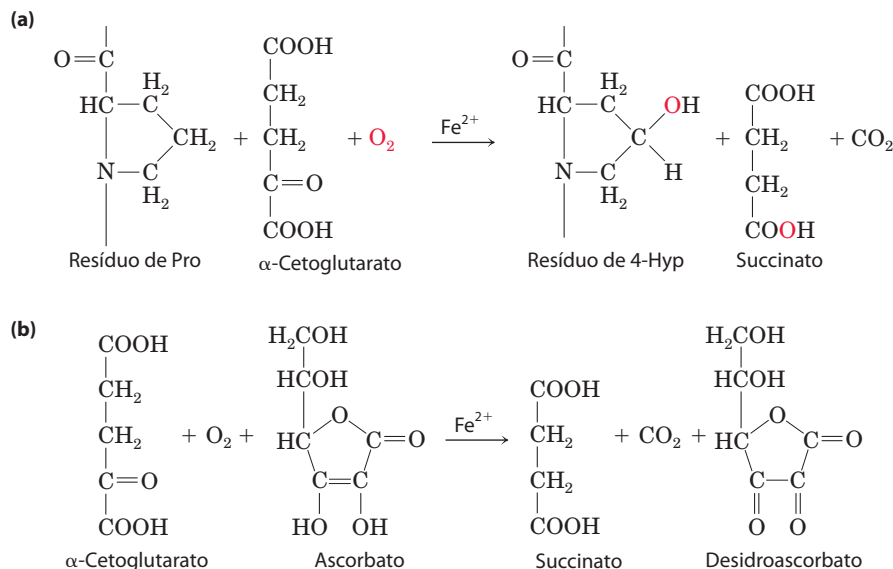
A hidroxilação de resíduos específicos de Pro no pró-colágeno, o precursor do colágeno, requer a ação da enzima prolil-4-hidroxilase. Essa enzima ( $M_r = 240.000$ ) apresenta-se como um tetrâmero  $\alpha_2\beta_2$  em todos os vertebrados. A atividade de hidroxilação da prolina está nas subunidades  $\alpha$ . Cada subunidade  $\alpha$  contém um átomo de ferro não hemínico ( $Fe^{2+}$ ), e a enzima faz parte de uma classe de hidroxilases que necessitam de  $\alpha$ -cetogluturato em suas reações.

Em uma reação normal da prolil-4-hidroxilase (Figura Q-2a), uma molécula de  $\alpha$ -cetogluturato e uma de  $O_2$

se ligam à enzima. O  $\alpha$ -cetogluturato é oxidativamente descarboxilado para formar  $CO_2$  e succinato. O átomo de oxigênio remanescente é então usado para hidroxilar o resíduo de Pro apropriado no pró-colágeno. Nenhum ascorbato é necessário nessa reação. Entretanto, a prolil-4-hidroxilase também catalisa uma descarboxilação oxidativa da prolina (Figura Q-2b). Durante essa reação, o  $Fe^{2+}$  do grupamento heme se oxida, inativando a enzima e impedindo a hidroxilação da prolina. O ascorbato consumido nessa reação é necessário para restaurar a atividade enzimática – pela redução do ferro do grupamento heme.

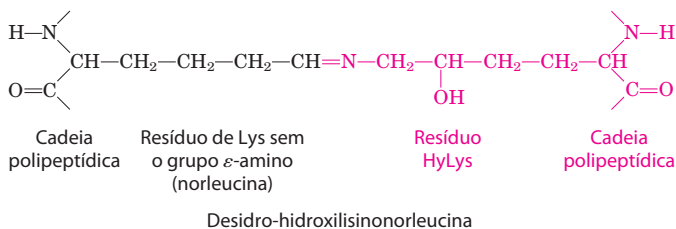
O escorbuto permanece sendo um problema ainda hoje, não somente em regiões remotas onde os alimentos nutritivos são escassos, mas, surpreendentemente, nos campi de universidades americanas. Os únicos vegetais consumidos pelos estudantes são aqueles em saladas refogadas, e os dias passam sem que esses jovens consumam frutas frescas. Um estudo de 1998, com 230 estudantes da Arizona State University, revelou que 10% deles tinham sérias deficiências de vitamina C, e dois estudantes tinham níveis de vitamina C tão baixos que provavelmente tinham escorbuto. Somente metade dos estudantes nessa pesquisa consumia a quantidade diária recomendada de vitamina C.

Então, consuma vegetais e frutas frescos.



**FIGURA Q-2** Reações catalisadas pela prolil-4-hidroxilase. (a) Reação usual, acoplada à hidroxilação da prolina, que não necessita de ascorbato. O destino dos dois átomos de oxigênio do  $O_2$  é mostrado em vermelho. (b) Reação não acoplada, na qual o  $\alpha$ -cetogluturato sofre descarboxilação oxidativa sem a hidroxilação da prolina. O ascorbato é consumido estequiometricamente neste processo, à medida que é convertido a desidroascorbato, prevenindo a oxidação do  $Fe^{2+}$ .

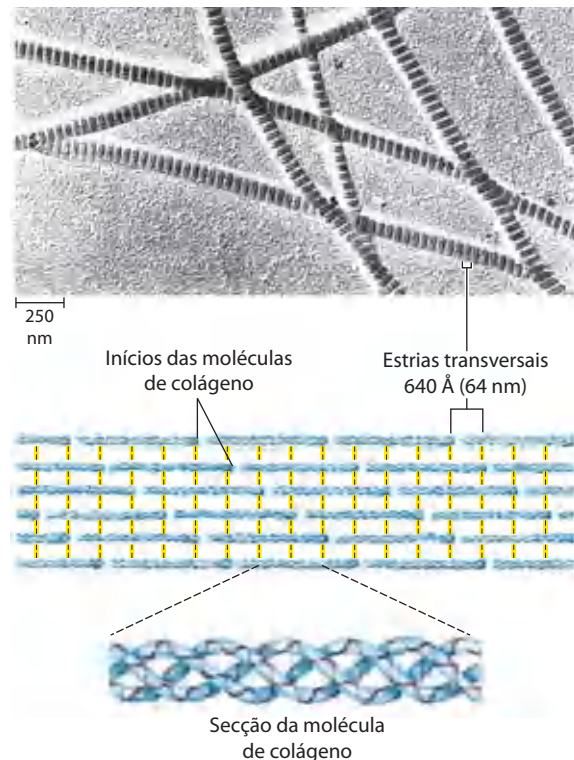
O empacotamento compacto das cadeias  $\alpha$  na hélice tripla de colágeno garante uma resistência elástica maior do que aquela de uma barra de aço de mesma seção transversal. As fibrilas de colágeno (Figura 4-13) são conjuntos de estruturas supramoleculares formados por triplas hélices de moléculas de colágeno (algumas vezes chamadas de moléculas de tropocolágeno) associadas em uma grande variedade de formas para garantir diferentes graus de resistência elástica. As cadeias  $\alpha$  e as fibrilas das moléculas de colágeno são interligadas por tipos incomuns de ligações covalentes envolvendo Lys, HyLys (5-hidroxisilisina; ver Figura 3-8a), ou resíduos His presentes em algumas posições X e Y. Essas ligações criam resíduos de aminoácidos incomuns, como o desidro-hidroxisilisonorleucina. O caráter cada vez mais rígido e quebradiço do tecido conectivo envelhecido resulta do acúmulo de ligações covalentes transversais nas fibrilas de colágeno.



Um mamífero tem mais de 30 variantes estruturais de colágeno, específicas para cada tecido e diferentes em sequência e função. Alguns defeitos genéticos na estrutura do colágeno em humanos ilustram a estreita relação entre a sequência de aminoácidos e a estrutura tridimensional nesta proteína. A osteogênese imperfeita é caracterizada pela formação anormal dos ossos em bebês; pelo menos oito variações dessa condição, em diferentes graus de severidade, ocorrem na população humana. A síndrome de Ehlers-Danlos é caracterizada por articulações soltas, e pelo menos seis variantes ocorrem em humanos. O compositor Niccolò Paganini (1782-1840) era famoso por sua destreza aparentemente impossível de tocar violino. Ele sofria de uma variante da síndrome Ehlers-Danlos que lhe rendeu efetivamente juntas duplas. Algumas variantes podem ser letais, outras causam problemas por toda vida.

Todas as variantes de ambas as condições resultam da substituição de um resíduo de aminoácido com um grupo R volumoso (como Cys ou Ser) por um resíduo de Gly de uma cadeia  $\alpha$  em uma ou outra proteína colágena (um resíduo Gly diferente é substituído em cada uma das doenças). Essas substituições de um único resíduo têm um efeito catastrófico na função do colágeno, pois interrompem a repetição de Gly-X-Y, que garante ao colágeno sua estrutura helicoidal única. Dada sua importância na hélice tripla do colágeno (Figura 4-12d), a Gly não pode ser substituída por outro resíduo de aminoácido sem um efeito substancialmente deletério na estrutura do colágeno. ■

**Fibroína da seda** A fibroína, a proteína da seda, é produzida por insetos e aranhas. Suas cadeias polipeptídicas estão predominantemente na conformação  $\beta$ . A fibroína é rica em resíduos Ala e Gly, permitindo um grande empacotamento das folhas  $\beta$  e um arranjo entrelaçado dos grupos R (Figura

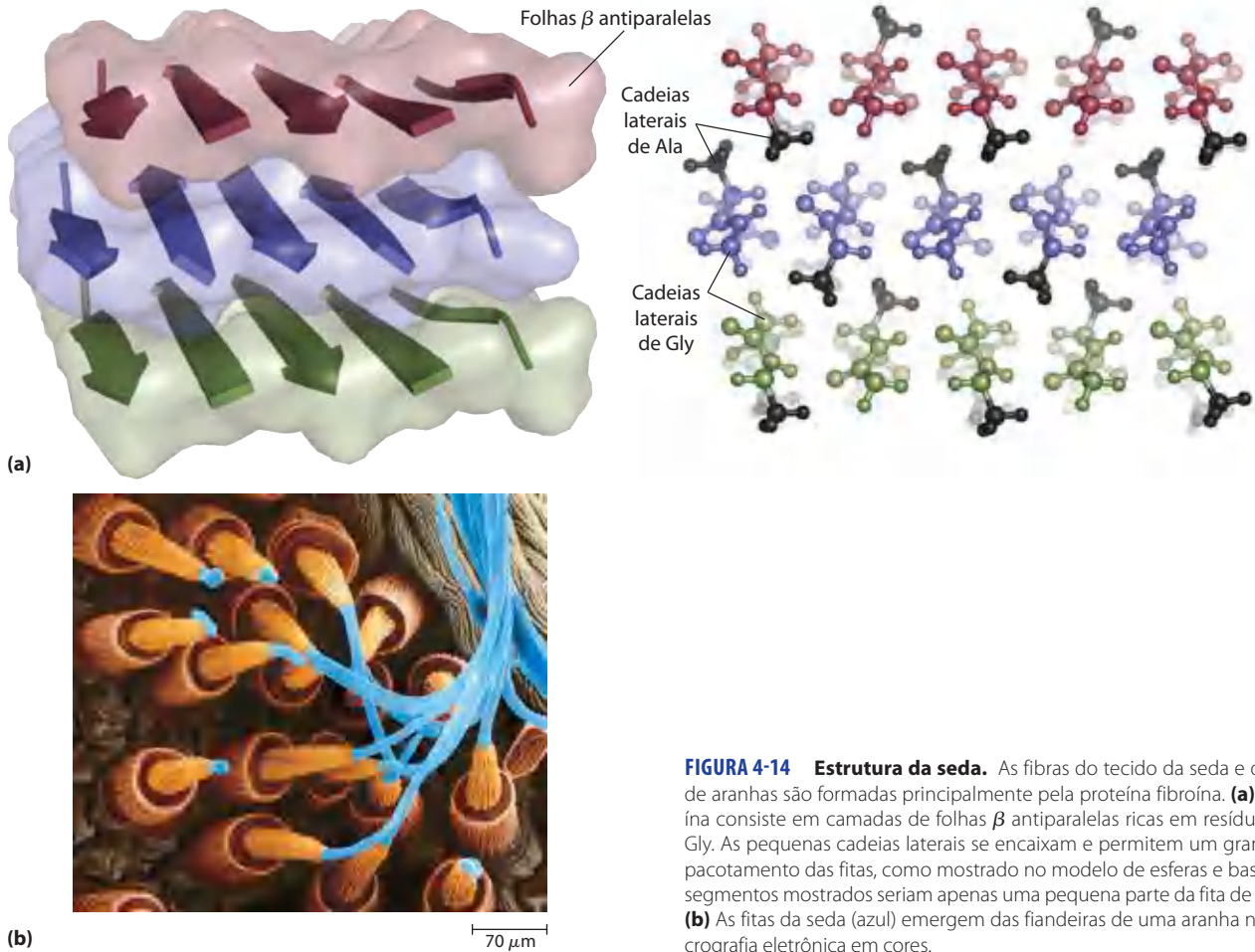


**FIGURA 4-13** Estrutura das fibrilas de colágeno. O colágeno ( $M_r$  300.000) é uma molécula em forma de bastão, com cerca de 3.000 Å de comprimento e apenas 15 Å de largura. Suas três cadeias  $\alpha$  helicoidais entrelaçadas podem ter sequências diferentes; cada cadeia tem aproximadamente 1.000 resíduos de aminoácidos. As fibrilas são feitas de moléculas de colágeno alinhadas de forma escalonada e com ligações cruzadas que garantem resistência. O alinhamento específico e o grau de ligações transversais variam com o tecido e formam estrias transversais características, vistas por microscopia eletrônica. No exemplo mostrado aqui, o alinhamento dos grupos iniciais de cada quatro moléculas produz estrias distantes 640 Å (64 nm) umas das outras.

ra 4-14). A estrutura global é estabilizada por extensivas ligações de hidrogênio entre todas as ligações peptídicas dos polipeptídeos de cada cadeia  $\beta$ , assim como pela otimização das interações de van der Waals entre as cadeias. A seda não é elástica porque sua conformação  $\beta$  já está bem estendida (Figura 4-6). Entretanto, a estrutura é flexível, pois as cadeias estão unidas por inúmeras interações fracas, em vez de ligações covalentes como ligações dissulfeto nas  $\alpha$ -queratinas.

### A diversidade estrutural reflete a diversidade funcional nas proteínas globulares

Em uma proteína globular, segmentos diferentes das cadeias polipeptídicas (ou de múltiplas cadeias polipeptídicas) se dobras uns sobre os outros, gerando uma forma mais compacta do que a observada para as proteínas fibrosas (Figura 4-15). O enovelamento também garante a diversidade estrutural necessária às proteínas para realizar um grande leque de funções biológicas. Proteínas globulares incluem enzimas, proteínas transportadoras, proteínas motoras, proteínas reguladoras, imunoglobulinas e proteínas com muitas outras funções.



**FIGURA 4-14 Estrutura da seda.** As fibras do tecido da seda e das teias de aranhas são formadas principalmente pela proteína fibroína. **(a)** A fibroína consiste em camadas de folhas  $\beta$  antiparalelas ricas em resíduos Ala e Gly. As pequenas cadeias laterais se encaixam e permitem um grande empacotamento das fitas, como mostrado no modelo de esferas e bastões. Os segmentos mostrados seriam apenas uma pequena parte da fita de fibroína. **(b)** As fitas da seda (azul) emergem das fiandeiras de uma aranha nesta micrografia eletrônica em cores.

A presente discussão sobre proteínas globulares começa com os princípios vislumbrados a partir das primeiras estruturas de proteínas elucidadas. Segue-se uma detalhada descrição e classificação comparativa de subestruturas proteicas. Tais discussões são possíveis somente devido à grande quantidade de informações disponíveis na Internet, a partir de bancos de dados públicos, particularmente do Banco de Dados de Proteínas (Quadro 4-4).

Conformação  $\beta$   
 $2.000 \times 5 \text{ \AA}$

Hélice  $\alpha$   
 $900 \times 11 \text{ \AA}$

Forma globular nativa  
 $100 \times 60 \text{ \AA}$

**FIGURA 4-15 As estruturas proteicas globulares são compactas e variadas.** A albumina sérica humana ( $M_r$ , 64.500) tem 585 resíduos em uma única cadeia. São dadas aqui as dimensões aproximadas de uma de suas cadeias polipeptídicas se ela ocorresse em uma conformação  $\beta$ , ou em uma hélice  $\alpha$ . Também é mostrado o tamanho da proteína em sua forma globular nativa, determinada por cristalografia por raios X; a cadeia polipeptídica deve estar com um enovelamento muito compacto para caber nestas dimensões.

### A mioglobina forneceu os indícios iniciais sobre a complexidade da estrutura globular proteica

**Arquitetura proteica – Estrutura terciária de proteínas globulares pequenas, II. Mioglobina** O primeiro avanço no entendimento da estrutura tridimensional de uma proteína globular veio com os estudos de difração por raios X da mioglobina, feitos por John Kendrew e seus colaboradores, em 1950. A mioglobina é uma proteína ligadora de oxigênio relativamente pequena ( $M_r$ , 16.700) das células musculares. Ela funciona tanto para a estocagem de oxigênio quanto para facilitar a difusão do oxigênio nos tecidos musculares em contração. A mioglobina contém uma única cadeia polipeptídica de 153 resíduos de aminoácidos de sequência conhecida e um único grupo ferro-protoporfirina, ou heme. O mesmo grupo heme encontrado na mioglobina é encontrado na hemoglobina, a proteína ligadora de oxigênio dos eritrócitos, sendo responsável pela coloração vermelha amarronzada tanto da mioglobina quanto da hemoglobina. A mioglobina é particularmente abundante nos músculos de mamíferos mergulhadores, como as baleias, as focas e os botos – tão abundante que os músculos destes animais são marrons. A estocagem e a distribuição do oxigênio pela mioglobina do músculo permitem que animais que mergulham permaneçam submersos por um longo período. As



**QUADRO 4-4 O banco de dados de proteínas**

Atualmente, o número de estruturas proteicas tridimensionais conhecidas está na casa das dezenas de milhares e mais do que dobra a cada dois anos. Essa abundância de informações está revolucionando o entendimento sobre a estrutura de proteínas, a relação estrutura-atividade e as rotas evolutivas pelas quais as proteínas chegaram ao seu estado atual, que podem ser vistas nas semelhanças que aparecem entre as famílias à medida que os bancos de dados de proteínas são examinados e classificados. Um dos recursos mais importantes disponíveis aos bioquímicos é o **Protein Data Bank (PDB)**, ou banco de dados de proteínas; [www.pdb.org](http://www.pdb.org)).

O PDB é um arquivo de estruturas tridimensionais de macromoléculas biológicas determinadas experimentalmente, contendo quase todas as estruturas macromoleculares (proteínas, RNA, DNA, etc.) elucidadas até o momento. A cada estrutura é atribuído um código de identificação (um código de quatro caracteres, chama-

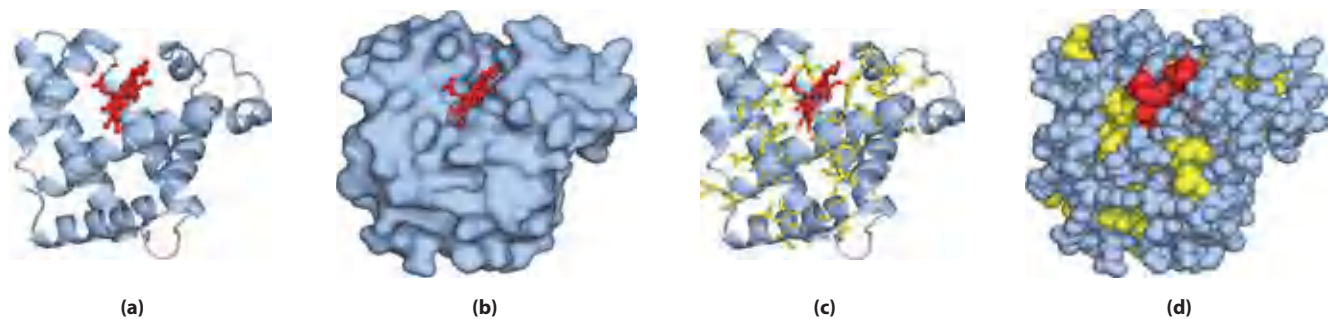
do de PDB ID). Tais identificadores são fornecidos nas legendas das figuras para cada uma das estruturas derivadas do PDB ilustradas nesse texto, de forma que os estudantes e os professores podem explorar essas mesmas estruturas. Os arquivos de dados no PDB descrevem as coordenadas espaciais de cada átomo cuja posição foi determinada (muitas das estruturas catalogadas não estão completas). Arquivos de dados adicionais fornecem informações de como as estruturas foram determinadas e sua precisão. As coordenadas atômicas podem ser convertidas em uma imagem da macromolécula com a ajuda de programas de visualização de estruturas. Os estudantes são incentivados a acessar o PDB e explorar as estruturas usando programas de visualização referenciados no banco de dados. Os arquivos de estruturas macromoleculares também podem ser baixados e explorados em seu computador, usando programas livres, tal como o Jmol.

atividades da mioglobina e de outras moléculas de globina são investigadas em detalhe no Capítulo 5.

A **Figura 4-16** mostra diversas representações estruturais da mioglobina, ilustrando como a cadeia polipeptídica é dobrada nas três dimensões – sua estrutura terciária. O grupo vermelho envolvido pela proteína é o heme. O esqueleto da molécula de mioglobina consiste em oito segmentos de hélices  $\alpha$  relativamente retas, interrompidas por dobras, algumas delas sendo voltas  $\beta$ . A hélice  $\alpha$  mais longa tem 23 resíduos de aminoácidos, a mais curta apenas 7. Todas as hélices são voltadas para a direita. Mais de 70% dos resíduos de mioglobina estão nessas regiões  $\alpha$ -helicoidais. As análises por raios X revelaram a posição precisa de cada um dos grupos R, que preenchem quase todo o espaço dentro da cadeia dobrada, não ocupada por átomos do esqueleto.

