

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

Proteínas 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.

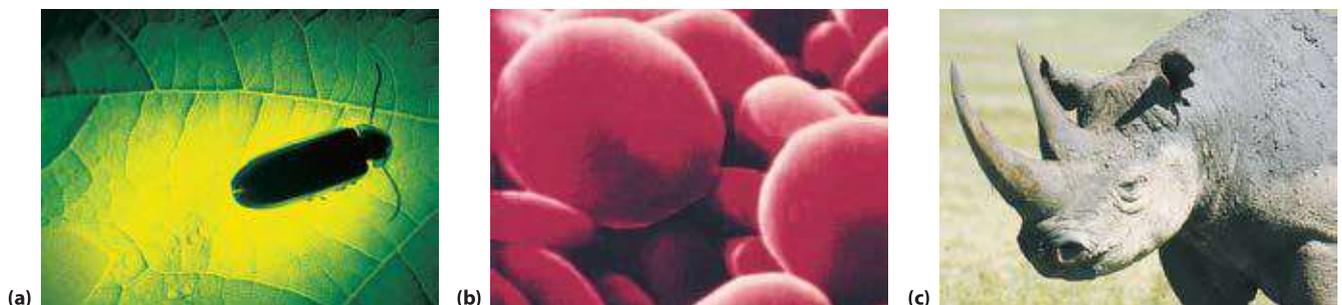


FIGURA 3-1 Algumas funções de proteínas. (a) A luz produzida por vaga-lumes é o resultado de uma reação envolvendo a proteína luciferina e ATP, catalisada pela enzima luciferase (ver Quadro 13-1). (b) Eritrócitos contêm grandes quantidades da proteína transportadora de oxigênio hemoglobina. (c) A proteína 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 α -aminoácidos. Eles têm um grupo carboxila e um grupo amino ligados ao mesmo átomo de carbono (o carbono α) (**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 polymerizados 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.

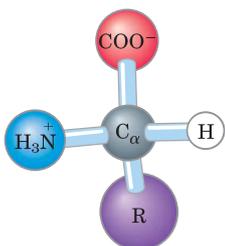


FIGURA 3-2 Estrutura geral de um aminoácido. Esta estrutura é comum a todos os tipos de α -aminoácidos, exceto um (a prolina, aminoácido cíclico, é a exceção). O grupo R, ou cadeia lateral (roxo), ligado ao carbono α (cinza) é diferente em cada aminoácido.



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 α 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 α é, portanto, um **centro quiral** (p. 17). Em decorrência do arranjo tetraédrico dos orbitais de ligação em volta do átomo de carbono α , 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**),

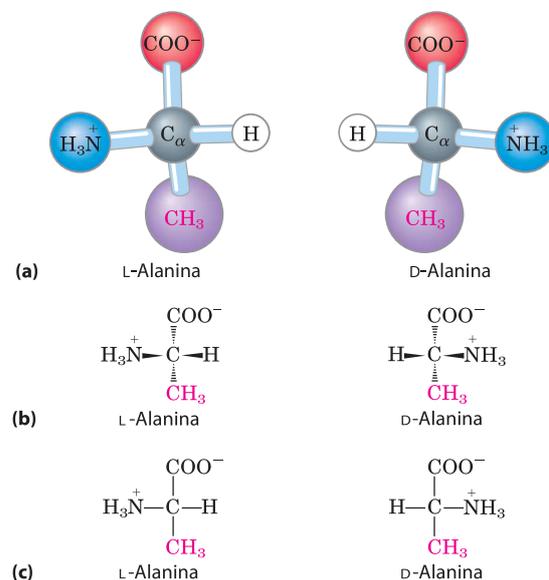


FIGURA 3-3 Estereoisomerismo em α -aminoácidos. (a) Os dois estereoisômeros da alanina, L- e D-alanina, são imagens especulares não sobrepostas um do outro (enantiômeros). (b, c) Duas convenções diferentes para representar as configurações espaciais dos estereoisômeros. Em fórmulas de perspectiva (b), 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 [†]	Ocorrência em proteínas (%) [‡]
			pK_1 (—COOH)	pK_2 (—NH ₃ ⁺)	pK_R (grupo R)			
Grupos R alifáticos, apolares								
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
Grupos R aromáticos								
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
Grupos R polares, não carregados								
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 [¶]	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
Grupos R carregados positivamente								
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
Grupos R carregados negativamente								
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.