Várias conclusões importantes podem ser tiradas a partir da estrutura da mioglobina. As posições das cadeias late-

rais dos aminoácidos refletem uma estrutura cuja estabilidade é resultante de suas interações hidrofóbicas. A maioria dos grupos R hidrofóbicos está no interior da molécula, escondidos da exposição à água. Exceto dois resíduos, todos os outros grupos R polares estão na superfície externa da molécula e estão hidratados. A molécula de mioglobina é tão compacta que em seu interior só há lugar para quatro moléculas de água. Esse denso núcleo hidrofóbico é típico das proteínas globulares. A fração de espaço ocupado pelos átomos em um líquido orgânico é de 0,4 a 0,6. Em uma proteína globular, essa fração é de cerca de 0,75, comparável com aquela do cristal (em um cristal típico, a fração é de 0,70 a 0,78, próximo do máximo teórico). Nesse ambiente compacto, as interações fracas são fortalecidas e reforçam umas às outras. Por exemplo, as cadeias apolares no núcleo estão tão próximas que as interações de curto alcance do tipo van der Waals têm uma contribuição significativa na estabilização das interações hidrofóbicas.



**FIGURA 4-16 Estrutura terciária da mioglobina de cachalote.** (PDB ID 1MBO) A orientação da proteína é igual de (a) a (d); o grupo heme é mostrado em vermelho. Além de ilustrar a estrutura da mioglobina, esta figura mostra exemplos de diversas formas de apresentar a estrutura de uma proteína. **(a)** Esqueleto polipeptídico na representação de fita, introduzido por Jane Richardson, destacando as estruturas secundárias. As regiões de hélice  $\alpha$  são evidentes. **(b)** Imagem da superfície da proteína; ela é útil na visualização de

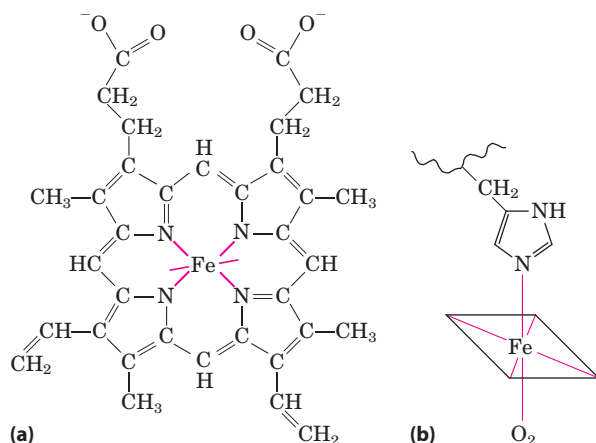
fechas, onde outras moléculas podem se ligar à proteína. **(c)** Representação da estrutura em fita, incluindo as cadeias laterais (em amarelo) dos resíduos hidrofóbicos Leu, Ile, Val e Phe. **(d)** Modelo de volume atômico com todas as cadeias laterais dos aminoácidos. Cada átomo está representado por uma esfera que cobre o seu raio de van der Waals. Os resíduos hidrofóbicos são novamente mostrados em amarelo; a maioria está no interior da proteína, e por isso não está visível.



A dedução da estrutura da mioglobina confirmou algumas expectativas e introduziu alguns elementos novos à estrutura secundária. Como predito por Pauling e Corey, todas as ligações peptídicas estão em configuração planar *trans*. As hélices  $\alpha$  da mioglobina forneceram a primeira evidência experimental direta da existência desse tipo de estrutura secundária. Três dos quatro resíduos Pro são encontrados em dobras. O quarto resíduo Pro ocorre em uma hélice  $\alpha$ , onde ele cria uma torção necessária para o firme empacotamento da hélice.

O grupo heme planar repousa sobre uma fenda na molécula de mioglobina. O átomo de ferro no centro do grupo heme apresenta duas posições de ligação (coordenação) perpendiculares ao plano do heme (Figura 4-17). Uma delas se liga ao grupo R de um resíduo de His na posição 93; a outra corresponde ao local onde a molécula de  $O_2$  se liga. Dentro dessa fenda, a acessibilidade do grupo heme ao solvente é bem restrita. Isso é importante para a função, pois grupos heme livres em uma solução oxigenada são rapidamente oxidados do estado ferroso ( $Fe^{2+}$ ), que é ativo na ligação reversível com o  $O_2$ , para o estado férrico ( $Fe^{3+}$ ), que não se liga ao  $O_2$ .

À medida que as diferentes estruturas da mioglobina foram resolvidas, os pesquisadores foram capazes de observar as mudanças estruturais que acompanham a ligação do oxigênio ou de outra molécula, e, assim, pela primeira vez, de entender a correlação entre a estrutura de uma proteína e sua função. Centenas de proteínas são agora submetidas a análises semelhantes. Hoje, a espectroscopia de ressonância magnética nuclear (RMN) e outras técnicas suplementam os dados de difração de raios X, fornecendo mais informações sobre as estruturas das proteínas (Quadro 4-5). Adicionalmente, o sequenciamento de DNA genômico de muitos organismos (Capítulo 9) identificou milhares de genes que codificam proteínas de sequências conhecidas, mas ainda de função desconhecida; esse trabalho continua a passos largos.



**FIGURA 4-17 O grupo heme.** Este grupo está presente na mioglobina, hemoglobina, citocromos e muitas outras proteínas (as proteínas heme). **(a)** O heme consiste em uma estrutura de anel orgânico complexo, a protoporfirina, que liga um átomo de ferro no seu estado ferroso ( $Fe^{2+}$ ). O átomo de ferro tem seis ligações coordenadas, quatro delas no plano da molécula da porfirina (e ligadas a ela) e duas perpendiculares a ela. **(b)** Na mioglobina e na hemoglobina, uma das ligações perpendiculares está ligada a um átomo de nitrogênio de um resíduo His. A outra está “vaga”, servindo de sítio de ligação para a molécula de  $O_2$ .

## As proteínas globulares têm uma diversidade de estruturas terciárias

A partir do que se sabe sobre estruturas terciárias de centenas de proteínas globulares, é claro que a mioglobina ilustra apenas uma das diversas formas em que a cadeia polipeptídica pode se dobrar. A Tabela 4-4 mostra as proporções de hélices  $\alpha$  e conformações  $\beta$  (expressas em porcentagem de resíduos em cada um dos tipos) em diversas proteínas globulares pequenas e de cadeia única. Cada uma dessas proteínas tem uma estrutura diferente, adaptada à sua função biológica particular, mas como um todo elas compartilham diversas propriedades importantes com a mioglobina. Cada uma delas está dobrada de forma compacta e, em cada caso, as cadeias laterais hidrofóbicas dos aminoácidos estão orientadas na direção do centro da proteína (longe da água), e suas cadeias laterais hidrofílicas se encontram na superfície. As estruturas também são estabilizadas por inúmeras ligações de hidrogênio e algumas interações iônicas.

Para os estudantes iniciantes, as estruturas terciárias muito complexas das proteínas globulares – algumas muito maiores do que a mioglobina – são mais facilmente acessadas concentrando-se nos padrões estruturais comuns, recorrentes em diferentes proteínas, geralmente não relacionadas. A estrutura tridimensional de uma proteína globular típica pode ser considerada um conjunto de segmentos polipeptídicos em hélices  $\alpha$  e conformações  $\beta$ , ligados por segmentos conectores. A estrutura pode ser, então, definida pela forma como os segmentos estão empilhados uns sobre os outros, e como estão arranjados os segmentos que os conectam.

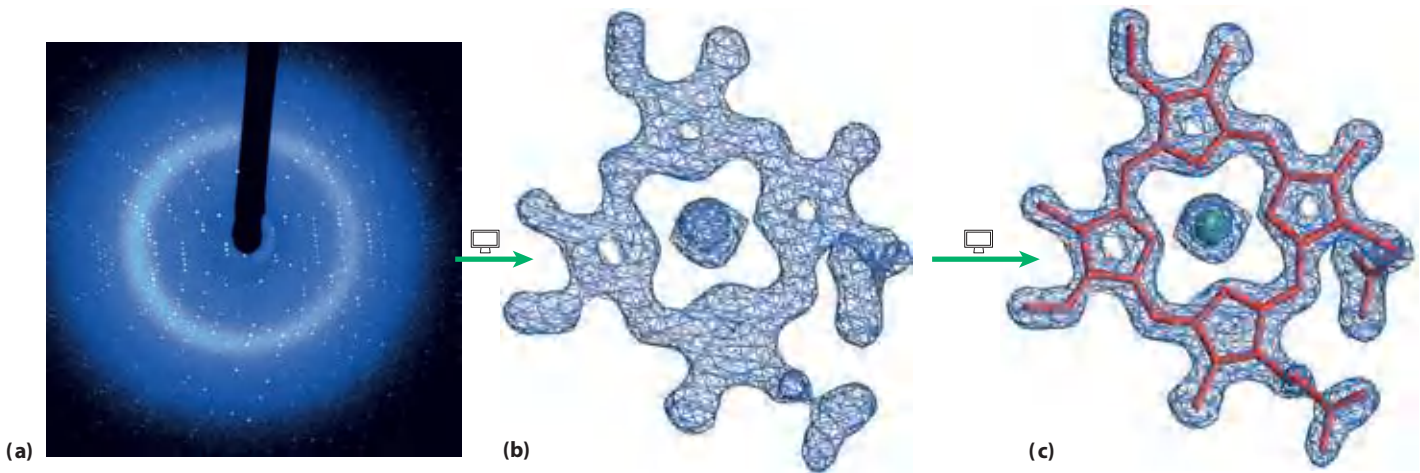
Para entender uma estrutura tridimensional completa, é preciso analisar seu padrão de enovelamento. Primeiro, definem-se dois importantes termos que descrevem os padrões de estruturas de proteínas ou elementos de uma cadeia polipeptídica, seguindo então para as regras de enovelamento.

**TABELA 4-4** Proporção aproximada de hélices  $\alpha$  e conformações  $\beta$  em algumas proteínas de cadeia única

Proteína (número total de resíduos)	Resíduos (%)*	
	Hélice $\alpha$	Conformação $\beta$
Quimotripsina (247)	14	45
Ribonuclease (124)	26	35
Carboxipeptidase (307)	38	17
Citocromo <i>c</i> (104)	39	0
Lisozima (129)	40	12
Mioglobina (153)	78	0

**Fonte:** Dados de Cantor, C.R. e Schimmel, P.R. (1980) *Biophysical Chemistry*, Part I: *The Conformation of Biological Macromolecules*, p. 100, W.H. Freeman and Company, New York.

\*As proporções das cadeias polipeptídicas que não fazem parte de hélices  $\alpha$  e conformações  $\beta$  consistem em curvaturas e espirais irregulares ou regiões estendidas. Segmentos de hélice  $\alpha$  e conformação  $\beta$  algumas vezes diferem um pouco de suas dimensões e geometrias usuais.

**QUADRO 4-5 MÉTODOS** Métodos de determinação da estrutura tridimensional de uma proteína**Difração de raios X**

O espaçamento dos átomos em um retículo cristalino pode ser determinado pela medida da localização e da intensidade dos pontos produzidos por um feixe de raios X, de um dado comprimento de onda, em um filme fotográfico, depois desse feixe ser difratado pelos elétrons dos átomos. Por exemplo, a análise por raios X de cristais de cloreto de sódio mostra que os íons  $\text{Na}^+$  e  $\text{Cl}^-$  estão dispostos em um arranjo cúbico simples. O espaçamento de diferentes tipos de átomos em moléculas orgânicas complexas, até mesmo as grandes como as proteínas, também pode ser analisado por métodos de difração de raios X. Entretanto, essa técnica de análise de cristais de moléculas complexas é muito mais trabalhosa do que a de cristais de sais simples. Quando o padrão de repetição do cristal é uma molécula tão grande quanto uma proteína, o número de átomos da molécula resulta em milhares de pontos de difração que precisam ser analisados por computador.

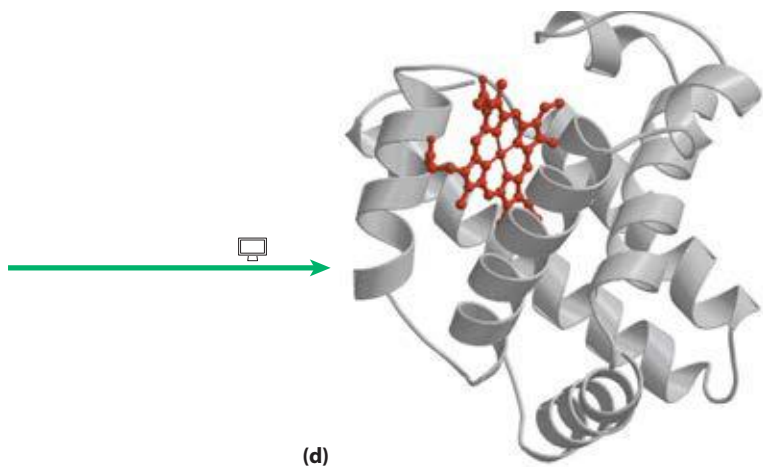
Considere como as imagens são geradas em um microscópio óptico. A luz de uma fonte pontual é focalizada em um objeto. O objeto espalha as ondas de luz, e estas ondas espalhadas são reagrupadas por uma série de lentes para gerar uma imagem aumentada do objeto. O tamanho mínimo do objeto cuja estrutura pode ser determinada desta forma – isto é, o poder de resolução do microscópio – é determinado pelo comprimento de onda de luz, neste caso, luz visível, com um comprimento de onda entre 400 e 700 nm. Objetos menores do que a metade do comprimento de onda da luz incidente não têm resolução. Para resolver objetos tão pequenos quanto as proteínas, é necessário usar raios X, com comprimentos de onda na faixa de 0,7 a 1,5 Å (0,07 a 0,15 nm). Entretanto, não há lentes que possam reagrupar os raios X para formar a imagem. Em vez disso, o padrão de difração dos raios X é coletado diretamente, e a imagem é reconstruída por técnicas matemáticas.

A quantidade de informação obtida em uma cristalografia por raios X depende do grau de organização estrutural da amostra.

Alguns parâmetros importantes foram obtidos nos primeiros estudos de padrão de difração de proteínas fibrosas arranjadas de forma regular no cabelo e na lã. Entretanto, os feixes ordenados formados pelas proteínas fibrosas não são cristais – as moléculas estão alinhadas lado a lado, mas não estão todas orientadas na mesma direção. Informações estruturais tridimensionais mais detalhadas sobre proteínas necessitam de um cristal de proteína altamente ordenado. As estruturas de muitas proteínas ainda permanecem desconhecidas, simplesmente porque se mostraram difíceis de cristalizar. Os profissionais comparam a preparação de cristais com manter unida uma pilha de bolas de boliche com uma fita de celofane.

Operacionalmente, existem vários passos em uma análise estrutural por raios X (Figura Q-1). Um cristal é colocado em um feixe de raios X, entre a fonte de raios X e o detector, e um conjunto de pontos, chamados de reflexões, é gerado. Os pontos são gerados pela difração do feixe de raios X, e cada átomo da molécula faz uma contribuição para cada ponto. Um mapa de densidade eletrônica da proteína é reconstruído a partir do padrão total de difração, por uma técnica matemática chamada de transformada de Fourier. O computador age, na verdade, como “lente computacional”. Um modelo da estrutura é então construído de acordo com o mapa de densidade eletrônica.

John Kendrew verificou que o padrão de difração de raios X da mioglobina cristalina (isolada de músculos do cachalote) é muito complexo, com cerca de 25.000 reflexões. A análise computacional dessas reflexões ocorreu em etapas. A resolução aumentou a cada etapa, até que em 1959 as posições de quase todos os átomos diferentes de hidrogênio foram determinadas. A sequência de aminoácidos da proteína, obtida por análise química, foi consistente com o modelo molecular. A estrutura de milhares de proteínas, a maioria delas mais complexa que a mioglobina, foi desde então determinada com um nível de resolução semelhante.



É claro que o ambiente físico em um cristal não é idêntico ao ambiente em solução ou em uma célula viva. O cristal impõe uma média temporal e espacial na estrutura deduzida de sua análise, e os estudos de difração de raios X fornecem poucas informações sobre os movimentos moleculares no interior da proteína. A conformação das proteínas em um cristal pode, a princípio, ser afetada por fatores não fisiológicos como contatos proteína-proteína dentro do cristal. Entretanto, quando estruturas derivadas da análise de cristais são comparadas com informações estruturais obtidas de outras formas (como RMN, descrita a seguir), as estruturas derivadas do cristal quase sempre representam a conformação funcional da proteína. A cristalografia por raios X pode ser aplicada com sucesso a proteínas grandes demais para serem estruturalmente analisadas por RMN.

### Ressonância magnética nuclear

Uma vantagem dos estudos de ressonância magnética nuclear (RMN) é que eles são realizados com as macromoléculas em solução, enquanto a cristalografia é limitada a moléculas que podem ser cristalizadas. A RMN também pode esclarecer o lado dinâmico da estrutura da proteína, incluindo mudanças conformacionais, enovelamento proteico e interação com outras moléculas.

A RMN é a manifestação do momento angular do *spin* nuclear, uma propriedade quântico-mecânica dos núcleos atômicos. Somente alguns átomos, incluindo  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{19}\text{F}$  e  $^{31}\text{P}$ , apresentam o tipo de *spin* nuclear que gera sinal na RMN. O *spin* nuclear gera um dipolo magnético. Quando um campo magnético estático e forte é aplicado em uma solução contendo um único tipo de macromolécula, os dipolos magnéticos são alinhados no campo em uma das duas orientações, paralela (baixa energia) ou antiparalela (alta energia). Um pequeno ( $\sim 10 \mu\text{s}$ ) pulso de energia eletromagnética de frequência adequada (a frequência de ressonância, que está dentro da faixa de radiofrequência) é aplicado, em certos ângulos, aos núcleos alinhados no campo magnético.

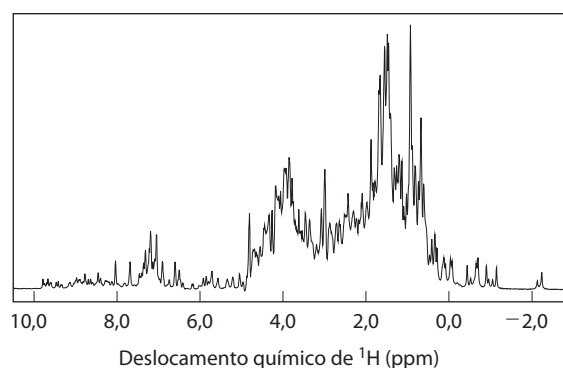
**FIGURA Q-1** Etapas da determinação da estrutura da mioglobina de cachalote via cristalografia por raios X. **(a)** O padrão da difração por raios X é gerado a partir de um cristal da proteína. **(b)** Os dados obtidos pelo padrão de difração são utilizados para calcular um mapa de densidade eletrônica tridimensional. Somente a densidade eletrônica do heme, uma pequena parte da estrutura, é mostrada aqui. **(c)** As regiões de maior densidade eletrônica revelam a localização do núcleo atômico, esta informação é utilizada para construir a estrutura final. Aqui, a estrutura do heme é modelada em seu mapa de densidade eletrônica. **(d)** A estrutura completa da mioglobina de cachalote, incluindo o heme (PDB ID 2MBW).

Alguma energia é absorvida à medida que o núcleo muda para o estado de maior energia, e o espectro de absorção resultante contém informações sobre a identidade do núcleo e seu ambiente químico nas imediações. Os dados de diversos desses experimentos com a amostra são reunidos, aumentando a relação sinal-ruído, e é gerado um espectro de RMN, como o da Figura Q-2.

O  $^1\text{H}$  é particularmente importante nos experimentos de RMN, pois é altamente sensível e naturalmente abundante. Para macromoléculas, o espectro de RMN de  $^1\text{H}$  pode ser bem complicado. Até mesmo uma pequena proteína pode conter centenas de átomos de  $^1\text{H}$ , normalmente gerando um espectro de RMN unidimensional muito complexo para ser analisado. A análise estrutural de proteínas se tornou possível com o advento das técnicas de RMN bidimensionais (Figura Q-3). Esses métodos permitem a medida do acoplamento, dependente da distância, dos *spins* nucleares de átomos espacialmente próximos (o efeito Overhauser nuclear [NOE], em um método chamado de NOESY) ou o acoplamento dos *spins* nucleares de átomos conectados por ligações covalentes (espectroscopia de correlação total, ou TOCSY).

A tradução de um espectro de RMN bidimensional em uma estrutura tridimensional completa é um processo muito trabalhoso. Os sinais de NOE fornecem algumas informações sobre as distâncias entre cada um dos áto-

(Continua na próxima página)



**FIGURA Q-2** Espectro de RMN unidimensional da globina do verme-de-sangue marinho (*Glycera* sp.). Esta proteína e a mioglobina de cachalote são análogos estruturais muito próximos, pertencendo à mesma família estrutural e compartilhando a função de transporte de oxigênio.



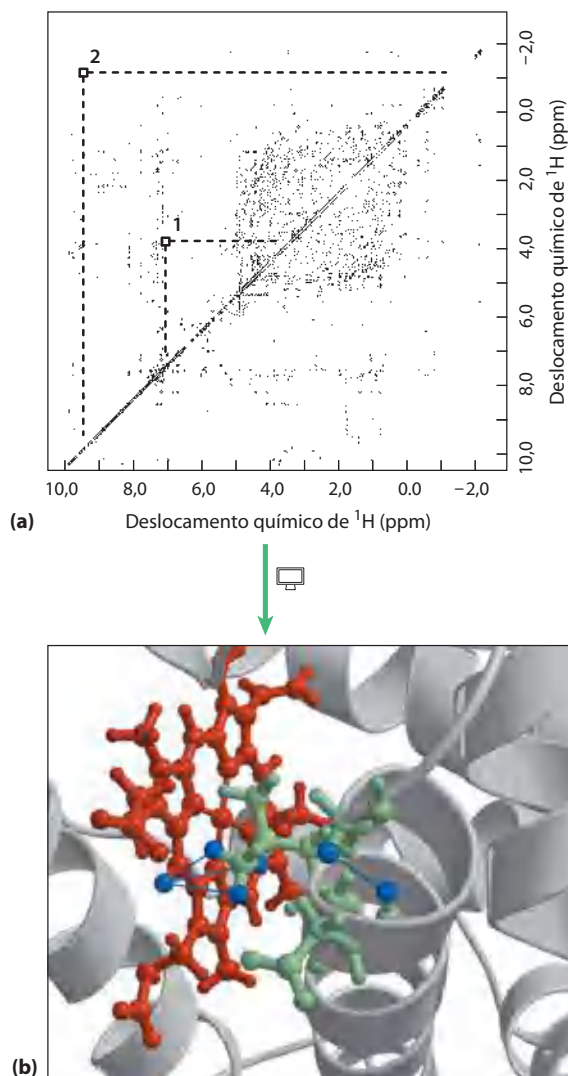
**QUADRO 4-5 MÉTODOS** Métodos de determinação da estrutura tridimensional de uma proteína (*Continuação*)

mos, mas para esses valores de distâncias serem úteis, os átomos que originam cada sinal devem ser identificados. Os experimentos complementares de TOCSY podem ajudar na identificação de quais sinais de NOE refletem átomos que estão ligados por ligações covalentes. Certos padrões de sinais de NOE têm sido associados com estruturas secundárias como hélices  $\alpha$ . A engenharia genética (Capítulo 9) pode ser usada para preparar proteínas que contenham isótopos raros como  $^{13}\text{C}$  ou  $^{15}\text{N}$ . Os novos sinais de RMN produzidos por esses átomos e o acoplamento com os sinais de  $^1\text{H}$  resultantes dessa substituição auxiliam na designação de cada um dos sinais de  $^1\text{H}$  NOE. O processo também é facilitado pelo conhecimento da sequência do polipeptídeo.

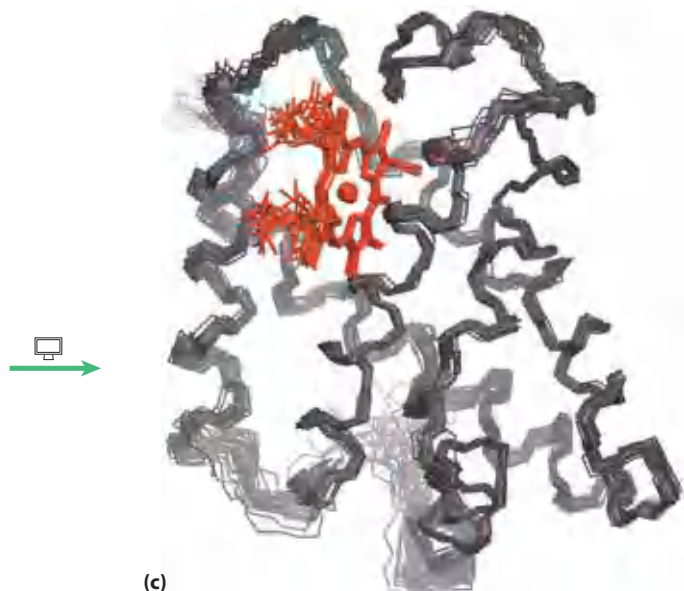
Para gerar a representação de uma estrutura tridimensional, os pesquisadores colocam em um computador as restrições de distância, junto com as restrições geométricas conhecidas como quiralidade, raios de van

der Waals e distância e ângulo de ligação. O computador gera, então, um conjunto de estruturas altamente relacionadas que representa um conjunto de conformações consistentes com os valores observados no NOE (Figura Q-3c). A incerteza das estruturas geradas por RMN é em parte um reflexo das vibrações moleculares dentro de uma proteína, discutidas em mais detalhes no Capítulo 5. O erro experimental normal também influi.

As estruturas proteicas determinadas tanto por cristalografia por raios X quanto por RMN geralmente são bem coincidentes. Em alguns casos, a localização exata de algumas cadeias laterais de aminoácidos específicos da superfície da proteína é diferente, normalmente devido aos efeitos relacionados com o empacotamento entre proteínas adjacentes no cristal. Juntas, as duas técnicas são as responsáveis pelo rápido aumento da disponibilidade de informações estruturais sobre as macromoléculas de células vivas.



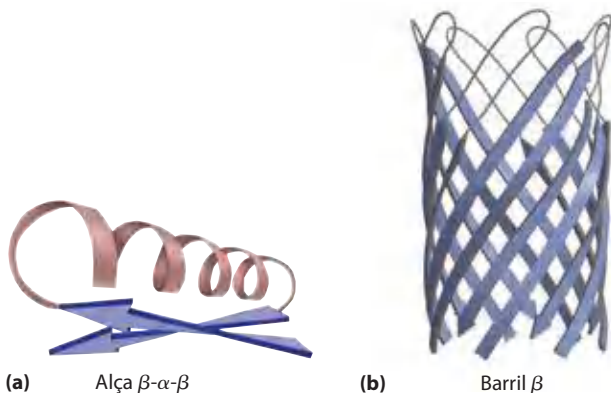
**FIGURA Q-3** Uso de RMN bidimensional para determinar a estrutura tridimensional da globina, a mesma proteína usada para gerar os dados da Figura Q-2. A diagonal em um espectro de RMN bidimensional é equivalente ao espectro unidimensional. Os picos fora da diagonal são sinais de NOE gerados pelas interações de curto alcance dos átomos de  $^1\text{H}$ , que podem gerar sinais bem distantes em um espectro unidimensional. Duas destas interações são identificadas em (a) e suas identidades estão mostradas em linhas azuis em (b) (PDB ID 1VRF). São desenhadas três linhas para a interação 2 entre um grupo metil da proteína e um hidrogênio do heme. O grupo metil gira rapidamente de forma que cada um dos seus três hidrogênios contribui igualmente para a interação e para o sinal de RMN. Tal informação é usada para determinar a estrutura tridimensional completa (PDB ID 1VRE), como em (c). As múltiplas linhas que mostram o esqueleto da proteína em (c) representam o conjunto de estruturas consistentes com os dados de distância da RMN. A semelhança estrutural com a mioglobina (Figura Q-1) é evidente. As proteínas estão orientadas da mesma forma nas duas figuras.





O primeiro termo é **motivo**, também chamado de **estrutura supersecundária** ou **enovelamento**. Um motivo ou enovelamento é um padrão de enovelamento identificável, envolvendo dois ou mais elementos da estrutura secundária e a conexão (ou conexões) entre eles. Um motivo pode ser muito simples, tal como dois elementos de estrutura secundária dobrados um sobre o outro e representa apenas uma pequena parte de uma proteína. Um exemplo é uma **alça  $\beta$ - $\alpha$ - $\beta$**  (Figura 4-18a). Um motivo também pode ter uma estrutura bem elaborada, envolvendo um grande número de segmentos proteicos dobrados juntos, como o barril  $\beta$  (Figura 4-18b). Em alguns casos, um único e grande motivo pode abranger toda a proteína. Os termos “motivo” e “enovelamento” são frequentemente usados como sinônimos, embora “enovelamento” seja mais comumente aplicado aos padrões de enovelamento ligeiramente mais complexos. Os termos incluem qualquer padrão favorável de enovelamento, sendo útil na descrição desses padrões. O segmento definido como um motivo pode, ou não, ser estável independentemente. Já foi mencionado um motivo bem estudado, a espiral enrolada da estrutura da  $\alpha$ -queratina, também encontrado em outras proteínas. O arranjo característico de oito hélices  $\alpha$  na mioglobina é reproduzido em todas as globinas e é chamado de enovelamento de globina. Observe que um motivo não é um elemento estrutural hierárquico que fica entre a estrutura secundária e a terciária. Ele é simplesmente um padrão de enovelamento. O termo sinônimo “estrutura supersecundária” pode, portanto, causar confusão, pois sugere uma hierarquia.

O segundo termo que descreve padrões estruturais é o **domínio**. Um domínio, como definido por Jane Richardson em 1981, é uma parte da cadeia polipeptídica que é independentemente estável ou pode se movimentar como uma entidade isolada em relação ao resto da proteína. Polipeptídeos com mais de algumas centenas de resíduos de aminoácidos normalmente se dobram em dois ou mais domínios, algumas vezes com diferentes funções. Muitas vezes, um domínio de uma proteína grande irá conservar sua estrutura tridimensional nativa mesmo quando separado (p. ex., por uma clivagem proteolítica) do resto da cadeia polipeptídica. Em uma proteína com múltiplos domínios, cada um



**FIGURA 4-18** Motivos estruturais. (a) Um motivo simples, a alça  $\beta$ - $\alpha$ - $\beta$ . (b) Um motivo mais elaborado, o barril  $\beta$ . Este barril  $\beta$  é o único domínio da  $\alpha$ -hemolisina (toxina que mata a célula, criando um orifício em sua membrana) da bactéria *Staphylococcus aureus* (derivada do PDB ID 7AHL).

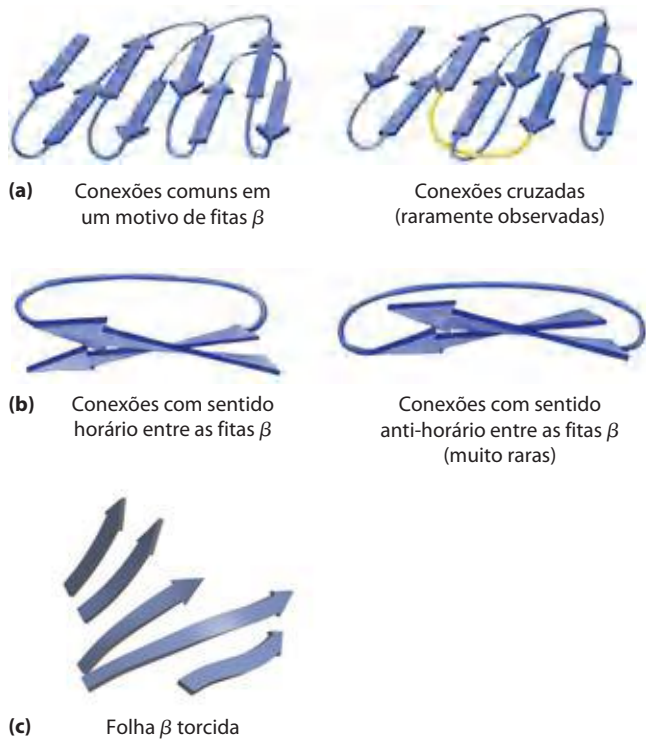


**FIGURA 4-19** Domínios estruturais do polipeptídeo troponina C. (PDB ID 4TNC) Esta proteína ligadora de cálcio encontrada no músculo tem dois domínios ligadores de cálcio separados, mostrados aqui em marrom e em azul.

deles pode aparecer como uma porção globular diferente (Figura 4-19); mais frequentemente, extensos contatos entre os domínios tornam difícil o discernimento entre eles. Domínios diferentes em geral têm funções diferentes, como a ligação de pequenas moléculas ou interação com outras proteínas. Proteínas pequenas normalmente têm somente um domínio (o domínio é a proteína).

O enovelamento dos polipeptídeos é sujeito a uma série de limitações físicas e químicas, e várias regras foram propostas a partir de estudos de padrões comuns de enovelamento proteico.

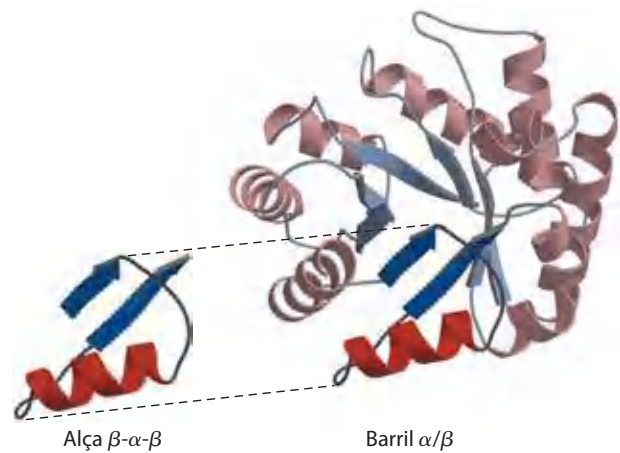
1. As interações hidrofóbicas dão uma grande contribuição para a estabilidade da estrutura de proteínas. O ocultamento dos grupos R dos aminoácidos hidrofóbicos, de modo a excluir a água, necessita de pelo menos duas camadas de estrutura secundária. Motivos simples como a alça  $\beta$ - $\alpha$ - $\beta$  (Figura 4-18a) criam essas camadas.
2. Quando ocorrem juntas em uma proteína, as hélices  $\alpha$  e as folhas  $\beta$  geralmente são encontradas em camadas estruturais diferentes. Isso porque o segmento do esqueleto polipeptídico em conformação  $\beta$  (Figura 4-6) não faz, com facilidade, ligações de hidrogênio com uma hélice  $\alpha$  adjacente a ele.
3. Segmentos adjacentes na sequência de aminoácidos normalmente se posicionam de forma adjacente na estrutura dobrada. Segmentos distantes do polipeptídeo podem se aproximar na estrutura terciária, mas não é a regra.
4. A conformação  $\beta$  é mais estável quando os segmentos individuais são levemente torcidos para a direita. Isso influencia tanto o arranjo das folhas  $\beta$  derivado dos segmentos supertorcidos quanto a forma da conexão do polipeptídeo entre elas. Duas folhas  $\beta$  paralelas, por exemplo, devem ser conectadas por uma fita cruzada (Figura 4-20a). Em princípio, essa fita pode ter uma conformação no sentido horário ou anti-horário, mas em proteínas o mais comum é no sentido horário. As conexões voltadas para a direita tendem a ser mais curtas do que as conexões voltadas para esquerda, e tendem a dobrar-se com um ângulo menor, o que as torna mais fáceis de serem formadas. A torção de folhas



**FIGURA 4-20 Padrões de enovelamento estável em proteínas.** (a) Conexões entre fitas  $\beta$  em arranjos de folhas  $\beta$ . As fitas aqui são visualizadas a partir de uma extremidade, sem torção. As linhas mais grossas representam as conexões nas extremidades mais próximas do observador; as linhas mais finas representam as conexões nas extremidades mais distantes das fitas  $\beta$ . As conexões em uma das extremidades (p. ex., aquelas próximas ao observador) raramente se cruzam umas com as outras. Um exemplo de tal cruzamento raro está ilustrado na fita amarela da estrutura a direita. (b) Devido à torção para a direita nas fitas  $\beta$ , as conexões entre as fitas em geral são no sentido horário. As conexões no sentido anti-horário devem assumir ângulos mais agudos e são mais difíceis de serem formadas. (c) Esta folha  $\beta$  torcida pertence a um domínio da fotoliase (proteína que repara certos tipos de danos ao DNA) de *E. coli* (derivado do PDB ID 1DNP). As alças de conexão foram removidas para que se foque no enovelamento da folha  $\beta$ .

$\beta$  também resulta em uma torção característica das estruturas formadas por vários desses segmentos, como pode ser observado no barril  $\beta$  (Figura 4-18b) e na folha  $\beta$  torcida (Figura 4-20c), que forma o núcleo de várias estruturas maiores.

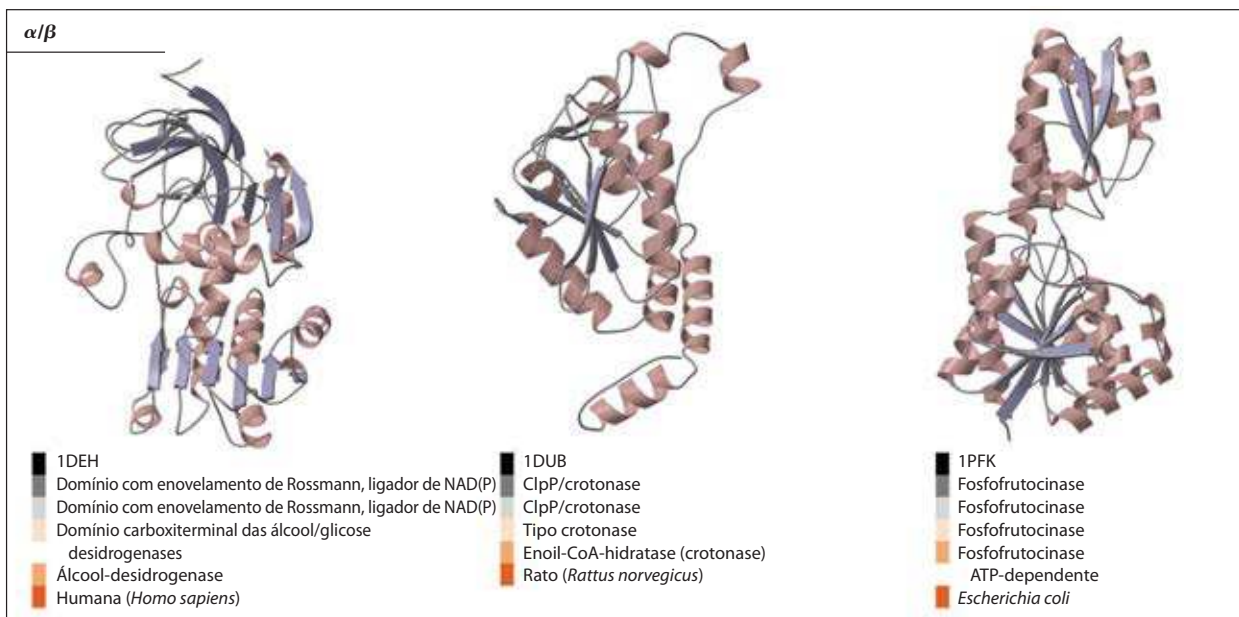
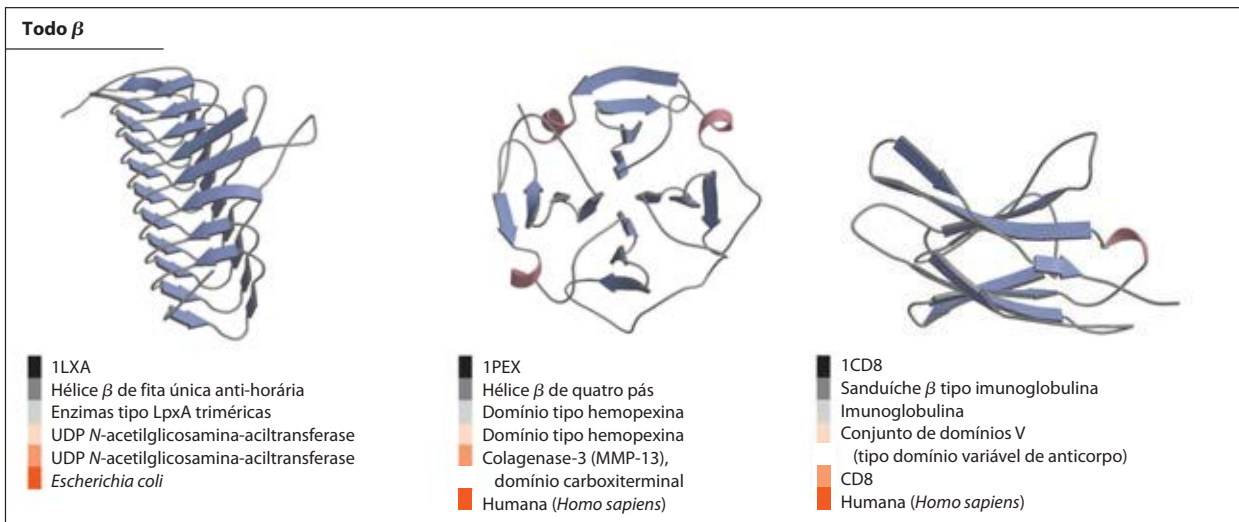
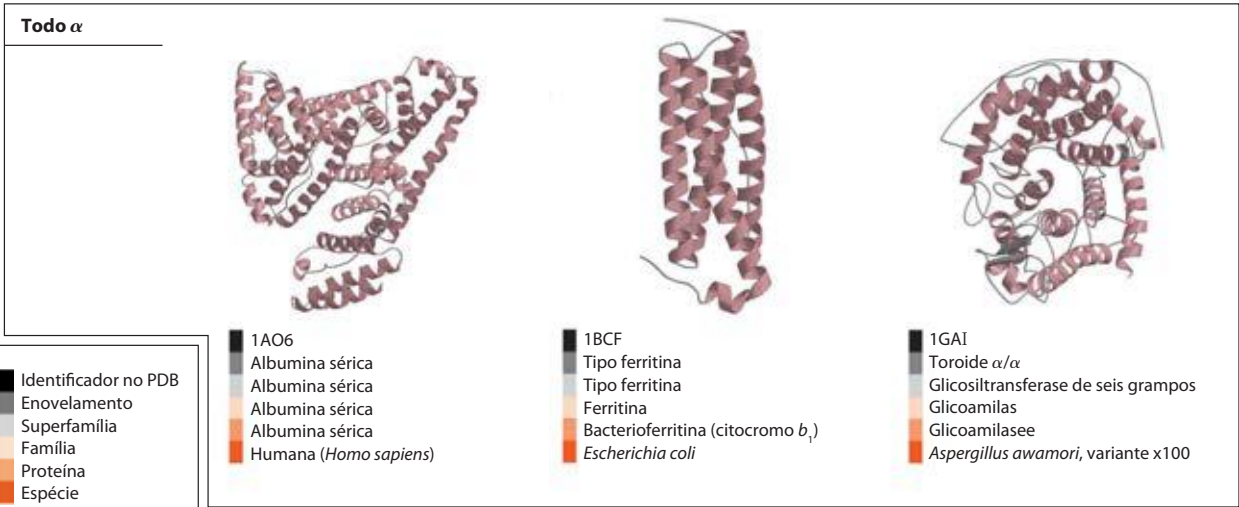
Seguindo essas regras, motivos complexos podem ser construídos a partir daqueles mais simples. Por exemplo, uma série de alças  $\beta$ - $\alpha$ - $\beta$  arranjadas de tal forma que as fitas  $\beta$  de um barril formem um motivo particularmente comum e estável, o **barril  $\alpha/\beta$**  (Figura 4-21). Nessa estrutura, cada segmento  $\beta$  paralelo é conectado ao seu vizinho por um segmento  $\alpha$ -helicoidal. Todas as conexões são voltadas para direita. O barril  $\alpha/\beta$  é encontrado em diversas enzimas, geralmente com um sítio de ligação (de um cofator ou substrato) na forma de uma fenda próxima a uma das extremidades do barril. Observe que domínios com padrão semelhante de enovelamento são ditos ter o mesmo motivo, mesmo que suas hélices  $\alpha$  e fitas  $\beta$  constituintes possam diferir em tamanho.