[†]Uma escala combinando hidrofobicidade e hidrofiliidade de grupos R. Os valores refletem a energia livre (ΔG) 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 carregadas 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.

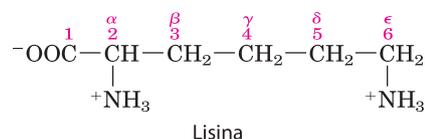
[‡]Ocorrência média em mais de 1.150 proteínas. De Doolittle, R.F. (1989) *Redundancies in protein sequences*. Em *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed.), pp. 599-623, Plenum Press, New York.

[¶]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 β , γ , δ , ϵ , e assim por diante, a partir do carbono α . 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 α 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.

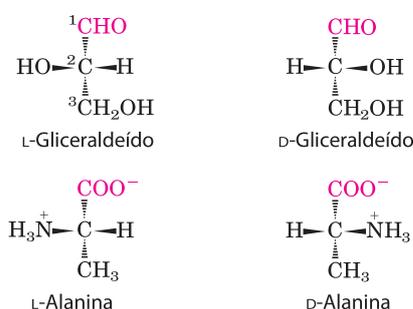


FIGURA 3-4 Relação estérica dos estereoisômeros de alanina à configuração absoluta do L-gliceraldeído e do D-gliceraldeído. Nestas fórmulas em perspectiva, os carbonos são alinhados verticalmente, com o átomo quiral no centro. Os carbonos nestas moléculas são numerados de 1 a 3, de cima para baixo, começando 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 α . Os L-aminoácidos são aqueles com o grupo α -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