**FIGURA 4-21 A construção de domínios maiores a partir dos menores.** O barril  $\alpha/\beta$  é um domínio que ocorre comumente, sendo formado por repetições do motivo alça  $\beta$ - $\alpha$ - $\beta$ . Este barril  $\alpha/\beta$  é um domínio da piruvato-cinase (enzima glicolítica) de coelho (derivada do PDB ID 1PKN).

## Motivos de proteínas são as bases da classificação estrutural

**Arquitetura proteica – Estrutura terciária de proteínas globulares grandes, IV. Classificação estrutural das proteínas.** Como se vê, a compreensão da complexidade da estrutura terciária é facilitada pela consideração das subestruturas. Seguindo essa ideia, os pesquisadores organizaram o conteúdo completo de bancos de dados proteicos, de acordo com uma hierarquia dos níveis estruturais. Todos esses bancos de dados se baseiam nos dados e nas informações depositados no Protein Data Bank (PDB, Banco de Dados de Proteínas). O banco de dados Structural Classification of Proteins (SCOP, ou Classificação Estrutural de Proteínas) é um bom exemplo desta importante tendência na bioquímica. No mais alto nível de classificação, o banco de dados SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop>) utiliza um esquema já em uso, com quatro classes de estruturas de proteínas: **toda  $\alpha$** , **toda  $\beta$** ,  **$\alpha/\beta$**  (com segmentos  $\alpha$  e  $\beta$  intercalados ou alternados) e  **$\alpha + \beta$**  (com regiões  $\alpha$  e  $\beta$  um pouco separadas). Cada classe inclui dezenas a centenas de diferentes combinações de enovelamentos (motivos), construídas a partir de subestruturas gradativamente identificáveis. Alguns desses arranjos subestruturais são bem comuns, outros foram encontrados em apenas uma proteína. A **Figura 4-22** mostra a variedade desses motivos distribuídos entre as quatro classes de estrutura de proteínas. Essa é apenas uma pequena amostra das centenas de motivos conhecidos. Entretanto, o número de padrões de enovelamento não é infinito. À medida que a velocidade na qual novas estruturas de proteínas são elucidadas, a fração dessas estruturas que contém novos motivos tem diminuído. Ao todo, devem existir menos de 1.000 motivos diferentes. A Figura 4-22 também mostra como as proteínas podem ser organizadas em função da presença dos vários motivos. Os níveis de organização mais superiores, **classe** e **enovelamento**, são puramente estruturais. Abaixo do nível de enovelamento (ver o código de cores na Figura 4-22), a classificação tem como base as relações evolutivas.



**FIGURA 4-22 Organização das proteínas com base nos motivos.** Mostrados aqui estão apenas alguns das centenas de motivos estáveis conhecidos, divididos em quatro classes: todo  $\alpha$ , todo  $\beta$ ,  $\alpha/\beta$  e  $\alpha + \beta$ . Os dados de classificação estrutural do banco de dados SCOP (Structural Classification of Proteins, <http://scop.mrc-lmb.cam.ac.uk/scop>) também são fornecidos

(ver código de cores). O identificador no PDB (listado inicialmente para cada estrutura) corresponde ao código único de acesso de cada estrutura depositada no Protein Data Bank ([www.pdb.org](http://www.pdb.org)). O barril  $\alpha/\beta$  (ver Figura 4-21) é outro motivo  $\alpha/\beta$  particularmente comum. **(Continua na próxima página)**



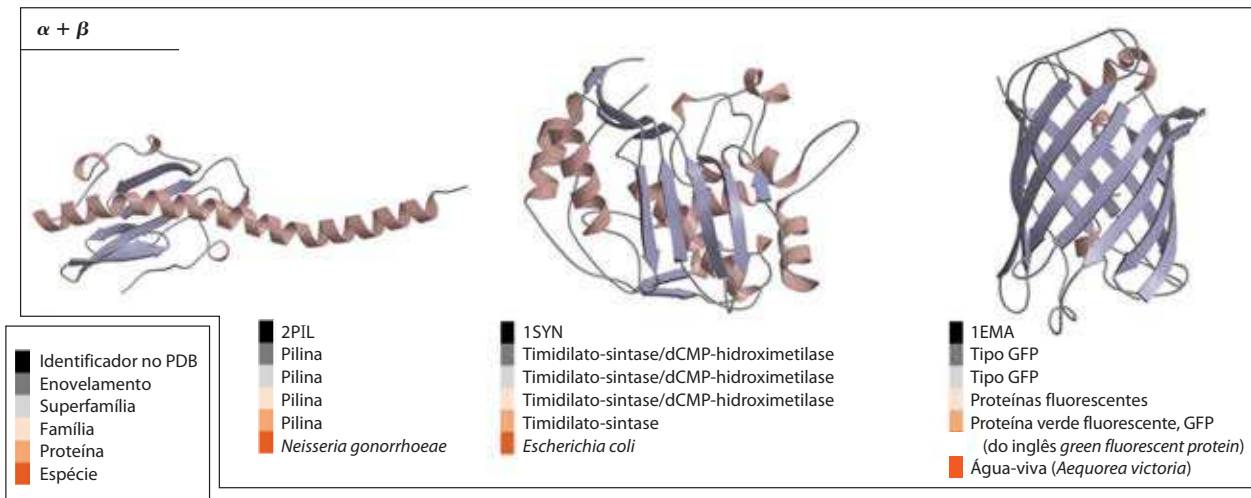


FIGURA 4-22 (Continuação)

Vários exemplos de domínios recorrentes ou motivos estruturais estão disponíveis, revelando que a estrutura terciária das proteínas é mais conservada do que suas sequências de aminoácidos. A comparação entre as estruturas de proteínas pode, então, fornecer muita informação sobre sua evolução. Proteínas com significativa semelhança na estrutura primária e/ou com estrutura terciária e funções semelhantes fazem parte de uma mesma **família proteica**. Uma relação evolutiva forte normalmente é evidente em uma família de proteínas. Por exemplo, a família globina apresenta muitas proteínas diferentes com similaridade estrutural e de sequência com a mioglobina (como visto nas proteínas usadas nos exemplos do Quadro 4-5 e no Capítulo 5). Duas ou mais famílias com pouca similaridade na sequência de aminoácidos algumas vezes utilizam o mesmo motivo estrutural geral e apresentam semelhanças funcionais; essas famílias são agrupadas como **superfamílias**. Uma relação evolutiva entre as famílias em uma superfamília é considerada provável, ainda que as diferenças temporais e funcionais – resultantes de diferentes pressões adaptativas – possam ter suprimido muitas das relações reveladoras entre sequências. Uma família de proteínas pode estar distribuída em todos os três domínios de vida celular: Bacteria, Archaea e Eukarya, sugerindo uma origem remota. Muitas proteínas envolvidas no metabolismo intermediário e no metabolismo dos ácidos nucleicos e proteínas se encaixam nessa categoria. Outras famílias podem estar presentes somente em um pequeno grupo de organismos, indicando que aquela estrutura surgiu mais recentemente. Rastrear a história natural dos motivos estruturais, usando as classificações estruturais de bancos de dados como o SCOP, fornece um grande complemento às análises de sequência na investigação da relação evolutiva. O banco de dados SCOP é manualmente gerenciado, com o objetivo de colocar as proteínas na sua rede evolutiva correta, com base nas características estruturais conservadas.

Motivos estruturais se tornaram especialmente importantes na definição de famílias e superfamílias de proteínas. Sistemas aperfeiçoados de classificação e elucidação de proteínas inevitavelmente levaram à elucidação de novas relações funcionais. Dada a importância das proteínas nos

sistemas vivos, essas comparações estruturais podem ajudar no esclarecimento de todos os aspectos da bioquímica, da evolução de proteínas individuais até a história evolutiva de rotas metabólicas completas.

### A estrutura quaternária varia de dímeros simples a grandes complexos

**Arquitetura proteica – Estrutura quaternária** Muitas proteínas têm múltiplas subunidades polipeptídicas (de duas a centenas). A associação das cadeias polipeptídicas pode servir a uma variedade de funções. Muitas proteínas com múltiplas subunidades podem ter funções reguladoras; a ligação de pequenas moléculas pode afetar a interação entre as subunidades, causando grandes mudanças na atividade da proteína em resposta a pequenas mudanças na concentração de substrato ou moléculas reguladoras (Capítulo 6). Em outros casos, subunidades distintas assumem funções diferentes, mas relacionadas, como catálise e regulação. Algumas associações, como nas proteínas fibrosas consideradas anteriormente neste capítulo e nas proteínas dos capsídeos dos vírus, têm, principalmente, funções estruturais. Algumas associações proteicas muito grandes formam sítios de reações complexas envolvendo múltiplas etapas. Por exemplo, cada ribossomo, o local onde ocorre a síntese de proteínas, incorpora dúzias de subunidades de proteínas com certo número de moléculas de RNA.

Uma proteína de múltiplas subunidades também é chamada de **multímero**. Um múltímero com apenas algumas subunidades é chamado de **oligômero**. Se um múltímero apresenta subunidades não idênticas, a estrutura total da proteína pode ser assimétrica e bem complicada. Entretanto, a maioria dos múltímeros apresenta subunidades idênticas, ou grupos de subunidades não idênticas que se repetem, geralmente em arranjos simétricos. Como apresentado no Capítulo 3, as unidades estruturais de repetição em uma proteína multimérica, tanto de um único tipo de subunidade quanto de um grupo de subunidades, são chamadas de **protômero**. Muitas vezes são utilizadas letras gregas para distinguir as diferentes subunidades que fazem parte do protômero.



A primeira proteína oligomérica que teve sua estrutura tridimensional determinada foi a hemoglobina ( $M_r$  64.500), que contém quatro cadeias polipeptídicas e quatro grupos prostéticos heme, nos quais os átomos de ferro se apresentam no estado ferroso ( $\text{Fe}^{2+}$ ) (Figura 4-17). A porção proteica, a globina, consiste em duas cadeias (141 resíduos em cada cadeia  $\alpha$ ) e duas cadeias  $\beta$  (146 resíduos em cada cadeia). Observe que neste caso,  $\alpha$  e  $\beta$  não se referem às estruturas secundárias. Em uma prática, que pode ser confusa para o aluno iniciante, as letras gregas  $\alpha$  e  $\beta$  ( $\gamma$  e  $\delta$ , e outras) são frequentemente usadas para distinguir dois tipos diferentes de subunidades em uma proteína com múltiplas subunidades, independentemente de quais tipos de estruturas secundárias podem predominar nas subunidades. Como a hemoglobina é quatro vezes maior do que a mioglobina, foi necessário muito mais tempo e esforço para resolver sua estrutura tridimensional por análise de raios X, o que foi finalmente obtido por Max Perutz, John Kendrew e seus colegas, em 1959. As subunidades da hemoglobina estão organizadas em pares simétricos (Figura 4-23). Cada par tem uma subunidade  $\alpha$  e uma subunidade  $\beta$ . A hemoglobina pode, portanto, ser descrita tanto como tetramero quanto dímero de protômeros  $\alpha\beta$ . O papel dessas subunidades distintas na função da hemoglobina é amplamente discutido no Capítulo 5.



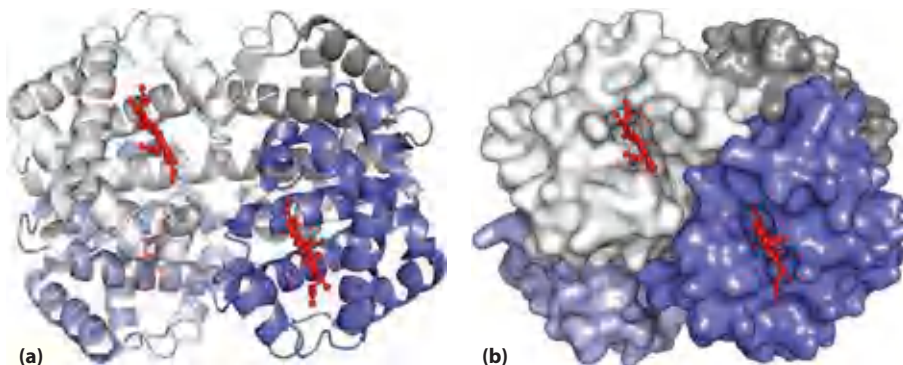
Max Perutz, 1914–2002 (esquerda) e John Kendrew, 1917–1997

### Algumas proteínas ou segmentos proteicos são intrinsecamente desordenados

Apesar de décadas de progresso no entendimento da estrutura proteica, muitas proteínas não podem ser cristalizadas, tornando difícil a determinação de suas estruturas tridimensionais por métodos considerados clássicos (ver Quadro 4-5). Mesmo quando cristalizadas, partes da proteína estão suficientemente desordenadas no cristal de forma que a estrutura determinada não inclui essas partes. Algumas vezes, isso se deve a características sutis da estrutura que tornam a cristalização difícil. No entanto, a razão pode ser mais simples: algumas proteínas ou segmentos proteicos carecem de estruturas ordenadas em solução.

O conceito de que algumas proteínas funcionam na ausência de uma estrutura definida é um produto da reavaliação dos dados envolvendo muitas proteínas diferentes. Um terço de todas as proteínas humanas pode ser desestruturado ou possuem segmentos desestruturados significativos. Todos os organismos possuem algumas proteínas que se encaixam nessa categoria. As **proteínas intrinsecamente desordenadas** têm propriedades distintas das proteínas estruturadas clássicas. Elas carecem de um núcleo hidrofóbico e, ao contrário, são caracterizadas por alta densidade de aminoácidos carregados, como Lys, Arg e Glu. Resíduos de Pro também são proeminentes, já que eles tendem a romper estruturas ordenadas.

A desordem estrutural e a alta densidade de cargas podem facilitar a atividade de algumas proteínas, como espaçadores, isoladores ou elementos de ligação em estruturas maiores. Outras proteínas desordenadas são sequestradoras, ligando íons e moléculas pequenas em solução e servindo de reservatórios ou depósitos de lixo. No entanto, muitas proteínas intrinsecamente desordenadas são o núcleo de importantes redes de interações proteicas. A falta de uma estrutura ordenada pode facilitar um tipo de promiscuidade funcional, permitindo a interação de uma proteína com múltiplos parceiros. Algumas proteínas intrinsecamente desordenadas atuam por inibir a atividade de outras proteínas por um mecanismo não usual: por envolver-se em torno de suas proteínas-alvo. Uma proteína desordenada pode ter algumas ou mesmo dúzias de proteínas parceiras. A desordem estrutural permite ao inibidor proteico se envolver em torno de múltiplos alvos de diferentes formas. A proteína

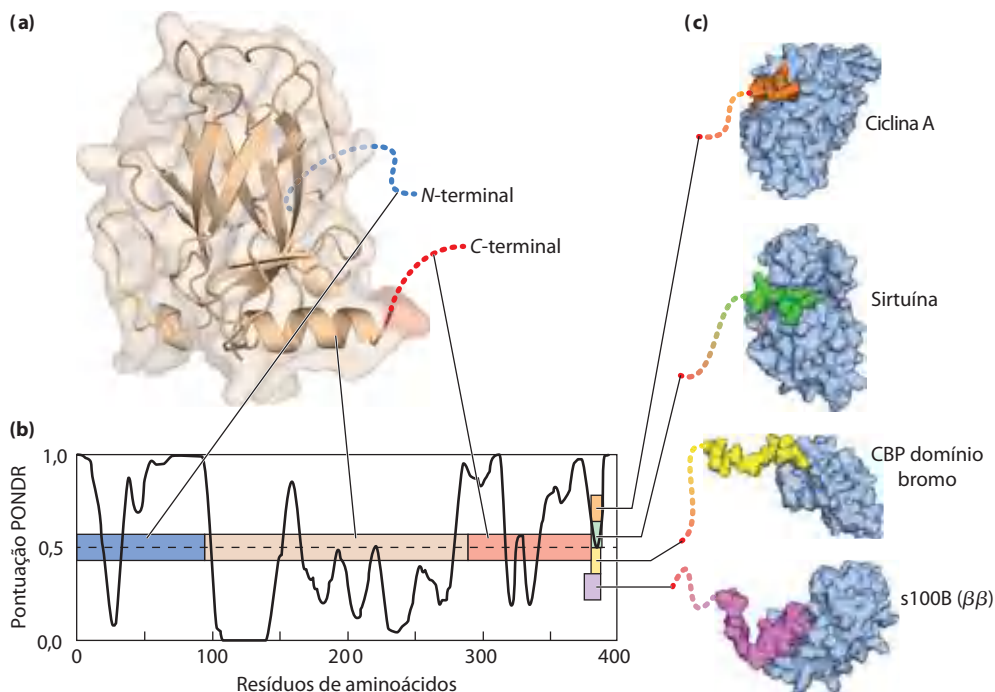


**FIGURA 4-23** Estrutura quaternária da desóxi-hemoglobina. (PDB ID 2HHB) A análise por difração de raios X da desóxi-hemoglobina (hemoglobina sem as moléculas de oxigênio ligadas aos grupos heme) mostra como as quatro subunidades polipeptídicas estão agrupadas. **(a)** Representação na forma de fitas revela os elementos estruturais secundários da estrutura e a posição de todos os cofatores hemes. **(b)** Modelo de contorno de superfície mostra a cavidade em que os cofatores hemes estão ligados e ajuda a visualizar o empacotamento das subunidades. As subunidades  $\alpha$  são mostradas em tons de cinza, e as subunidades  $\beta$ , em tons de azul. Observe que os grupos heme (em vermelho) estão relativamente distantes entre si.

intrinsecamente desordenada p27 exerce um papel-chave no controle da divisão celular em mamíferos. Essa proteína carece de estrutura definida quando em solução. Ela se enrola em torno e assim inibe a atividade de várias enzimas chamadas proteínas-quinasas (ver Capítulo 6) que facilitam a divisão. A estrutura flexível da p27 a permite se acomodar às suas proteínas-alvo distintas. Células tumorais humanas, que são simplesmente células que perderam a capacidade de controlar a divisão celular normalmente, geralmente possuem níveis reduzidos de p27; quanto menor o nível de p27, pior é o prognóstico para o paciente de câncer. Da mesma forma, as proteínas intrinsecamente desordenadas estão frequentemente presentes como concentradores ou suportes no centro de redes proteicas que constituem as vias de sinalização. Essas proteínas ou partes delas podem interagir com muitos parceiros diferentes. Elas frequentemente assumem uma estrutura quando interagem com outras proteínas, mas as estruturas que elas assumem podem variar de acordo com o parceiro de ligação. A proteína de mamíferos p53 também é fundamental no controle da divisão celular. Ela tem segmentos estruturados e não estruturados, e os segmentos distintos interagem com dúzias de outras proteínas. Uma região não estruturada da p53 na extremidade carboxil interage com pelo menos quatro parceiros diferentes e assume uma estrutura distinta em cada complexo (**Figura 4-24**).

### RESUMO 4.3 Estruturas terciária e quaternária das proteínas

- ▶ A estrutura terciária é a estrutura tridimensional da cadeia polipeptídica. Muitas proteínas se encaixam em uma ou duas classes de proteínas em geral, com base na estrutura terciária: fibrosa e globular.
- ▶ As proteínas fibrosas, utilizadas principalmente para funções estruturais, são formadas por elementos de estrutura secundária que se repetem.
- ▶ As proteínas globulares têm estruturas terciárias mais complicadas, geralmente contendo diversos tipos de estruturas secundárias na mesma cadeia polipeptídica. A primeira estrutura de proteína globular a ser determinada, por métodos de difração de raios X, foi a estrutura da mioglobina.
- ▶ As estruturas complexas das proteínas globulares podem ser analisadas pela observação dos padrões de enovelamento chamados de motivos (também chamados de enovelamentos ou estruturas supersecundárias). Os milhares de estruturas proteicas conhecidos geralmente são formados por um repertório de apenas poucas centenas de motivos. Os domínios são regiões de uma cadeia polipeptídica que podem se dobrar de forma estável e independente.
- ▶ A estrutura quaternária resulta de interações entre as subunidades de proteínas com múltiplas subunidades



**FIGURA 4-24** Ligação da região carboxiterminal intrinsecamente desordenada da proteína p53 aos seus ligantes. (a) A proteína p53 é composta por vários segmentos diferentes (PDB ID 1XQH). Apenas o domínio central é bem organizado. (b) A sequência linear da proteína p53 está representada na barra colorida. O gráfico sobreposto apresenta um diagrama da pontuação PONDR (Predictor of Natural Disordered Regions; Previsão de Regiões Naturalmente Desordenadas) versus a sequência da proteína. PONDR é um dos melhores algoritmos disponíveis para prever a probabilidade que um determinado resíduo de aminoácido está em uma região de desordem intrínseca, com base na sequência de aminoácidos ao redor e na

composição de aminoácidos. Uma pontuação de 1,0 indica uma probabilidade de 100% que a proteína estará desordenada. Nesta estrutura proteica, o domínio central em bege está organizado. As regiões aminoterminais (azul) e carboxiterminais (vermelho) estão desordenadas. A extremidade da região carboxiterminal tem múltiplos ligadores e ela se dobra quando se liga a cada um deles; no entanto, a estrutura tridimensional assumida quando ocorre a ligação é diferente para cada uma das interações mostradas, e por isso o segmento carboxiterminal (resíduos de 11 a 20) está mostrado em cor diferente em cada complexo (ciclina A, PDB ID 1H26; sirtuína, PDB ID 1MA3; CBP domínio bromo, PDB ID 1JSP; s100B (ββ), PDB ID 1DT7).

(multiméricas) ou grandes associações de proteínas. Algumas proteínas multiméricas possuem unidades repetidas, formadas por uma única subunidade ou grupos de subunidades, cada unidade chamada protômero.

- ▶ Algumas proteínas ou segmentos proteicos são intrinsecamente desordenados, carecendo de estruturas definidas. Essas proteínas têm composições distintas de aminoácidos que permitem uma estrutura mais flexível. Algumas dessas proteínas desordenadas funcionam como componentes estruturais ou sequestradoras; outras podem interagir com muitas proteínas parceiras, servindo como inibidores versáteis ou como componentes centrais de redes de interação de proteínas.

#### 4.4 Desnaturação e enovelamento das proteínas

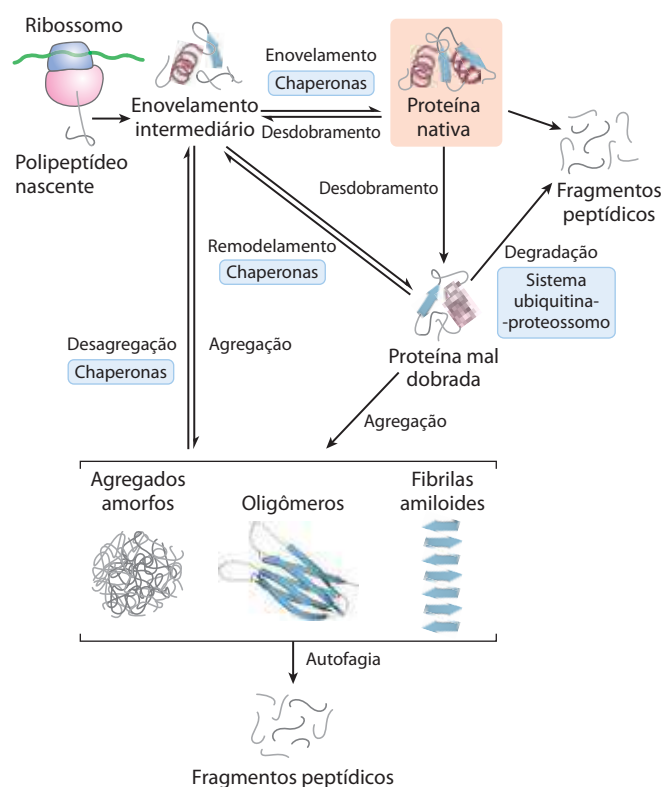
As proteínas têm uma existência surpreendentemente precária. Como foi visto, a conformação de uma proteína nativa é apenas marginalmente estável. Além disso, a maioria das proteínas deve manter certa flexibilidade conformacional para funcionar. A manutenção contínua do grupo ativo de proteínas celulares, necessárias em um dado conjunto de condições, é chamada **proteostase**. A proteostase celular requer a atividade coordenada de vias para síntese e enovelamento de proteínas, o redobramento de proteínas parcialmente desdobradas e o sequestro e degradação de proteínas irreversivelmente desdobradas. Em todas as células, essas redes envolvem centenas de enzimas e proteínas especializadas.

Como visto na **Figura 4-25**, a vida de uma proteína engloba muito mais do que sua síntese e degradação. A estabilidade marginal da maioria das proteínas pode produzir um balanço tênue entre os estados dobrados e desdobrados. À medida que as proteínas são sintetizadas nos ribossomos (Capítulo 27), elas devem dobrar-se em sua conformação nativa. Algumas vezes isso ocorre espontaneamente, porém mais frequentemente isso ocorre com a assistência de enzimas e complexos especializados chamados chaperonas. Muitos desses mesmos auxiliares do enovelamento atuam para redobrar proteínas que se tornaram transitivamente desdobradas. As proteínas inapropriadamente dobradas frequentemente expõem superfícies hidrofóbicas que as tornam “pegajosas”, conduzindo à formação de agregados inativos. Esses agregados podem perder suas funções normais, mas não são inertes; seu acúmulo nas células situa-se no centro de doenças que vão de diabetes a doenças de Parkinson e Alzheimer. Não surpreendentemente, todas as células elaboraram vias de reciclagem e/ou degradação de proteínas irreversivelmente deformadas.

Agora serão abordadas as transições entre os estados dobrados e desdobrados, e a rede de vias que controlam essas transições.

#### A perda de estrutura da proteína resulta na perda de função

As estruturas proteicas evoluíram para atuar em determinados ambientes celulares. Condições diferentes daquelas da célula podem resultar em mudanças estruturais grandes ou pequenas na proteína. A perda de estrutura tridimensional suficiente para causar a perda de função é chamada de

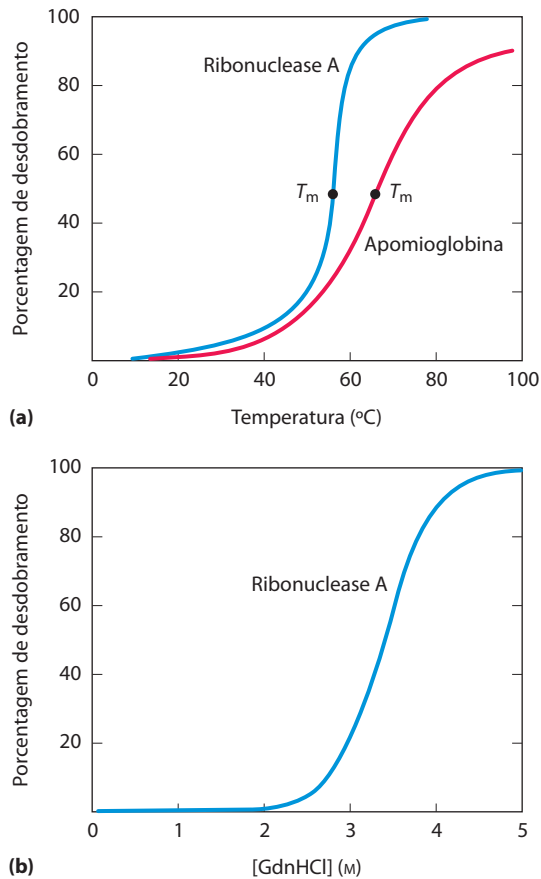


**FIGURA 4-25 Vias que contribuem para proteostase.** Três tipos de processos contribuem para proteostase. Primeiro, as proteínas são sintetizadas no ribossomo. Segundo, múltiplas vias contribuem para o enovelamento proteico, muitas dessas envolvem a atividade de complexos chamados chaperonas. As chaperonas (incluindo as chaperoninas) também contribuem para o redobramento de proteínas parcialmente ou transitivamente desdobradas. Finalmente, proteínas irreversivelmente desdobradas estão sujeitas ao sequestro e à degradação por várias vias adicionais. As proteínas desdobradas e os intermediários proteicos enovelados que escapam do controle de qualidade das chaperonas e das vias de degradação podem agregar, formando tanto agregados desordenados quanto agregados organizados do tipo amiloide que contribuem para doenças e processos de envelhecimento.

**desnaturação.** O estado desnaturado não necessariamente corresponde ao desdobramento completo da proteína e à randomização da conformação. Na maioria das condições, as proteínas desnaturadas existem como um conjunto de estados parcialmente dobrados.

A maioria das proteínas pode ser desnaturada pelo calor, que tem efeitos complexos nas muitas interações fracas da proteína (principalmente sobre as ligações de hidrogênio). Se a temperatura é aumentada lentamente, a conformação da proteína em geral permanece intacta até que, em uma estreita faixa de temperatura, ocorre uma perda abrupta da estrutura (e da função) (**Figura 4-26**). A mudança repentina sugere que o desdobramento é um processo cooperativo: a perda de estrutura em uma parte da proteína desestabiliza as outras partes. Os efeitos do aquecimento sobre as proteínas não são facilmente preditos. As proteínas altamente termorresistentes de bactérias termofílicas e de arqueobactérias evoluíram para funcionar na temperatura das fontes termais ( $\sim 100^{\circ}\text{C}$ ). Mesmo assim, as estruturas dessas proteínas diferem pouco de suas proteínas homólogas de bactérias como a *Escherichia coli*. Não se sabe ao





**FIGURA 4-26** Desnaturação de proteínas. Os resultados correspondem a proteínas desnaturadas por duas modificações diferentes em seus ambientes. Em cada caso, a transição entre o estado dobrado e o estado não dobrado é abrupta, sugerindo cooperatividade neste processo. **(a)** Desnaturação térmica da apomioglobina (mioglobina sem o grupo prostético heme) de cavalo e da ribonuclease A (com suas ligações dissulfeto intactas; ver Figura 4-27). O ponto médio da faixa de temperatura em que ocorre a desnaturação é chamado de temperatura de fusão, ou  $T_m$ . A desnaturação da apomioglobina foi monitorada por dicroísmo circular (ver Figura 4-10), que mede a quantidade de estrutura helicoidal na proteína. A desnaturação da ribonuclease A foi acompanhada pelo monitoramento das mudanças na fluorescência intrínseca da proteína, que é afetada pelas mudanças no ambiente dos resíduos de Trp. **(b)** Desnaturação da ribonuclease A, com suas ligações dissulfeto intactas, pelo hidrócloro de guanidina (GdnHCl), monitorado por dicroísmo circular.

certo como essas pequenas diferenças garantem a estabilidade estrutural em altas temperaturas.

As proteínas também podem ser desnaturadas por pHs extremos, por certos solventes orgânicos miscíveis, como álcool ou acetona, por certos solutos como ureia e hidrócloro de guanidina, ou por detergentes. Cada um desses agentes desnaturantes representa um tratamento relativamente brando, já que nenhuma ligação covalente da cadeia polipeptídica é rompida. Solventes orgânicos, ureia e detergentes atuam, principalmente, rompendo as interações hidrofóbicas que mantêm o núcleo estável das proteínas globulares; a ureia também rompe as ligações de hidrogênio; extremos de pH alteram a carga líquida da proteína, causando repulsão eletrostática e o rompimento de algumas ligações de hidrogênio. As estruturas desnaturadas resultantes desses vários tratamentos não são, necessariamente, as mesmas.

A desnaturação frequentemente leva à precipitação de proteínas, uma consequência da formação de agregados proteicos pela exposição de superfícies hidrofóbicas associadas. Os agregados são, com frequência, altamente desordenados. O precipitado proteico visto após ferver um ovo é um exemplo. Agregados mais ordenados também são observados em algumas proteínas, como será visto.

### A sequência de aminoácidos determina a estrutura terciária

A estrutura terciária de uma proteína globular é determinada por sua sequência de aminoácidos. A prova mais importante disso vem dos experimentos que mostram que a desnaturação de algumas proteínas é reversível. Certas proteínas globulares desnaturadas por temperatura, extremos de pH ou agentes desnaturantes reassem suas estruturas nativas e suas atividades biológicas se retornarem às condições nas quais a conformação nativa é estável. Esse processo é chamado de **renaturação**.

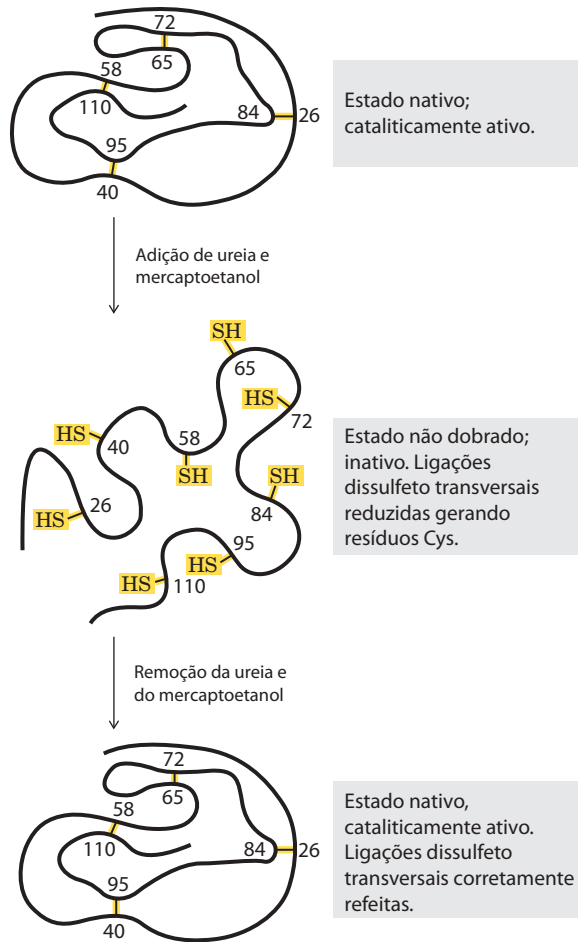
Um exemplo clássico é a desnaturação e a renaturação da ribonuclease A, demonstradas por Christian Anfinsen nos anos 1950. A ribonuclease A purificada desnatura completamente em uma solução concentrada de ureia, na presença de um agente redutor. O agente redutor rompe as quatro ligações dissulfeto, resultando em oito resíduos Cys, e a ureia rompe as interações hidrofóbicas de estabilização, desta forma liberando todo o polipeptídeo de sua conformação dobrada. A desnaturação da ribonuclease é acompanhada por uma completa perda de atividade catalítica. Quando a ureia e o agente redutor são removidos, a ribonuclease desnaturada, aleatoriamente enrolada, se dobra de modo espontâneo em sua estrutura terciária correta, com restauração total de sua atividade catalítica (**Figura 4-27**). O reenovelamento da ribonuclease é tão preciso que as quatro ligações dissulfeto intramoleculares são restabelecidas na molécula renaturada nas mesmas posições da ribonuclease nativa. Mais tarde, resultados semelhantes foram obtidos utilizando a ribonuclease A cataliticamente ativa quimicamente sintetizada. Isso elimina a possibilidade de que algum contaminante minoritário da preparação da ribonuclease purificada por Anfinsen tenha contribuído para a renaturação da enzima, afastando assim qualquer dúvida que poderia restar de que essa enzima se dobra espontaneamente.

O experimento de Anfinsen forneceu a primeira evidência de que a sequência de aminoácidos de uma cadeia polipeptídica contém todas as informações necessárias para o enovelamento da cadeia em sua estrutura tridimensional nativa. Trabalhos posteriores demonstraram que somente uma minoria de proteínas, muitas delas pequenas e estáveis por natureza, se dobrará espontaneamente em sua forma nativa. Apesar de todas as proteínas terem o potencial de se dobrar em sua estrutura nativa, muitas delas necessitam de alguma assistência.

### Os polipeptídeos dobram-se rapidamente por um processo gradual

Nas células vivas, as proteínas são construídas, a partir dos aminoácidos, em uma velocidade muito alta. Por exemplo, as células de *E. coli* podem fazer uma molécula de proteína



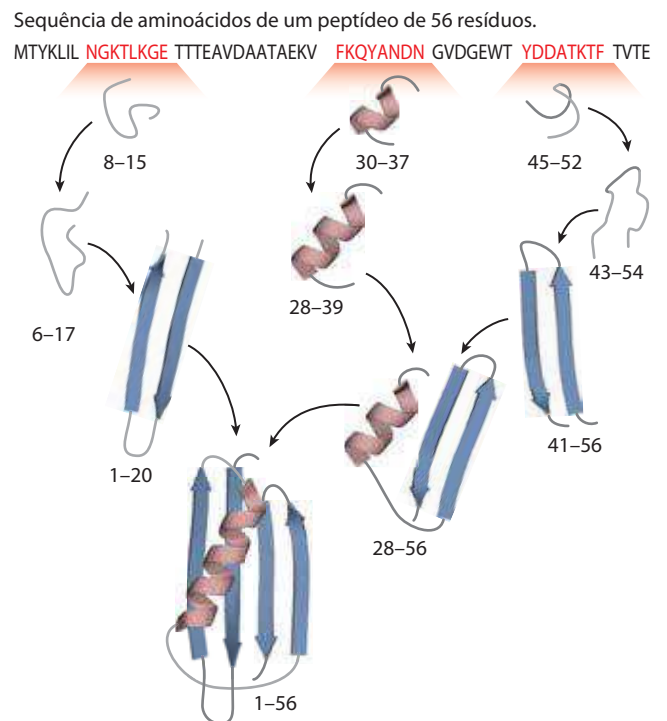


**FIGURA 4-27 Renaturação da ribonuclease desnaturada e desdobrada.** A ureia desnatura a ribonuclease, e o mercaptoetanol (HOCH<sub>2</sub>CH<sub>2</sub>SH) a reduz, rompendo as ligações dissulfeto e liberando os oito resíduos de Cys. A renaturação envolve o restabelecimento correto das ligações dissulfeto transversais.

completa, biologicamente ativa, com 100 resíduos de aminoácidos, em 5 segundos a 37°C. No entanto, a síntese das ligações peptídicas no ribossomo não é suficiente; a proteína deve dobrar-se.

Como a cadeia polipeptídica chega à sua conformação nativa? Suponha-se, de forma conservadora, que cada um dos resíduos de aminoácidos pode assumir, em média, 10 conformações diferentes, resultando em 10<sup>100</sup> conformações diferentes do polipeptídeo. Suponha-se, também, que a proteína se dobra espontaneamente por um processo aleatório no qual ela testa todas as conformações possíveis em torno de cada uma das ligações de seu esqueleto até encontrar sua forma nativa, biologicamente ativa. Se cada conformação fosse testada no menor tempo possível (~10<sup>-13</sup> segundos, ou o tempo necessário para uma única vibração molecular), levaria aproximadamente 10<sup>77</sup> anos para testar todas as possíveis conformações. Claramente, o enovelamento de proteínas não é um processo completamente aleatório, de tentativa e erro. Deve haver atalhos. Esse problema foi apontado pela primeira vez por Cyrus Levinthal, em 1968, sendo algumas vezes chamado de paradoxo de Levinthal.

A rota de enovelamento de uma cadeia polipeptídica grande é muito complicada. No entanto, rápido progresso tem sido feito nesse campo, o suficiente para produzir algoritmos robustos que podem frequentemente prever a estrutura de proteínas menores, com base em suas sequências de aminoácidos. As principais vias de enovelamento são hierárquicas. As estruturas secundárias locais se formam primeiro. Certas sequências de aminoácidos se dobras prontamente em hélices  $\alpha$  ou folhas  $\beta$ , guiadas por restrições como aquelas revisadas na discussão sobre estruturas secundárias. Interações iônicas, envolvendo grupos carregados que normalmente estão próximos na sequência linear da cadeia polipeptídica, podem ter um papel importante no direcionamento dos primeiros passos de enovelamento. O arranjo de estruturas locais é seguido por interações de longo alcance entre, por exemplo, elementos da estrutura secundária que se aproximam para formar estruturas enoveladas estáveis. As interações hidrofóbicas exercem um papel significativo ao longo do processo, já que a agregação das cadeias laterais de aminoácidos apolares fornece uma estabilização entrópica a intermediários e, por fim, à estrutura enovelada final. O processo continua até a formação de domínios completos e até que todo o polipeptídeo esteja dobrado (Figura 4-28). Proteínas nas quais predominam as interações de curto alcance (entre pares de resíduos geralmente localizados próximos um do outro na sequência polipeptídica) tendem a se dobrar mais rapidamente do que



**FIGURA 4-28 Uma via de enovelamento proteico definida para uma proteína pequena.** Uma via hierárquica é mostrada com base em modelagem computacional. Primeiro, pequenas regiões de estrutura secundária são agrupadas e, então, gradualmente incorporadas em estruturas maiores. O programa usado para este modelo tem tido muito êxito na predição da estrutura tridimensional de pequenas proteínas a partir de sua sequência de aminoácidos. Os números indicam os resíduos de aminoácidos desse peptídeo de 56 resíduos que adquiriu sua estrutura final em cada uma das etapas mostradas.

proteínas com padrões de enovelamento mais complexos e com mais interações de longo alcance entre os diferentes segmentos. À medida que as proteínas com múltiplos domínios são sintetizadas, domínios próximos à extremidade aminoterminal (sintetizadas primeiro) podem se dobrar antes do polipeptídeo inteiro ter sido montado.

Termodinamicamente, o processo de enovelamento pode ser visto como um tipo de funil de energia livre (Figura 4-29). Os estados não dobrados são caracterizados por um alto grau de entropia conformacional e energia livre relativamente alta. À medida que o enovelamento ocorre, o estreitamento do funil reflete a diminuição do espaço conformacional que deve ser procurado, à medida que a proteína se aproxima de seu estado nativo. As pequenas depressões ao longo das paredes do funil de energia livre representam intermediários semiestáveis, que podem tornar o processo de enovelamento um pouco mais lento. No fundo do funil, o conjunto de intermediários dobrados é reduzido a uma única conformação nativa (ou a um pequeno conjunto de conformações nativas). Os funis podem ter uma variedade de formas dependendo da complexidade da via de enovelamento, da existência de intermediários semiestáveis, e do potencial de intermediários em particular se reunirem em agregados de proteínas erroneamente dobradas (Figura 4-29).