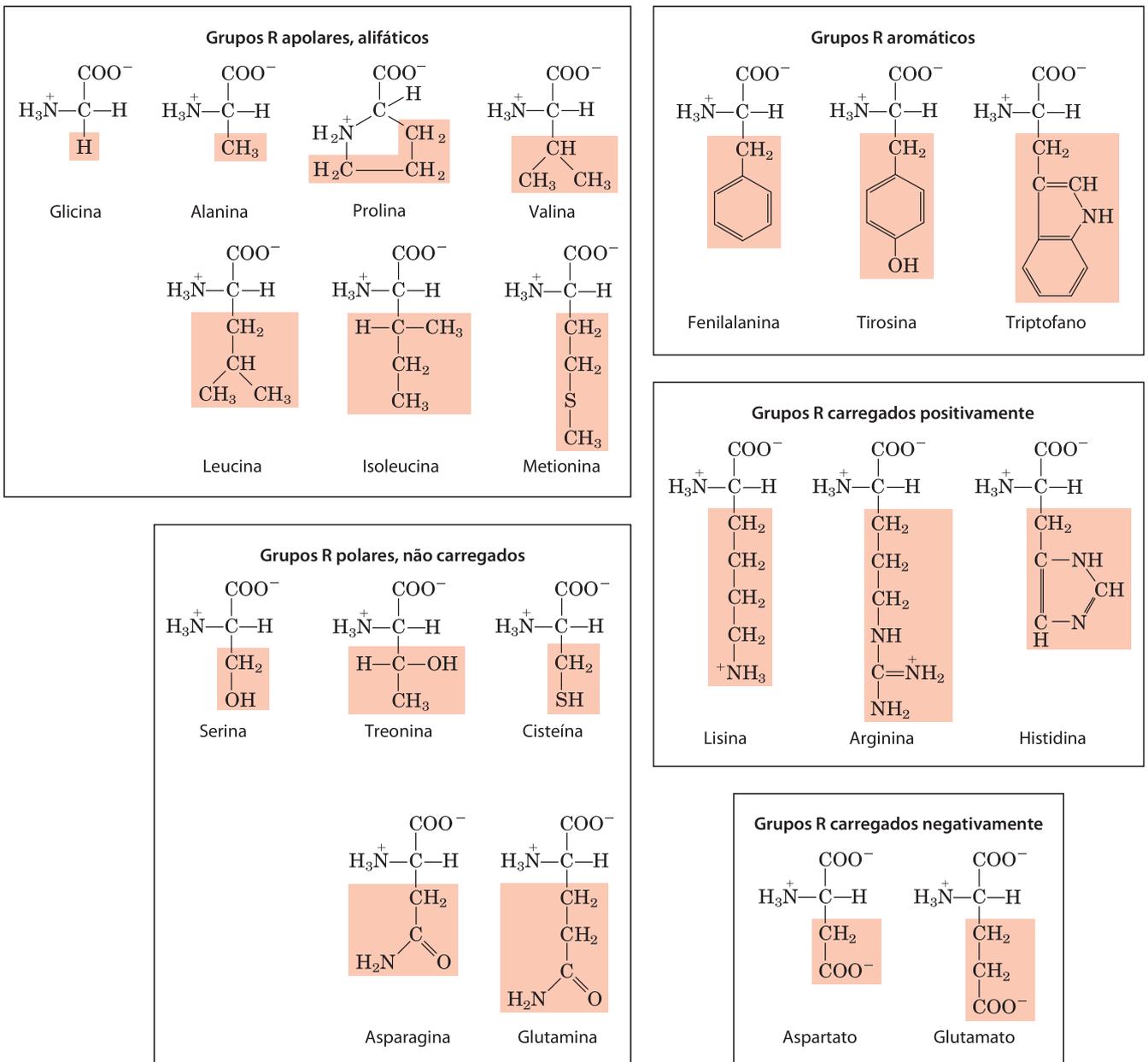


FIGURA 3-5 Os 20 aminoácidos comuns de proteínas. As fórmulas estruturais mostram o estado de ionização que predomina em pH 7,0. As porções não sombreadas são aquelas comuns a todos os aminoácidos; aquelas sombreadas são os grupos R. Embora o grupo R da histidina seja mostrado

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.

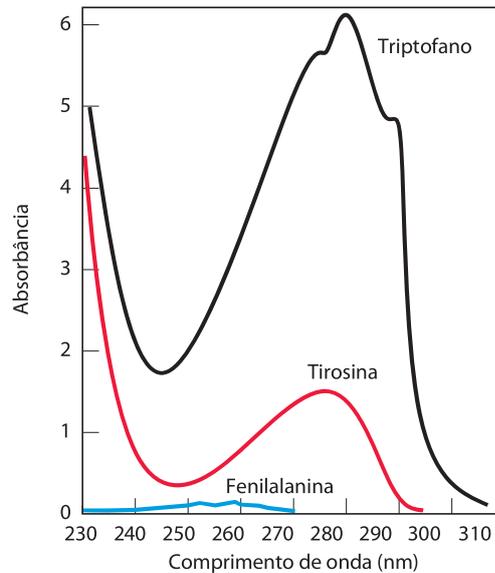
Grupos R aromáticos Fenilalanina, tirosina e triptofano, com suas cadeias laterais aromáticas, são relativamente apolares (hidrofóbicos). Todos podem participar em in-

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, ϵ 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.

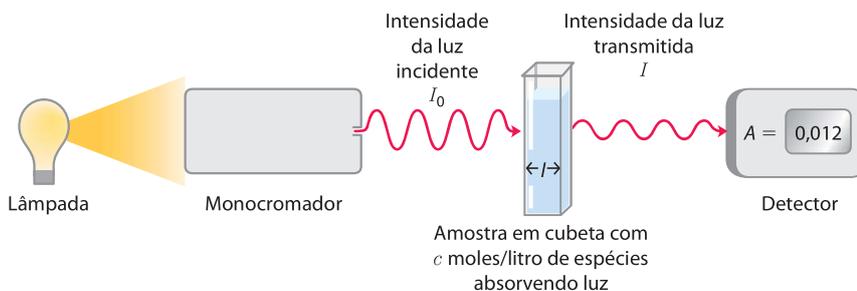


FIGURA Q-1 Os principais componentes de um espectrofotômetro. A fonte de luz emite luz em um amplo espectro, quando o monocromador seleciona e transmite luz de um comprimento de onda específico. A luz monocromática passa através da amostra em uma cubeta de tamanho l e é absorvida pela amostra como uma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

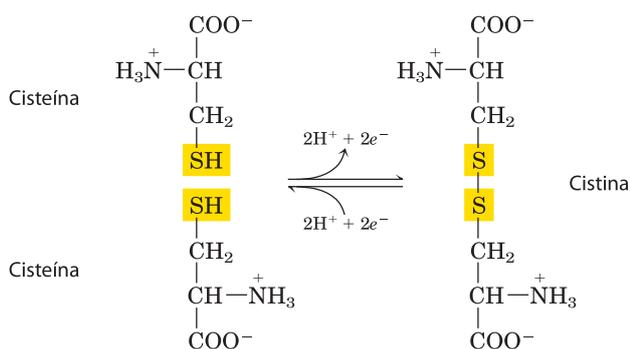


FIGURA 3-7 Formação reversível de uma ligação dissulfeto pela oxidação de duas moléculas de cisteína. Ligações de dissulfeto entre resíduos Cys estabilizam as estruturas de muitas proteínas.

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 ε 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 **γ-carboxiglutamato**, encontrado na proteína de coagulação protrombina e em algumas outras proteínas que se ligam ao Ca²⁺ como parte de suas funções biológicas. Mais complexa é a **desmosina**, derivada de quatro resíduos Lys, encontrada na proteína fibrosa elastina.

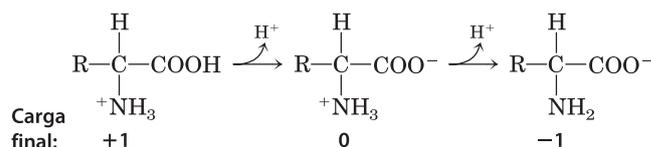
A **selenocisteína** é um caso especial. Esse raro resíduo de aminoácido é introduzido durante a síntese proteica, em vez de criado por uma modificação pós-sintética. Contém selênio em vez do enxofre da cisteína. Na verdade derivada de serina, a selenocisteína é um constituinte de apenas algumas poucas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

Alguns resíduos de aminoácidos em uma proteína podem ser modificados transitóriamente 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 α-aminoácido monoamino monocarboxílico, como a alanina, é um ácido diprótico quando completamente protonado; ele tem dois grupos, o grupo —COOH e o grupo —NH₃⁺, 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.