A estabilidade termodinâmica não é igualmente distribuída na estrutura da proteína – a molécula tem regiões de relativamente alta estabilidade e outras de estabilidade baixa ou desprezível. Por exemplo, uma proteína pode ter dois domínios estáveis ligados por um segmento inteiramente desordenado. Regiões de menor estabilidade podem permitir à proteína alterar sua conformação entre dois ou mais estados. Como será visto nos próximos dois capítulos, variações na estabilidade em regiões de uma proteína são, com frequência, essenciais para sua função. Proteínas ou

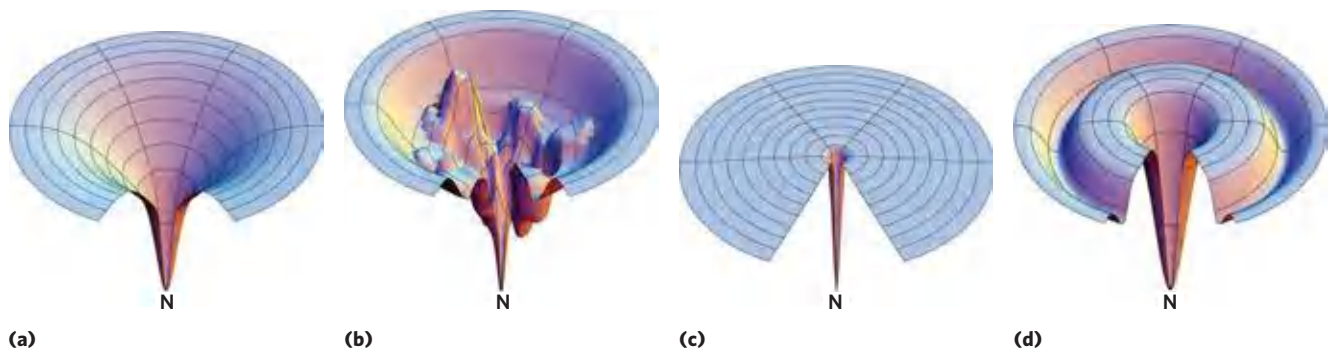
segmentos proteicos intrinsecamente desordenados não se dobram de forma alguma.

À medida que o entendimento sobre estrutura e enovelamento de proteínas evolui, programas computacionais sofisticados para a predição da estrutura das proteínas a partir da sequência de aminoácidos têm sido desenvolvidos. A predição da estrutura tridimensional das proteínas é uma das especialidades da bioinformática, e os progressos nessa área são monitorados com um teste bianual chamado de competição CASP (Critical Assessment of Structural Prediction, ou Avaliação Crítica da Predição Estrutural). Grupos de diversas partes do mundo tentam prever a estrutura de certa proteína (cuja estrutura já foi determinada, mas ainda não foi publicada). Os grupos que têm mais sucesso são convidados a apresentar seus resultados na conferência da CASP. O sucesso desses esforços está melhorando rapidamente.

### Algumas proteínas se dobram de forma assistida

Nem todas as proteínas se dobram espontaneamente à medida que são sintetizadas dentro da célula. O enovelamento de muitas proteínas necessita de **chaperonas**, proteínas que interagem com polipeptídeos parcialmente dobrados ou dobrados de forma incorreta, facilitando os mecanismos de enovelamento correto ou garantindo um microambiente adequado para ocorrer o enovelamento. Vários tipos de chaperonas moleculares foram encontrados em organismos que vão desde bactérias até humanos. As duas principais famílias de chaperonas, ambas bem estudadas, são a família da **Hsp70** e as **chaperoninas**.

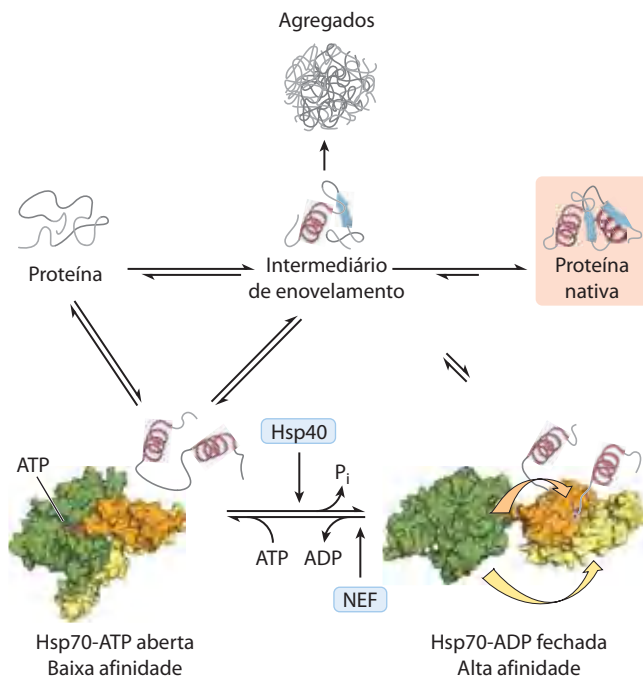
A família de proteínas Hsp70 geralmente tem massa molecular próxima de 70.000 e são abundantes em células submetidas a altas temperaturas (por isso, a denomina-



**FIGURA 4-29** Termodinâmica do enovelamento proteico, mostrado como funil de energia livre. À medida que as proteínas se dobram, o espaço conformacional que pode ser explorado pela estrutura fica restrito. Este é modelado como funil termodinâmico tridimensional, com  $\Delta G$  representado como profundidade e com a estrutura nativa (N) no fundo (ponto de menor energia-livre). O funil para determinada proteína pode ter uma variedade de formas, dependendo do número e dos tipos de intermediários de enovelamento nas vias de enovelamento. Qualquer intermediário de enovelamento com estabilidade significativa e um tempo de vida finita estaria representado como um mínimo de energia-livre local – uma depressão na superfície do funil. **(a)** Um funil simples, mas relativamente amplo e suave representa uma proteína com múltiplas vias de enovelamento (ou seja, a ordem em que as partes distintas da proteína se dobram seria de alguma forma aleatória), mas que assume suas estruturas tridimensionais sem

intermediários de enovelamento com estabilidade significativa. **(b)** Este funil representa uma proteína mais comum com múltiplos intermediários de enovelamento possíveis com estabilidade significativa nas várias vias que levam à estrutura nativa. **(c)** Uma proteína com estrutura nativa estável, essencialmente sem intermediários de enovelamento com estabilidade significativa, e apenas uma ou muito poucas vias de enovelamento produtivo está mostrada como um funil com uma depressão estreita levando à forma nativa. **(d)** Uma proteína com intermediários de enovelamento com estabilidade substancial em praticamente qualquer via que leva ao estado nativo (ou seja, uma proteína na qual um motivo ou domínio particular sempre se dobra rapidamente, mas as outras partes da proteína se dobram mais lentamente e em ordem aleatória) está representada por um funil com uma depressão principal ao redor da depressão que leva à forma nativa.

ção “proteínas de choque térmico”, do inglês *heat shock proteins*, de  $M_r$  70.000, ou Hsp70). As proteínas Hsp70 se ligam a regiões não dobradas do peptídeo rico em resíduos hidrofóbicos. Dessa forma, essas chaperonas “protegem” da desnaturação pela temperatura tanto as moléculas de proteína quanto as novas moléculas de peptídeo que estão sendo sintetizadas (e ainda não estão dobradas). As proteínas Hsp70 também bloqueiam o enovelamento de certas proteínas que devem permanecer não dobradas até que sejam deslocadas através da membrana (como descrito no Capítulo 27). Algumas chaperonas também facilitam o arranjo quaternário de proteínas oligoméricas. As proteínas Hsp70 se ligam aos polipeptídeos e os liberam em um ciclo que utiliza a energia da hidrólise do ATP e envolve diversas outras proteínas (incluindo uma classe chamada de Hsp40). A **Figura 4-30** ilustra um enovelamento assistido por chaperona, como o elucidado para as chaperonas Hsp70 e Hsp40 eucarióticas. A ligação de um



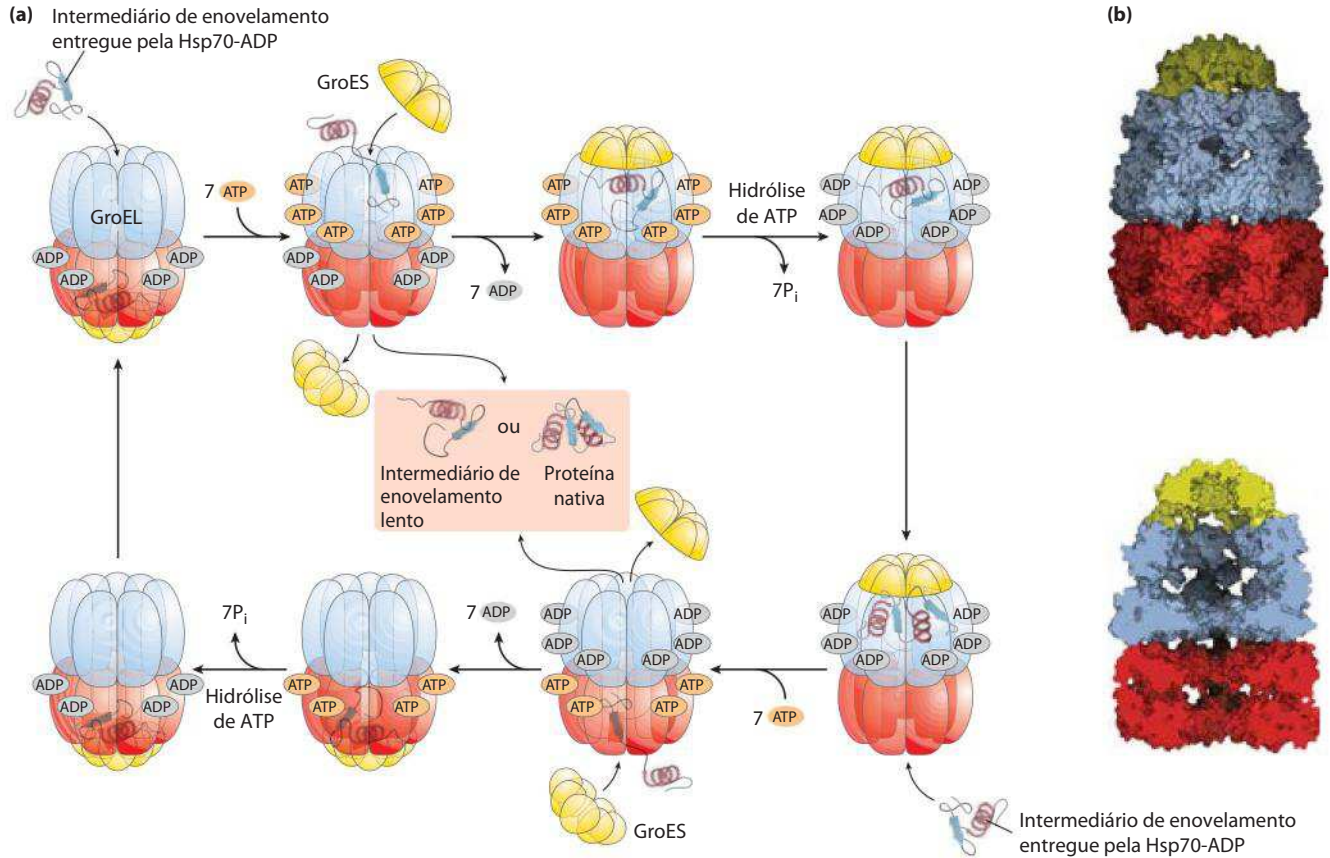
**FIGURA 4-30** As chaperonas no enovelamento proteico. A rota pela qual as chaperonas da classe Hsp70 se ligam aos polipeptídeos e os liberam é ilustrada pelas chaperonas eucarióticas Hsp70 e Hsp40. As chaperonas não promovem diretamente o enovelamento da proteína, mas evitam a formação de agregados de peptídeos não dobrados. As proteínas não dobradas ou parcialmente dobradas primeiro se ligam à forma aberta da Hsp70 ligada ao ATP (PDB ID 2QXL). A Hsp40 então interage com esse complexo e induz a hidrólise do ATP produzindo a forma fechada do complexo (derivado de PDB ID 2KHO e 1DKZ), na qual os domínios coloridos em cor de laranja e amarelo estão juntos como as duas partes de uma mandíbula, mantendo presas partes da proteína desdobrada. A dissociação do ADP e a reciclagem de Hsp70 requer a interação com outra proteína, um fator de troca de nucleotídeo (NEF). Em uma população de moléculas de polipeptídeos, alguma fração das moléculas liberadas após a ligação transitória das proteínas parcialmente dobradas pela Hsp70 assumirá a conformação nativa. As remanescentes são novamente ligadas pela Hsp 70 ou desviadas para o sistema das chaperoninas (Hsp60; ver Figura 4-31). Em bactérias, as chaperonas Hsp70 e Hsp40 são chamadas DnaK e DnaJ, respectivamente. DnaK e DnaJ foram primeiro identificadas como proteínas necessárias para a replicação *in vitro* de certas moléculas de DNA viral (por isso a designação “Dna”).

peptídeo não dobrado à chaperona Hsp70 pode quebrar um agregado proteico ou prevenir a formação de um novo agregado. Quando o peptídeo ligado é liberado, ele tem a chance de retomar o enovelamento à sua estrutura nativa. Se o enovelamento não ocorrer rápido o suficiente, o polipeptídeo pode ser ligado de novo e o processo se repete. Alternativamente, o polipeptídeo ligado a Hsp70 pode ser entregue a uma chaperonina.

Chaperoninas são complexos elaborados de proteínas necessários para o enovelamento de algumas proteínas celulares que não se doam espontaneamente. Em *E. coli*, estima-se que de 10 a 15% das proteínas celulares necessitam do sistema de chaperoninas residentes, chamado de GroEL/GroES, para o enovelamento nas condições normais (mais de 30% precisam de assistência quando as células são submetidas ao estresse por calor). Em eucariotos, o sistema de chaperoninas análogo chama-se Hsp60. As chaperoninas se tornaram conhecidas quando foram consideradas necessárias para o crescimento de certos vírus bacterianos (por isso, a designação “Gro”, de *growth*). Essa família de proteínas é estruturada como uma série de anéis de múltiplas subunidades, formando duas câmaras orientadas uma de costas para a outra. Uma proteína desdobrada primeiro se liga a uma superfície hidrofóbica exposta próxima a extremidade apical de uma câmara GroEL. A proteína é então presa dentro da câmara, fechada transitóriamente, pela “tampa” GroES (**Figura 4-31**). A GroEL sofre mudanças conformacionais substanciais, associadas à hidrólise lenta de ATP que também regula a ligação e a liberação de GroES. Dentro da câmara, uma proteína tem cerca de 10 segundos para se dobrar – o tempo necessário para a hidrólise do ATP ligado. A restrição de uma proteína dentro da câmara previne a agregação proteica inapropriada e também restringe o espaço conformacional que a cadeia polipeptídica pode explorar à medida que ela se enovela. A proteína é liberada quando GroES se dissocia, mas pode religar rapidamente para outro ciclo se o enovelamento não tiver sido completo. As duas câmaras em um complexo GroEL se alternam na ligação e liberação dos substratos polipeptídicos. Em eucariotos, o sistema Hsp60 utiliza um sistema semelhante para o enovelamento de proteínas. No entanto, em vez da tampa GroES, protrusões dos domínios apicais das subunidades se flexionam e fecham a câmara. O ciclo hidrolítico do ATP também é mais lento nos complexos Hsp60, dando mais tempo à proteína retida na câmara para o enovelamento.

Finalmente, as rotas de enovelamento de algumas proteínas requerem duas enzimas que catalisam reações de isomerização. A **proteína dissulfeto-isomerase (PDI)** é uma enzima amplamente distribuída que catalisa a interconversão, ou modificação, de ligações dissulfeto até que as ligações da conformação nativa sejam formadas. Entre suas funções, a PDI catalisa a eliminação de intermediários dobrados com ligações dissulfeto transversais inadequadas. A **peptídeo-prolil-*cis-trans*-isomerase (PPI)** catalisa a interconversão de isômeros *cis* e *trans* de ligações peptídicas de resíduos de Pro (Figura 4-8), que pode representar um passo lento no enovelamento de proteínas que contêm algumas ligações peptídicas com Pro na conformação *cis*.






**FIGURA 4-31 As chaperoninas no enovelamento de proteínas. (a)** Mecanismo de ação proposto para as chaperoninas GroEL (membro da família de proteínas Hsp60) e GroES de *E. coli*. Cada complexo GroEL consiste em dois anéis heptaméricos que formam duas câmaras grandes (cada subunidade com  $M_r$  57.000). A GroES, também um heptâmero (subunidades de  $M_r$  10.000), bloqueia a entrada de uma das câmaras na GroEL depois da ligação, dentro da câmara, de uma proteína desdobrada. A câmara com a proteína desdobrada é designada como *cis*; o oposto é designado *trans*. O enove-

lamento ocorre no interior da câmara *cis*, durante o tempo que leva para hidrolisar os 7 ATP ligados às subunidades no anel heptamérico. As moléculas de GroES e de ADP então se dissociam, e a proteína é liberada. As duas câmaras do sistema GroEL/Hsp60 alternam a ligação e facilitam o enovelamento das proteínas clientes. **(b)** Imagens da superfície e corte longitudinal do complexo GroEL/GroES (PDB ID 1AON). O corte (abaixo) ilustra o grande espaço interno no qual as outras proteínas se ligam.

### Defeitos no enovelamento proteico fornecem a base molecular para uma ampla gama de doenças genéticas em humanos

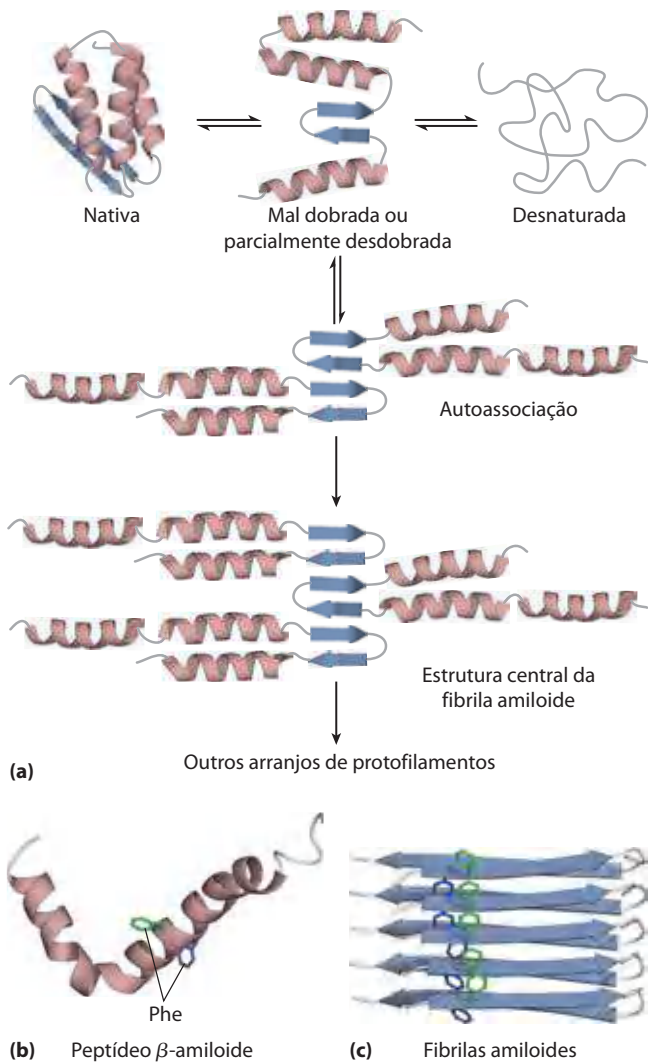
 Apesar da participação de muitos processos que assistem o enovelamento proteico, erros podem ocorrer. De fato, o enovelamento proteico errado é um problema substancial para a célula, e um quarto ou mais de todos os polipeptídeos sintetizados é destruído por não se dobrar corretamente. Em alguns casos, os erros de enovelamento causam ou contribuem para o desenvolvimento de doenças graves.

Muitas condições, incluindo diabetes do tipo 2, doença de Alzheimer, doença de Huntington e doença de Parkinson, estão associadas com um mecanismo de enovelamento errôneo: uma proteína solúvel normalmente secretada pela célula é secretada em um estado de enovelamento errado, sendo convertida em uma fibra **amiloide** extracelular insolúvel. As doenças são coletivamente chamadas de **amiloidoses**. As fibras são altamente ordenadas e não ramificadas, com um diâmetro de 7 a 10 nm e alto conteúdo de estruturas do tipo folha  $\beta$ . Os segmentos  $\beta$  são orientados de forma perpendicular ao eixo da fibra. Em algumas fibras amiloides, a estrutura

geral apresenta duas camadas de folhas  $\beta$  longas, como aquela mostrada para o peptídeo  $\beta$ -amiloide, na **Figura 4-32**.

Muitas proteínas podem assumir a estrutura da fibrila amiloide como alternativa para sua conformação dobrada normal, e a maioria dessas proteínas tem grande concentração de resíduos de aminoácidos aromáticos em uma região central das folhas  $\beta$  ou hélice  $\alpha$ . As proteínas são secretadas em uma conformação não totalmente dobrada. O núcleo (ou parte dele) se dobra em uma folha  $\beta$  antes que o resto da proteína se dobre corretamente, e as folhas  $\beta$  de duas ou mais moléculas de proteínas com enovelamento incompleto se associam para começar a formação de uma fibrila amiloide. A fibrila cresce no espaço extracelular. Outras partes da proteína, então, se dobras de forma diferente, ficando externas ao núcleo de folhas  $\beta$  da fibrila em formação. O efeito dos aminoácidos aromáticos na estabilização da estrutura é mostrado na Figura 4-32c. A maioria das moléculas proteicas se dobra corretamente; por isso, o surgimento dos sintomas das amiloidoses geralmente é muito lento. Se uma pessoa herda uma mutação, como a substituição de um resíduo aromático em uma posição que favorece a formação de fibrilas amiloides, os sintomas da doença podem começar mais cedo.