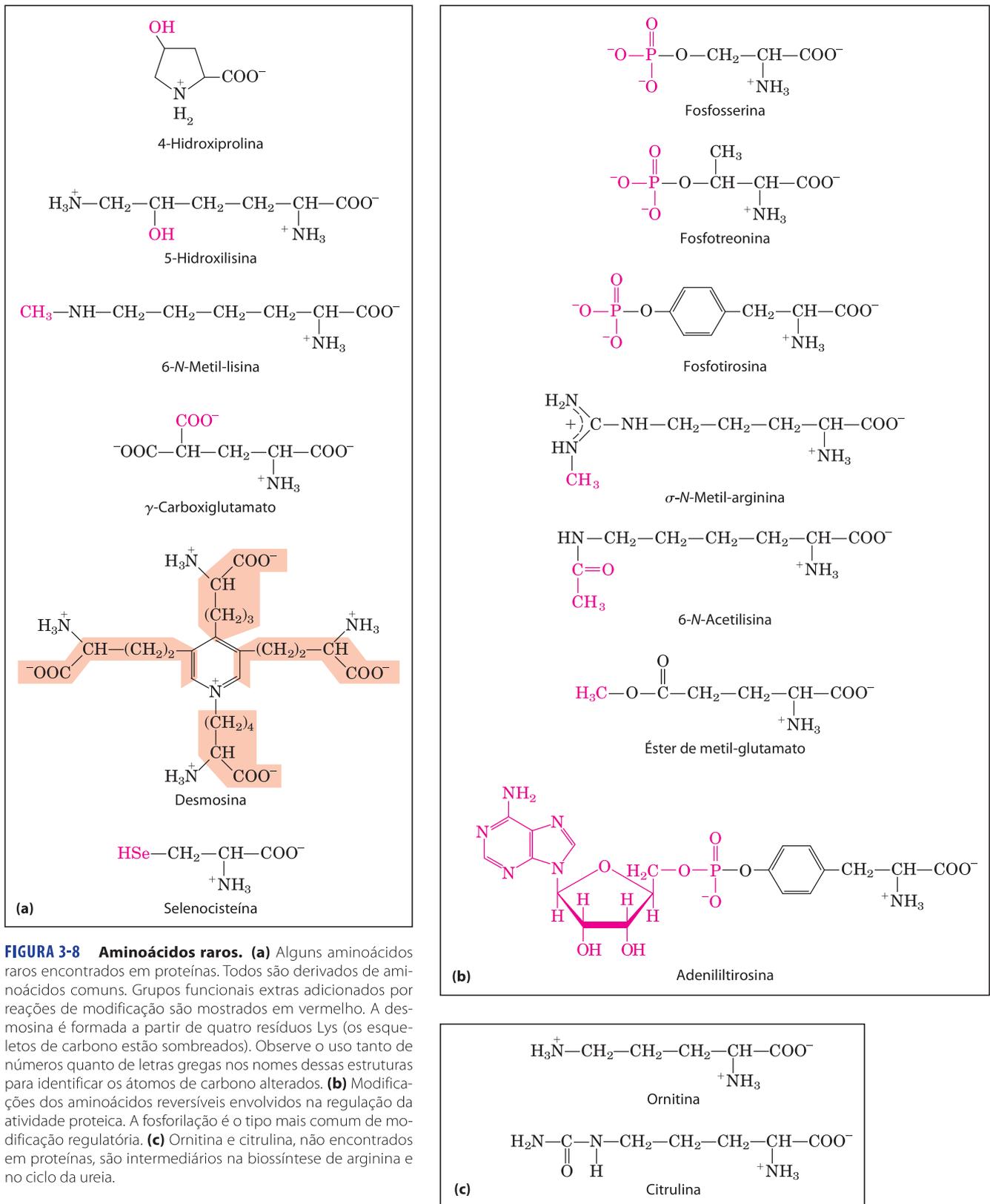


FIGURA 3-8 Aminoácidos raros. (a) Alguns aminoácidos raros encontrados em proteínas. Todos são derivados de aminoácidos comuns. Grupos funcionais extras adicionados por reações de modificação são mostrados em vermelho. A desmosina é formada a partir de quatro resíduos Lys (os esqueletos de carbono estão 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

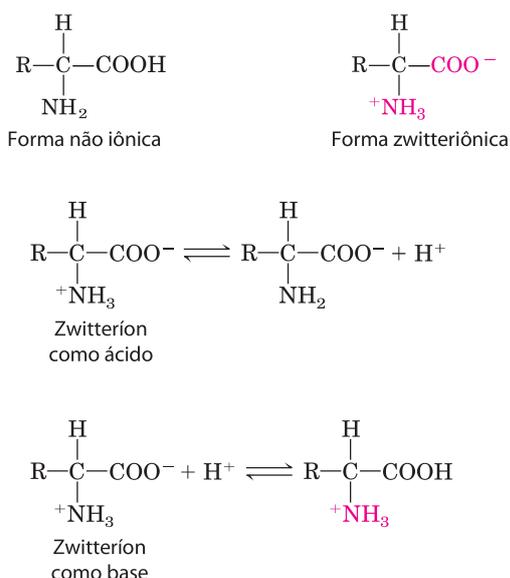


FIGURA 3-9 Formas não iônicas e zwitteriônicas de aminoácidos. A forma não iônica não ocorre em quantidades significativas em soluções aquosas. O zwitterion predomina em pH neutro. Um zwitterion pode atuar tanto como ácido (doador de prótons) quanto como base (aceptor de prótons).

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 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 pK_a (marcado pK_1 na Figura 3-10) de 2,34 (lembre-se do Capítulo 2 que pH e pK_a são simplesmente notações convenientes para concentração de prótons e a constante de equilíbrio para ionização, respectivamente. O pK_a é uma medida da tendência de um grupo doar um próton, com essa tendência diminuindo dez vezes à medida que o 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 pK_a (marcado 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 $\text{H}_2\text{N}-\text{CH}_2-\text{COO}^-$.