**FIGURA 4-32** Formação das fibrilas amiloides causadoras de doenças.

(a) Moléculas de proteínas cuja estrutura normal inclui regiões de folhas  $\beta$  passam por enovelamento parcial. Em um pequeno número de moléculas, antes do enovelamento completo, as regiões de folhas  $\beta$  de um polipeptídeo se associam com a mesma região de outro polipeptídeo, formando o núcleo de um amiloide. Outras moléculas de proteína lentamente se associam ao amiloide e o estendem para formar uma fibrila. (b) O peptídeo  $\beta$ -amiloide inicia como dois segmentos  $\alpha$ -helicoidais de uma proteína maior. A clivagem proteolítica desta proteína maior libera o peptídeo  $\beta$ -amiloide relativamente instável, que perde sua estrutura  $\alpha$ -helicoidal. Ele pode, então, lentamente ser organizado em fibrilas amiloides (c), que contribuem para as placas características no exterior do tecido nervoso das pessoas com doença de Alzheimer. As cadeias laterais aromáticas mostradas aqui exercem um papel significativo na estabilização da estrutura amiloide. A estrutura amiloide é rica em folhas  $\beta$ , com as fitas  $\beta$  arranjadas perpendicularmente ao eixo da fibrila amiloide. O peptídeo  $\beta$ -amiloide assume a forma de duas camadas de folhas  $\beta$  paralelas estendidas. Alguns peptídeos formadores de amiloide podem se dobrar na forma de hélices  $\beta$  voltadas para esquerda (ver Figura 4-22).

Em eucariotos, as proteínas destinadas à secreção iniciam seu enovelamento no retículo endoplasmático (RE; conferir via no Capítulo 27). Quando ocorrem condições de estresse ou quando a síntese proteica ameaça sobrecarregar a capacidade de enovelamento proteico do RE, pode ocorrer o acúmulo de proteínas desdobradas. Essas condições disparam a resposta a proteínas não dobradas (UPR,

de *unfolded protein response*). Um conjunto de reguladores transcricionais que constituem a UPR alinha os vários sistemas por aumentar a concentração de chaperonas no RE ou por diminuir a taxa global de síntese proteica, ou ambos. Os agregados amiloides que se formam antes da UPR podem ser removidos. Alguns são degradados por **autofagia**. Nesse processo, eles são primeiro encapsulados em uma membrana, e, então, o conteúdo da vesícula resultante é degradado após a fusão com um lisossomo citosólico. Alternativamente, as proteínas erroneamente dobradas podem ser degradadas por um sistema de proteases chamado sistema ubiquitina-proteossomo (descrito no Capítulo 27). Defeitos em qualquer um desses sistemas diminuem a capacidade de lidar com proteínas erroneamente dobradas e aumentam a propensão para o desenvolvimento de doenças amiloides.

Algumas amiloidoses são sistêmicas, envolvendo vários tecidos. A mais simples é causada pela deposição de fibrilas formadas por cadeias leves de imunoglobulinas com enovelamento incorreto (ver Capítulo 5), ou fragmentos de cadeias leves derivados da degradação proteolítica. A idade média de aparecimento é de 65 anos. Os pacientes têm sintomas como fadiga, rouquidão, inchaço e perda de peso, e muitos morrem dentro de um ano após o diagnóstico. Os rins e o coração geralmente são os órgãos mais afetados. Algumas amiloidoses estão associadas com outros tipos de doenças. Indivíduos com certas infecções crônicas ou doenças inflamatórias como artrite reumatoide, tuberculose, fibrose cística e alguns cânceres podem ter um grande aumento de secreção de um polipeptídeo pró-amiloide chamado de proteína amiloide sérica A (SAA, do inglês *serum amyloid A*). Essa proteína, ou fragmentos dessa proteína, se deposita no tecido conectivo do baço, dos rins, do fígado e ao redor do coração. Indivíduos com essa condição, conhecida como amiloidose sistêmica secundária, apresentam inúmeros sintomas, dependendo dos órgãos inicialmente afetados. A doença geralmente é fatal dentro de poucos anos. Mais de 80 amiloidoses estão associadas com mutações na transtirretina (proteína que se liga aos hormônios da tireoide e os transporta, distribuindo-os pelo corpo e cérebro). Uma grande variedade de mutações nessa proteína leva à deposição amiloide, concentrada ao redor de diferentes tecidos, produzindo assim diferentes sintomas. As amiloidoses também estão associadas com mutações hereditárias nas proteínas lisozima, cadeia  $\alpha$  do fibrinogênio A e apolipoproteínas A-I e A-II, todas descritas nos capítulos posteriores.

Algumas doenças amiloides estão associadas com órgãos específicos. A proteína pró-amiloide geralmente é secreta somente pelos tecidos afetados, e sua alta concentração local leva à deposição de amiloide ao redor do tecido (embora algumas proteínas possam estar sistematicamente distribuídas). Um lugar comum de deposição de amiloide é próximo às células das ilhotas  $\beta$  pancreáticas, responsáveis pela secreção de insulina e regulação do metabolismo de glicose (ver Figura 23-26). A secreção pelas células  $\beta$  de um pequeno peptídeo (37 aminoácidos), chamado de polipeptídeo amiloide das ilhotas (PPAI, ou IAPP, do inglês *islet amyloid polypeptide*), pode levar à deposição de amiloide ao redor das ilhotas, destruindo gradualmente as células. Um adulto saudável tem de 1 a 1,5 milhão de células  $\beta$  pancreáticas. Com a perda progressiva dessas células, a home-

## QUADRO 4-6



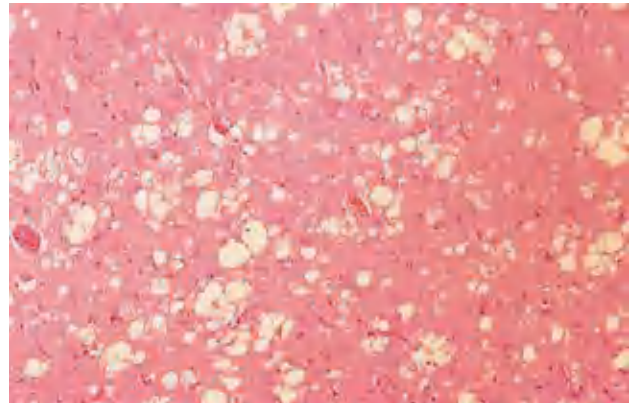
## MEDICINA

## Morte por envelhecimento errado: as doenças priônicas

Uma proteína cerebral dobrada de forma errada parece ser o agente causador de doenças cerebrais neurodegenerativas raras em mamíferos. Talvez a mais conhecida seja a encefalopatia espongiforme bovina (EEB, ou BSE, do inglês *bovine spongiform encephalopathy*; também conhecida como doença da vaca louca). Doenças relacionadas incluem a kuru e a doença de Creutzfeldt-Jakob em humanos, *scrapie* em ovinos, e doença debilitante crônica em cervos e alces. Essas doenças também são conhecidas como encefalopatias espongiformes porque o cérebro doente frequentemente se torna cheio de buracos (Figura Q-1). A deterioração progressiva do cérebro leva a um espectro de sintomas neurológicos, incluindo perda de peso, comportamento errático, problemas de postura, equilíbrio e coordenação, e perda da capacidade cognitiva. Essas doenças são fatais.

Nos anos de 1960, pesquisadores descobriram que amostras de agentes causadores de doença pareciam não conter ácidos nucleicos. Naquela época, Tikvah Alper sugeriu que o agente fosse uma proteína. Inicialmente, a ideia pareceu uma heresia. Todos os agentes causadores de doenças conhecidos até aquele momento – vírus, bactérias, fungos, e assim por diante – continham ácidos nucleicos, e sua virulência estava relacionada à reprodução genética e propagação. Entretanto, quatro décadas de investigações, realizadas mais notavelmente por Stanley Prusiner, forneceram evidências de que as encefalopatias espongiformes são diferentes.

Os agentes infecciosos foram identificados como uma única proteína ( $M_r$  28.000), que Prusiner apelidou de proteína **prion** (PrP). O nome foi derivado de *proteinaceous infectious* (proteína infecciosa), mas Prusiner



**FIGURA Q-1** Secção corada de córtex cerebral, da necropsia de um paciente com a doença de Creutzfeldt-Jakob, mostra a degeneração espongiforme (vacuolar), a característica neuro-histológica mais comum. Os vacúolos amarelados são intracelulares e ocorrem, majoritariamente, em processos pré e pós-sinápticos de neurônios. Os vacúolos nesta secção variam de 20 a 100  $\mu\text{m}$  em diâmetro.

achou que “prion” soava melhor do que “proin”. A proteína prion é um constituinte normal do tecido cerebral em todos os mamíferos. Seu papel não é conhecido em detalhes, mas deve ter uma função de sinalização molecular. Linhagens de camundongos sem o gene para PrP (e, por isso, sem a proteína) sofrem efeitos maléficis não óbvios. A doença ocorre somente quando a PrP celular normal, ou PrP<sup>C</sup>, apresenta uma conformação alterada chamada de PrP<sup>Sc</sup> (Sc significa *scrapie*). A estrutura de PrP<sup>C</sup> caracteriza-se por duas hélices  $\alpha$ . A estrutura de

ostase de glicose é afetada e, no final, quando 50% ou mais células são perdidas, a condição evolui para diabetes melito do tipo 2 (independente de insulina).

As doenças de deposição amiloide que induzem neurodegeneração, especialmente em adultos idosos, são uma classe especial de amiloidoses localizadas. A doença de Alzheimer está associada com deposição amiloide extracelular pelos neurônios, envolvendo o peptídeo amiloide  $\beta$  (Figura 4-32b), derivado de uma grande proteína transmembrana (proteína precursora de amiloide  $\beta$ ) encontrada na maioria dos tecidos humanos. Quando ele faz parte de uma proteína maior, o peptídeo é composto por dois segmentos  $\alpha$ -helicoidais que atravessam a membrana. Quando os domínios externos e internos são clivados por proteases específicas, o peptídeo  $\beta$ -amiloide relativamente instável deixa a membrana e perde sua estrutura  $\alpha$ -helicoidal. Ele pode então tomar a forma de duas camadas de folhas  $\beta$  paralelas estendidas, que lentamente podem se reunir em fibrilas amiloides (Figura 4-32c). O depósito dessas fibras amiloides parece ser a causa primária da doença de Alzheimer, mas um segundo tipo de agregado do tipo amiloide, envolvendo a proteína chamada tau,

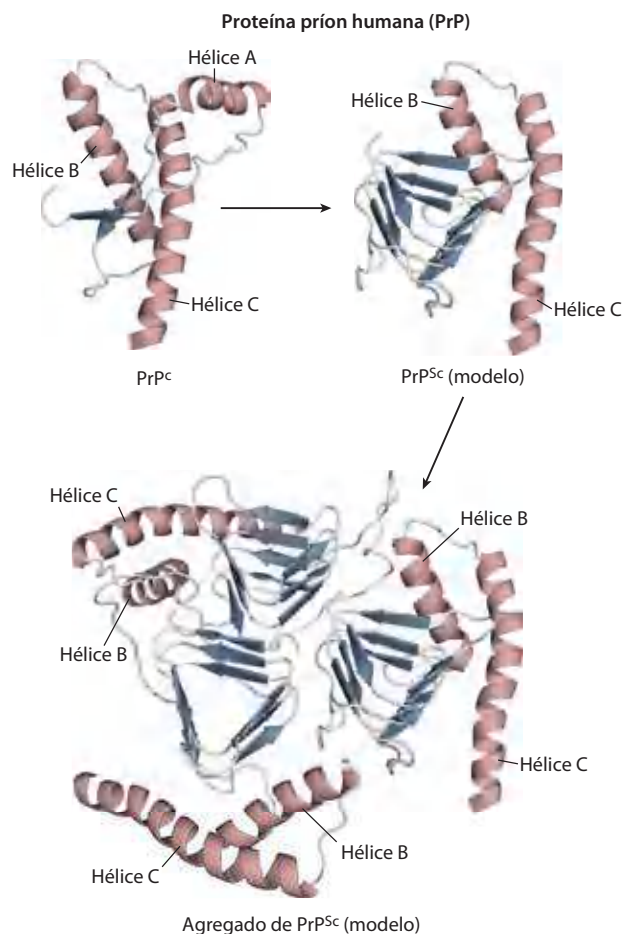
também ocorre intracelularmente (nos neurônios) em pessoas com doença de Alzheimer. Mutações hereditárias na proteína tau não resultam em Alzheimer, mas causam demência frontotemporal e parkinsonismo (condição com sintomas que lembram a doença de Parkinson), que pode ser igualmente devastadora.

Várias outras condições neurodegenerativas envolvem agregação intracelular de proteínas com envelhecimento errado. Na doença de Parkinson, a forma mal dobrada da proteína  $\alpha$ -sinucleína se agrega em massas esféricas filamentosas, chamadas de corpos de Lewy. A doença de Huntington envolve a proteína huntingtina, que tem uma longa repetição de poliglutaminas. Em alguns indivíduos, essa repetição é maior do que o normal, ocorrendo um tipo de agregação intracelular mais sutil. Notavelmente, quando proteínas mutantes humanas envolvidas nas doenças de Parkinson e Huntington são expressas em *Drosophila melanogaster*, as moscas demonstram degeneração expressa como deterioração dos olhos, tremores e morte precoce. Todos esses sintomas são altamente suprimidos se a expressão da chaperona Hsp70 também estiver aumentada.

PrP<sup>Sc</sup> é muito diferente, com a maior parte da estrutura convertida em folhas  $\beta$  semelhante à amiloide (Figura Q-2). A interação da PrP<sup>Sc</sup> com a PrP<sup>C</sup> converte a última em PrP<sup>Sc</sup>, iniciando um efeito dominó no qual cada vez mais proteínas do cérebro se convertem na forma causadora da doença. O mecanismo pelo qual a presença de PrP<sup>Sc</sup> resulta em encefalopatia espongiforme não é conhecido.

A forma hereditária das doenças priônicas, uma mutação no gene que codifica a PrP, produz uma mudança em um resíduo de aminoácido que acredita-se tornar a conversão de PrP<sup>C</sup> em PrP<sup>Sc</sup> mais favorável. Um completo entendimento das doenças priônicas aguarda novas informações de como uma proteína príon afeta as funções cerebrais. Informações estruturais sobre PrP estão começando a fornecer indicativos quanto ao processo molecular que permite às proteínas priônicas interagirem de forma a alterar sua conformação (Figura Q-2).

**FIGURA Q-2** Estrutura do domínio globular de uma PrP humana (PDB ID 1QLX) e dos modelos de PrP<sup>Sc</sup> mal dobrada, na conformação causadora de doença e de um agregado PrP<sup>Sc</sup>. As hélices  $\alpha$  estão marcadas para ajudar a ilustrar a mudança conformacional. A hélice A é incorporada na estrutura da folha  $\beta$  da conformação mal dobrada.



O mal enovelamento proteico não necessariamente significa que a formação de amiloide cause doenças sérias. Por exemplo, a fibrose cística é causada por defeitos em uma proteína de membrana chamada de reguladora transmembrana da fibrose cística (CFTR, do inglês *cystic fibrosis transmembrane conductance*), que atua como canal de íons cloreto. A mutação mais comum que causa a fibrose cística é uma deleção de um resíduo Phe na posição 508 da CFTR, que causa o enovelamento inapropriado da proteína. A maior parte dessa proteína é então degradada, e sua função normal é perdida (ver Quadro 11-2). Muitas das doenças relacionadas com mutações no colágeno (p. 130) também causam um enovelamento defeituoso. Um tipo particular e famoso de enovelamento errado de proteína pode ser visto nas doenças priônicas (Quadro 4-6). ■

#### RESUMO 4.4 Desnaturação e enovelamento das proteínas

▶ A manutenção do estado estável da coleção de proteínas celulares ativas necessárias em um conjunto de condições específicas – chamada proteoestase – envolve um conjunto elaborado de vias e processos que dobram, redobram e degradam cadeias polipeptídicas.

- ▶ A estrutura tridimensional e a função da maioria das proteínas podem ser destruídas pela desnaturação, demonstrando uma relação entre estrutura e função. Algumas proteínas desnaturadas podem renaturar espontaneamente para formar proteínas biologicamente ativas, mostrando que as estruturas terciárias são determinadas pela sequência de aminoácidos.
- ▶ O enovelamento de proteínas nas células é geralmente hierárquico. Inicialmente, regiões de estrutura secundária podem se formar, o que é seguido pelo enovelamento em motivos e domínios. Grandes conjuntos de intermediários dobrados são rapidamente conduzidos a uma única conformação nativa.
- ▶ Para muitas proteínas, o enovelamento é facilitado por chaperonas Hsp70 e por chaperoninas. A formação de ligações dissulfeto e a isomerização *cis-trans* de ligações peptídicas contendo Pro são catalisadas por enzimas específicas.

▶ O enovelamento errado é a base molecular de uma grande variedade de doenças humanas, incluindo as amiloidoses.



## Termos-chave

Os termos em **negrito** estão definidos no glossário.

<b>conformação</b> 115	<b>motivo</b> 137
<b>conformação nativa</b> 116	<b>enovelamento</b> 137
<b>interações</b>	<b>domínio</b> 137
<b>hidrofóbicas</b> 116	família proteica 140
camada de solvatação 116	multímero 140
grupo peptídico 118	<b>oligômero</b> 140
diagrama de	<b>protômero</b> 140
Ramachandran 119	<b>proteínas intrinsecamente</b>
<b>estrutura secundária</b> 119	<b>desordenadas</b> 141
<b>hélice <math>\alpha</math></b> 120	<b>proteostase</b> 143
<b>conformação <math>\beta</math></b> 123	<b>desnaturação</b> 143
folha $\beta$ 123	<b>renaturação</b> 144
<b>volta <math>\beta</math></b> 123	<b>chaperonas</b> 146
<b>espectroscopia de</b>	Hsp70 146
<b>dicroísmo circular</b>	<b>chaperonina</b> 146
<b>(CD)</b> 125	proteína dissulfeto-
<b>estrutura terciária</b> 125	-isomerase (PDI) 147
<b>estrutura quaternária</b> 125	peptidil-prolil- <i>cis-trans</i> -
<b>proteínas fibrosas</b> 125	-isomerase (PPI) 147
<b>proteínas globulares</b> 125	amilóide 148
$\alpha$ -Queratina 126	<b>amiloidoses</b> 148
colágeno 127	<b>autofagia</b> 149
fibroína da seda 130	príon 150
<b>Protein Data Bank</b>	
<b>(PDB)</b> 132	

## Leituras adicionais

### Geral

**Anfinsen, C.B.** (1973) Principles that govern the folding of protein chains. *Science* **181**, 223-230.

O autor revisa seu trabalho clássico sobre ribonuclease.

**Creighton, T.E.** (1993) *Proteins: Structure and Molecular Properties*, 2<sup>nd</sup> edn, W. H. Freeman and Company, New York. Fonte abrangente e confiável.

**Kendrew, J.C.** (1961) The three-dimensional structure of a protein molecule. *Sci. Am.* **205** (December), 96-111.

Descreve como a estrutura da mioglobina foi determinada e o que foi descoberto a partir dela.

**Richardson, J.S.** (1981) The anatomy and taxonomy of protein structure. *Adv. Protein. Chem.* **34**, 167-339.

Excelente resumo sobre padrões e princípios estruturais de proteínas. O autor iniciou a representação em "fitas", amplamente utilizada para estruturas de proteínas.

### Estruturas secundária, terciária e quaternária

**Beeby, M., O'Connor, B.D., Ryttersgaard, C., Boutz, D.R., Perry, L.J. & Yeates, T.O.** (2005) The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol.* **3**, e 309.

**Brown, J.H.** (2006) Breaking symmetry in protein dimmers: designs and function. *Protein Sci* **15**, 1-13.

**Dunker, A.K. & Kriwacki, R.W.** (2011) The orderly chaos of proteins. *Sci. Am.* **304** (April), 68-73.

Um bom resumo do trabalho sobre proteínas que perderam sua estrutura intrínseca.

**Herráez, A.** (2006) Biomolecules in the computer. *Biochem. Mol. Biol. Educ.* **34**, 255-261.

**McPherson, A.** (1989) Macromolecular crystals. *Sci. Am.* **260** (march), 62-69.

Uma descrição de como macromoléculas como as proteínas são cristalizadas.

**Milner-White, E.J.** (1997) The partial charge of the nitrogen atom in peptide bonds. *Protein Sci.* **6**, 2477-2482.

**Ponting, C.P. & Russel, R.R.** (2002) The natural history of protein domains. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 45-71.

Uma explicação de como bancos de dados estruturais podem ser utilizados para explorar a evolução.

### Desnaturação e enovelamento de proteínas

**Chiti, F & Dobson, C.M.** (2006) Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366.

**Dill, K.A., Ozkan, S.B., Shell, M.S., & Weikl, T.R.** (2008) The protein folding problem. *Annu. Rev. Biophys.* **37**, 289-316.

**Grazit, E.** (2005) Mechanisms of amyloid fibril self-assembly and inhibition. *FEBS J.* **272**, 5971-5978.

**Hartl, F.U., Bracher, A., & Hayer-Hartl, M.** (2011) Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324-332.

**Hoppener, J.W.M. & Lips, C.J.M.** (2006) Role of islet amyloid in type 2 diabetes mellitus. *Int. J. Biochem. Cell Biol.* **38**, 726-736.

**Kapinga, H.H. & Craig, E.A.** (2010) The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell. Biol.* **11**, 579-592.

**Norrby, E.** (2011) Prions and protein-folding diseases. *J. Intern. Med.* **270**, 1-14.

**Prusiner, S.B.** (1995) The prion disease. *Sci. Am.* **272** (January), 48-57.

Um bom resumo da evidência que levou à hipótese do príon.

**Selkoe, D.J.** (2003) Folding proteins in fatal ways. *Nature* **426**, 900-904.

Um bom resumo sobre amiloidoses.

**Tang, Y., Chang, H., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M., Hartl, F., & Hayer-Hartl, M.** (2006) Structural features of the GroEL-GroES nanocage required for rapid folding of encapsulated protein. *Cell* **125**, 903-914.

**Tydemers, J., Mogk, A., & Bukau, B.** (2010) Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.* **11**, 777-788.

## Problemas

**1. Propriedades da ligação peptídica.** Nos estudos de raios X de peptídeos cristalizados, Linus Pauling e Robert Corey constataram que a ligação C—N da ligação peptídica tem um comprimento intermediário (1,32 Å) entre uma ligação simples C—N (1,49 Å) e uma ligação dupla C=N (1,27 Å). Também observaram que a ligação peptídica é planar (todos os quatro átomos ligados ao grupo C—N estão no mesmo plano), e que os dois átomos de carbono ligados ao C—N são sempre *trans* um em relação ao outro (em lados opostos da ligação peptídica).

(a) O que o comprimento da ligação C—N da ligação peptídica indica sobre sua força e sua ordem de ligação (isto é, se ela é simples, dupla ou tripla)?

(b) O que as observações de Pauling e Corey nos dizem a respeito da facilidade de rotação sobre o eixo C—N da ligação peptídica?



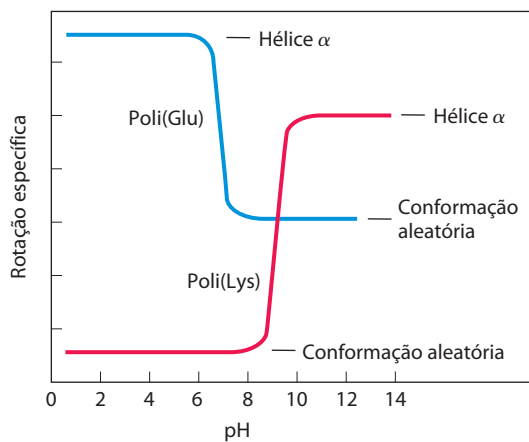
**2. Relações estruturais e funcionais em proteínas fibrosas.** William Astbury descobriu que o padrão de difração de raios X da lã mostra uma unidade estrutural repetida espaçada de 5,2 Å ao longo do comprimento da fibra de lã. Quando ele fervia e esticava a lã, o padrão de raios X mostrava uma nova unidade estrutural que se repetia, com um espaçamento de 7,0 Å. Ferver e esticar a lã e depois deixar que ela se retraia novamente resulta em um padrão de raios X consistente com o espaçamento original de 5,2 Å. Apesar de as observações fornecerem pistas importantes sobre a estrutura molecular da lã, Astbury não foi capaz de interpretá-las naquela época.

(a) Levando em conta o atual conhecimento sobre a estrutura da lã, interprete as observações de Astbury.

(b) Quando blusões e meias de lã são lavados em água quente ou aquecidos em uma máquina de secar, eles encolhem. Por outro lado, a seda não encolhe em condições iguais. Explique.

**3. Velocidade de síntese da  $\alpha$ -queratina do cabelo.** O cabelo cresce com a velocidade de 15 a 20 cm/ano. Todo esse crescimento está concentrado na base da fibra de cabelo, onde os filamentos de  $\alpha$ -queratina são sintetizados dentro das células vivas da epiderme, sendo arranjados em estruturas com formato de cordas (ver Figura 4-11). O elemento estrutural principal da  $\alpha$ -queratina é a hélice  $\alpha$ , que tem 3,6 resíduos de aminoácidos por volta e avança 5,4 Å por volta (ver Figura 4-4a). Assumindo que a biossíntese de cadeias de queratina  $\alpha$ -helicoidal é o fator limitante da velocidade de crescimento do cabelo, calcule a velocidade na qual as ligações peptídicas das cadeias de  $\alpha$ -queratina devem ser sintetizadas (ligações peptídicas por segundo) para justificar o crescimento anual observado para o cabelo.

**4. Efeito do pH na conformação das estruturas secundárias  $\alpha$ -helicoidais.** O desdobraimento da hélice  $\alpha$  de um polipeptídeo para formar uma estrutura enrolada aleatoriamente é acompanhado por um grande decréscimo em uma propriedade chamada de rotação específica, que mede a capacidade de uma solução de desviar a luz circularmente polarizada. O poliglutamato, polipeptídeo formado somente de resíduos L-Glu, tem uma conformação  $\alpha$ -helicoidal em pH 3. Quando o pH sobe para 7, há uma grande perda da rotação específica da solução. Da mesma forma, a polilisina (resíduos L-Lys) forma uma hélice  $\alpha$  em pH 10, mas, quando o pH é diminuído para 7, a rotação específica também diminui, como mostrado no gráfico a seguir.



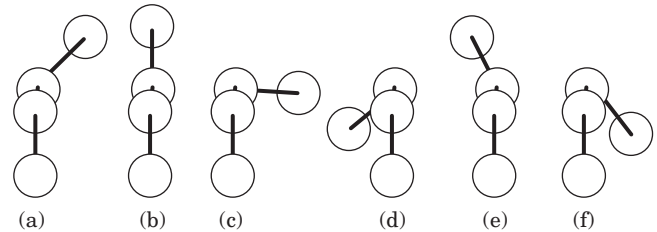
Qual é a explicação para o efeito da mudança de pH nas conformações da poli(Glu) e da poli(Lys)? Por que a transição ocorre em uma faixa tão estreita de pH?

**5. Ligações dissulfeto determinam as propriedades de diversas proteínas.** Algumas proteínas naturais são ricas em ligações dissulfeto, e suas propriedades mecânicas (tensão elástica, viscosidade, dureza, etc.) estão correlacionadas com o grau de ligações dissulfeto.

(a) A glutenina, proteína do trigo rica em ligações dissulfeto, é responsável pelo caráter aderente e elástico da massa feita com farinha de trigo. Similarmente, a natureza dura e resistente do casco da tartaruga é devida às inúmeras ligações dissulfeto de sua  $\alpha$ -queratina. Qual é a base molecular para a correlação entre o conteúdo de ligações dissulfeto e as propriedades mecânicas da proteína?

(b) A maioria das proteínas globulares é desnaturada e perde sua atividade quando aquecida brevemente a 65°C. Entretanto, as proteínas globulares que contêm múltiplas ligações dissulfeto geralmente têm que ser aquecidas por mais tempo e a temperaturas mais altas para que desnaturem. Uma dessas proteínas é o inibidor de tripsina pancreática bovina (BPTI, de *bovine pancreatic trypsin inhibitor*), que tem 58 resíduos de aminoácidos em uma única cadeia e contém três ligações dissulfeto. No resfriamento de uma solução de BPTI desnaturada, a atividade da proteína é restaurada. Qual é a base molecular para essa propriedade?

**6. Ângulos diedros.** Uma série de ângulos de torção,  $\phi$  e  $\psi$ , que podem ser adotados pelo esqueleto peptídico está mostrada abaixo. Qual desses corresponde ao  $\phi$  e  $\psi$  para uma hélice tripla de colágeno ideal? Consulte a Figura 4-9 como guia.



**7. Sequência de aminoácidos e estrutura de proteínas.** O crescente entendimento de como as proteínas se doam permite que os pesquisadores façam previsões sobre a estrutura de uma proteína, com base em dados sobre sua sequência primária de aminoácidos. Considere a seguinte sequência de aminoácidos:

1 2 3 4 5 6 7 8 9 10  
 Ile-Ala-His-Thr-Tyr-Gly-Pro-Phe-Glu-Ala-  
 11 12 13 14 15 16 17 18 19 20  
 Ala-Met-Cys-Lys-Trp-Glu-Ala-Gln-Pro-Asp-  
 21 22 23 24 25 26 27 28  
 Gly-Met-Glu-Cys-Ala-Phe-His-Arg

(a) Onde devem ocorrer dobras ou voltas  $\beta$ ?

(b) Onde devem se formar ligações dissulfeto intramoleculares?

(c) Assumindo que esta sequência é parte de uma proteína globular maior, indique a localização provável (na superfície externa ou no interior da proteína) dos seguintes resíduos de aminoácido: Asp, Ile, Thr, Ala, Gln, Lys. Explique sua resposta. (Dica: observe o índice de hidropatia na Tabela 3-1.)

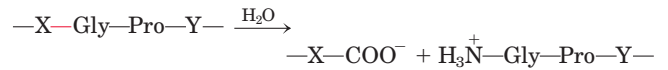
**8. A bacteriorrodopsina em proteínas púrpuras de membrana.** Sob condições ambientais adequadas, a arqueobactéria *Halobacterium halobium* sintetiza uma proteína de membrana ( $M_r$  26.000) conhecida como bacteriorrodopsina, que é púrpura devido a seu conteúdo em retinal (ver Figura 10-21). Moléculas dessa proteína se agregam em “manchas púrpuras” na membrana celular. A bacteriorrodopsina atua como bomba de prótons ativada pela luz, que fornece energia para as funções da célula. A análise por raios X dessa proteína revela que ela consiste em sete segmentos  $\alpha$ -helicoidais paralelos, cada um deles atravessando a membrana celular da bactéria (espessura de 45 Å). Calcule o número mínimo de resíduos de aminoácidos necessários para um segmento de hélice  $\alpha$  atravessar completamente a membrana. Estime a fração proteica da bacteriorrodopsina que está envolvida em hélices transmembrana. (Use uma massa média de 110 para os resíduos de aminoácido.)

**9. Terminologia da estrutura de proteínas.** A mioglobina é um motivo, um domínio ou uma estrutura tridimensional completa?

**10. Interpretando os diagramas de Ramachandran.** Examine as duas proteínas marcadas (a) e (b) abaixo. Quais dos dois diagramas de Ramachandran, marcados (c) e (d), é mais provável ser derivado de qual das proteínas? Por quê?

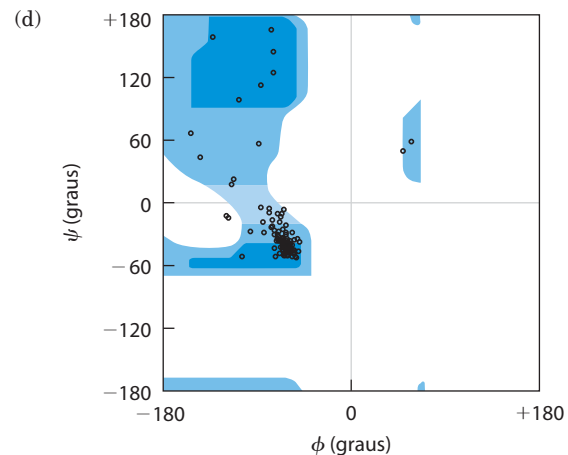
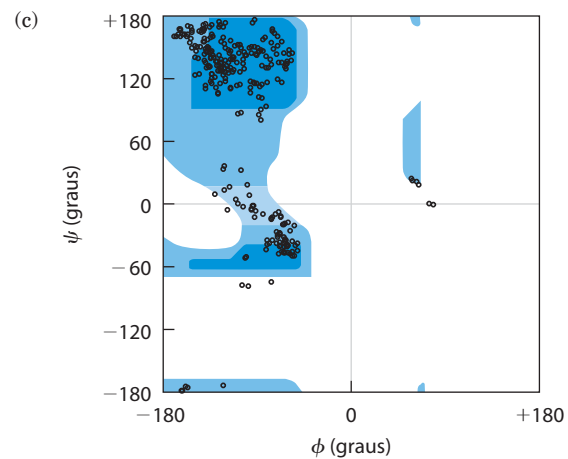
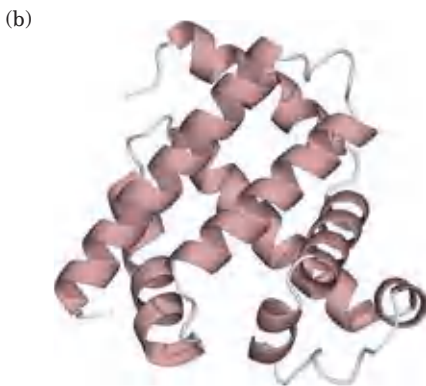
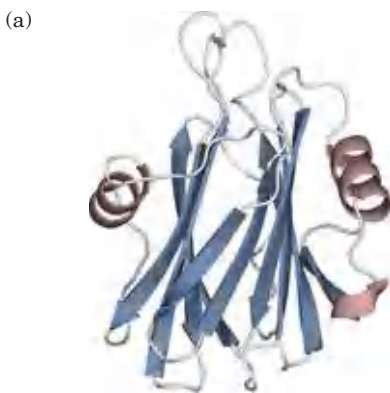
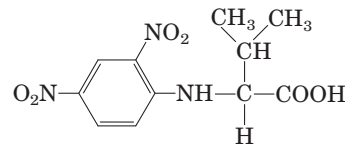
**11. Ação patogênica da bactéria que causa gangrena gasosa.** A bactéria anaeróbia altamente patogênica *Clostridium perfringens* é responsável pela gangrena gasosa, condição na qual o tecido animal tem sua estrutura destruída. Essa bactéria secreta uma enzima que catalisa de maneira eficiente a hidrólise das ligações peptídicas indicadas em vermelho:

*diuim perfringens* é responsável pela gangrena gasosa, condição na qual o tecido animal tem sua estrutura destruída. Essa bactéria secreta uma enzima que catalisa de maneira eficiente a hidrólise das ligações peptídicas indicadas em vermelho:



em que X e Y são qualquer um dos 20 aminoácidos comuns. Como a secreção dessa enzima contribui para a capacidade de invasão desta bactéria nos tecidos humanos? Por que essa enzima não afeta a própria bactéria?

**12. Número de cadeias polipeptídicas em uma proteína com múltiplas subunidades.** Uma amostra (660 mg) de uma proteína oligomérica de  $M_r$  132.000 foi tratada com um excesso de 1-fluoro-2,4-dinitrobenzeno (reagente de Sanger) sob condições levemente alcalinas até que a reação química se completasse. As ligações peptídicas da proteína foram, então, completamente hidrolisadas pelo aquecimento com HCl concentrado. O hidrolisado continha 5,5 mg do seguinte composto:



Derivados 2,4-dinitrofenil dos grupos  $\alpha$ -amino de outros aminoácidos não foram encontrados.

(a) Explique como essa informação pode ser usada para determinar o número de cadeias polipeptídicas na proteína oligomérica.

(b) Calcule o número de cadeias polipeptídicas nesta proteína.

(c) Que outra técnica de análise de proteínas você poderia utilizar para determinar se as cadeias dessa proteína são iguais ou diferentes?

**13. Predição de estrutura secundária.** Qual dos seguintes peptídeos é mais propenso a assumir uma estrutura  $\alpha$ -helicoidal? Por quê?

- (a) LKAENDEAARAMSEA  
(b) CRAGGFWDQPGTNS



**14. Fibras amiloides em doenças.** Diversas pequenas moléculas aromáticas, como o vermelho de fenol (utilizado como modelo de fármaco não tóxico), são conhecidas por inibir a formação de amiloides em sistemas modelo em laboratório. Um objetivo da pesquisa desses pequenos compostos aromáticos é a identificação de um fármaco que iniba de modo eficiente a formação de amiloides no cérebro das pessoas com a doença de Alzheimer incipiente.

(a) Sugira por que moléculas com substituintes aromáticos evitariam a formação de amiloides.

(b) Alguns pesquisadores sugeriram que um fármaco usada para tratar a doença de Alzheimer também poderia ser efetivo no tratamento de diabetes melito do tipo 2 (independente de insulina). Por que um único fármaco poderia ser efetivo no tratamento dessas duas doenças diferentes?

## Bioquímica na internet

**15. Modelagem de proteínas na internet.** Um grupo de pacientes com a doença de Crohn (doença inflamatória do intestino) foi submetido à biópsia das mucosas intestinais para identificar o agente causador da doença. Os pesquisadores identificaram uma proteína que estava presente em níveis mais altos nos pacientes com a doença de Crohn do que em pacientes com outras doenças inflamatórias do intestino, ou no grupo controle. A proteína foi isolada, e a seguinte sequência *parcial* de aminoácidos foi obtida (leia da esquerda para a direita):

EAELCPDRCI	HSFQNLGIQC	VKKRDLEQAI
SQRIQTNNNP	FQVPIEEQRG	DYDLNAVRLC
FQVTVRDPSG	RPLRLPPVLP	HPIFDNRAPN
TAEKICRVN	RNSGSCLGGD	EIFLLCDKVQ
KEDIEVYFTG	PGWEARGSFS	QADVHRQVAI
VFRTPPYADP	SLQAPVRVSM	QLRRPSDREL
SEPMEFQYLP	DTDDRHRIEE	KRKRTYETFK
SIMKKSPPFSG	PTDPRPPRRR	IAVPSRSSAS
VPKPAPQPYP		

(a) Você pode identificar esta proteína utilizando um banco de dados de proteínas da Internet. Alguns bons lugares para começar sua busca incluem o Protein Information Resource (PIR; <http://pir.georgetown.edu>), o Structural Classification of Proteins (SCOP; <http://scop.mrc-lmb.cam.ac.uk/scop>) e o Prosite (<http://prosite.expasy.org>).

No banco de dados selecionado por você, siga as indicações para a ferramenta de comparação de seqüências. Coloque aproximadamente 30 resíduos da seqüência da proteína

no campo de busca adequado, e submeta-o para análise. O que essa análise lhe diz sobre a identidade da proteína?

(b) Tente utilizar diferentes porções da seqüência de aminoácidos. Você sempre obtém o mesmo resultado?

(c) Uma grande variedade de sites da *web* fornece informações sobre a estrutura tridimensional das proteínas. Encontre informações sobre estruturas secundárias, terciárias e quaternárias da proteína utilizando bancos de dados como o Protein Data Bank (PDB; [www.pdb.org](http://www.pdb.org)) ou o SCOP.

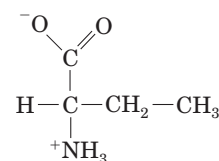
(d) Durante sua busca na internet, o que você aprendeu sobre a função celular da proteína?

## Problema de análise de dados

**16. Proteínas imagem-espaciais.** Como apresentado no Capítulo 3, “os resíduos de aminoácidos em uma molécula de proteína são exclusivamente estereoisômeros L”. Não está claro se essa seletividade é necessária para o funcionamento correto da proteína ou se é um acidente da evolução. Para explorar essa questão, Milton e colaboradores (1992) publicaram um estudo de uma enzima inteiramente formada de estereoisômeros D. A enzima que eles escolheram foi a HIV-protease, enzima proteolítica sintetizada pelo HIV, que converte as pré-proteínas virais inativas em suas formas ativas.

Anteriormente, Wlodawer e colaboradores (1989) relataram a síntese completa da protease do HIV a partir de aminoácidos L (a enzima L), usando o processo mostrado na Figura 3-32. As proteases do HIV normais contêm dois resíduos Cys nas posições 67 e 95. Como a síntese química de proteínas contendo Cys é tecnicamente difícil, Wlodawer e colaboradores substituíram os dois resíduos Cys da proteína pelo aminoácido sintético ácido L- $\alpha$ -amino-*n*-butírico (Aba). Segundo os autores, isso foi feito para “reduzir as dificuldades sintéticas associadas com a desproteção da Cys e para facilitar o manuseio do produto”.

(a) A estrutura do Aba é mostrada abaixo. Por que essa é uma substituição adequada para um resíduo Cys? Sob quais circunstâncias ela não seria adequada?



Ácido L- $\alpha$ -amino-*n*-butírico

Wlodawer e colaboradores desnaturaram a proteína recém-sintetizada por dissolução em 6 M de hidróxido de guanidina, e então permitiram que ela se dobrasse lentamente, retirando a guanidina por diálise contra um tampão neutro (glicerol 10%, 25 mM NaPO<sub>4</sub>, pH 7).

(b) Existem várias razões para prever que uma proteína sintetizada, desnaturada e dobrada dessa maneira não seja ativa. Dê três razões para isso.

(c) De maneira interessante, a L-protease resultante foi ativa. O que essa descoberta lhe informa sobre o papel das ligações dissulfeto na molécula de protease do HIV nativa?

Em seus estudos, Milton e colaboradores sintetizaram a protease do HIV a partir de D-aminoácidos, utilizando o mesmo protocolo do estudo anterior (Wlodawer et al.). Formalmente, existem três possibilidades para o enovelamento da D-protease: resultaria em uma molécula com (1) o mesmo formato da

L-protease; (2) a imagem especular da L-protease; ou (3) alguma outra coisa, possivelmente inativa.

(d) Para cada uma das possibilidades, decida se elas são prováveis ou não e defenda a sua posição.

Na verdade, a D-protease foi ativa: ela hidrolisou um substrato sintético específico e foi inibida por inibidores específicos. Para examinar a estrutura das D e L-enzimas, Milton e colaboradores testaram a atividade das duas formas com as formas D e L de substratos peptídicos quirais, e a capacidade de inibição pelas formas D- e L de peptídeos quirais análogos ao substrato. As duas formas também foram testadas quanto à inibição pelo inibidor aquiral azul de Evans. Os resultados aparecem na tabela.

HIV- -protease	Hidrólise do substrato		Inibição		
	D-substrato	L-substrato	Peptídeo inibidor		Azul de Evans (aquiral)
			D-inibidor	L-inibidor	
L-protease	–	+	–	+	+
D-protease	+	–	+	–	+

(e) Qual dos três modelos propostos é sustentado pelos dados? Explique sua resposta.

(f) Por que o azul de Evans inibe as duas formas de protease?

(g) Você esperaria que a quimotripsina pudesse digerir a D-protease?

(h) Você esperaria, para qualquer enzima, que a síntese total a partir de D-aminoácidos, seguida de renaturação, resultasse em uma enzima ativa? Explique sua resposta.

## Referências

**Milton, R.C., Milton, S.C., & Kent, S.B.** (1992) Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show demonstration of reciprocal chiral substrate specificity. *Science* **256**, 1445-1448.

**Wlodawer, A. Miller, M., Jaskólski, M., Sathyanarayana, B.K., Baldwin, E., Webwe, I.T., Selk, L.M., Clawson, L., Schneider, J., & Kent, S.B.** (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **245**, 616-621.