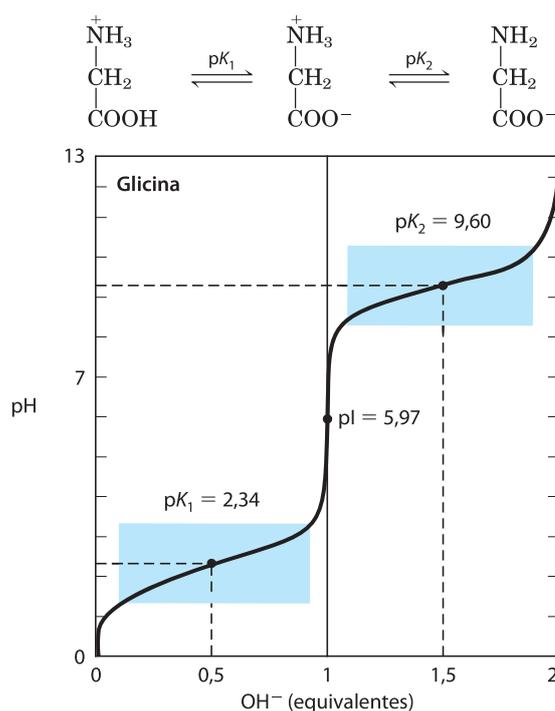


FIGURA 3-10 Titulação de um aminoácido. Aqui é mostrada a curva de titulação de 0,1 M de glicina a 25°C. As espécies iônicas que predominam em pontos-chave na titulação são mostradas acima do gráfico. Os retângulos sombreados, centrados em torno de $\text{pK}_1 = 2,34$ e $\text{pK}_2 = 9,60$, indicam 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 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 pK_a de 4,76 – próximo da média para um grupo carboxila ligado a um hidrocarboneto alifático não substituído. O 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 α , como descrito na Figura 3-11. As cargas opostas no zwitterion resultante estão estabilizadas. De modo semelhante, o pK_a do grupo amino na glicina é alterado para baixo em relação ao 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 α -amino tem um pK_a menor do que o de um de uma amina alifática, como a metilamina (Figura 3-11). Em resumo, o 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 pK_a perturbados de grupos doadores/aceptores de prótons de resíduos específicos.

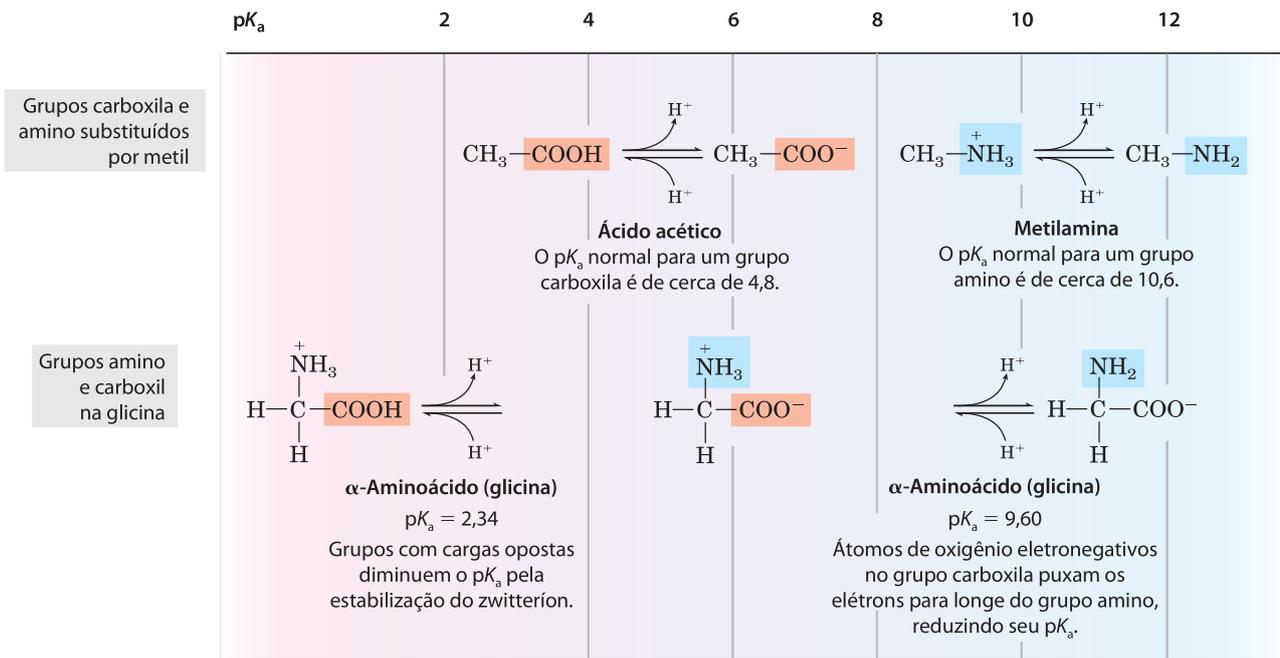


FIGURA 3-11 Efeito do ambiente químico no pK_a . Os valores de pK_a para os grupos ionizáveis na glicina são mais baixos do que aqueles dos grupos simples de carboxila e o amino substituídos por metil. Essas perturbações do pK_a se devem a interações intramoleculares. Efeitos semelhantes podem ser causados 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.

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.

Curvas de titulação predizem a carga elétrica dos aminoácidos

Aminoácidos diferem em suas propriedades acidobásicas

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 :

As propriedades compartilhadas de muitos aminoácidos permitem algumas generalizações simplificadas sobre seu comportamento acidobásico. Em primeiro lugar, todos os aminoácidos com um único grupo α -amino, um único grupo α -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).

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

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-

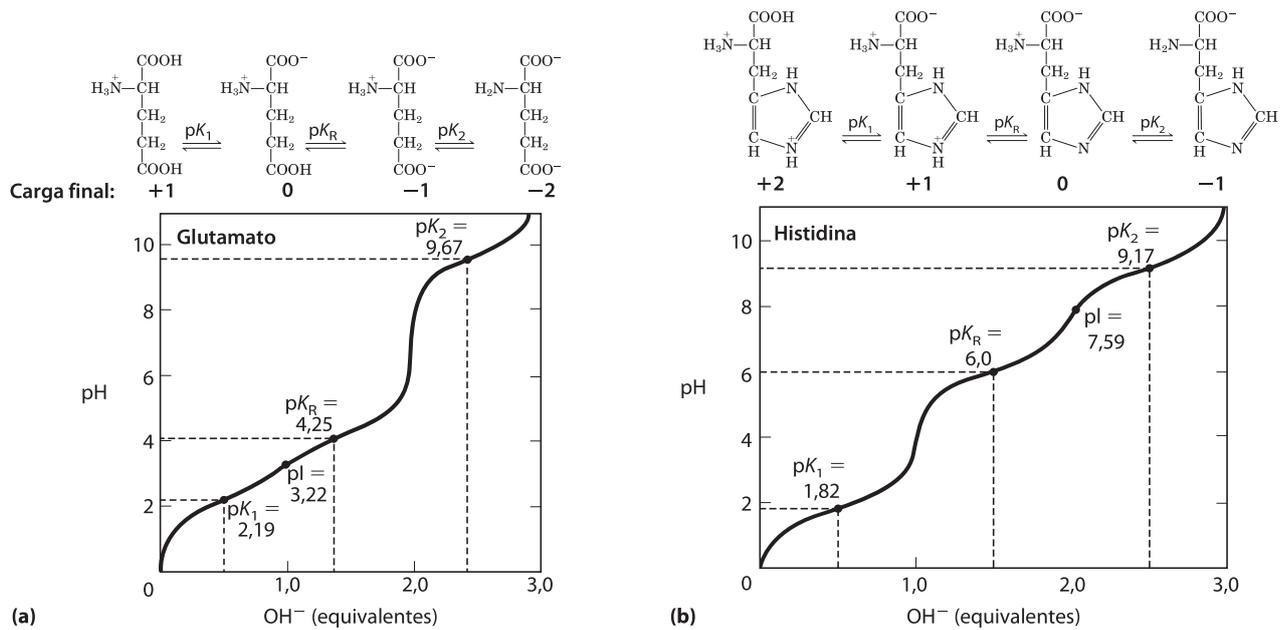


FIGURA 3-12 Curvas de titulação para (a) glutamato e (b) histidina. O grupo R do pK_a é designado aqui como pK_R .

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 α -carboxila, para a titulação do grupo α -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 α -carboxila, um grupo α -amino e um grupo R característico substituído no átomo do carbono α . O átomo de carbono α 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 α -carboxila de um aminoácido e do grupo α -amino do outro (**Figura 3-13**). A formação da ligação peptídica é um exemplo de

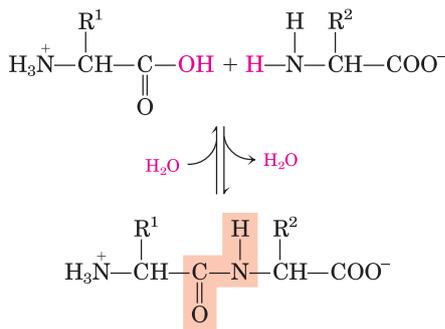


FIGURA 3-13 Formação de uma ligação peptídica por condensação.

O grupo α -amino de um aminoácido (com grupo R^1) atua como nucleófilo para deslocar o grupo hidroxila de outro aminoácido (com grupo R^2), 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 α -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. ■

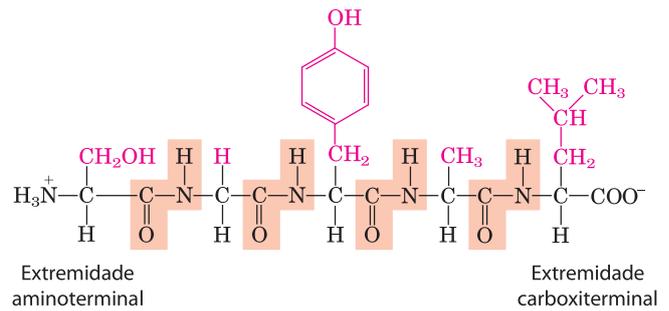


FIGURA 3-14 O pentapeptídeo seril-glicil-tirosil-alanil-leucina, **Ser-Gly-Tyr-Ala-Leu**, ou **SGYAL**. Os peptídeos são nomeados a partir do resíduo aminoterminal, que por convenção é colocado à esquerda. As ligações peptídicas são sombreadas; os grupos R estão em cor-de-rosa.

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 α -amino e um grupo α -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 α . Os grupos α -amino e α -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

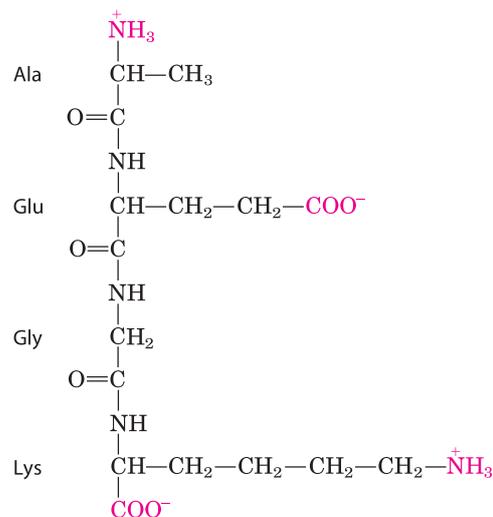


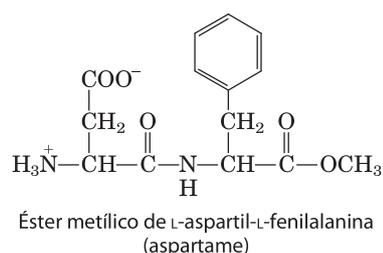
FIGURA 3-15 Alanil-glutamil-glicil-lisina. Este tetrapeptídeo tem um grupo α -amino livre, um grupo α -carboxila livre e dois grupos R ionizáveis. Os grupos ionizados em pH 7,0 estão em cor-de-rosa.

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 α -amino e α -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 α -carboxila e α -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 α idênticas e duas cadeias β idênticas, todas as quatro mantidas unidas por interações não covalentes. Cada subunidade α é pareada de modo idêntico com uma subunidade β 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 bovinos, 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	β_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.

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-

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 ((NH₄)₂SO₄) 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 semi-permeável. 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

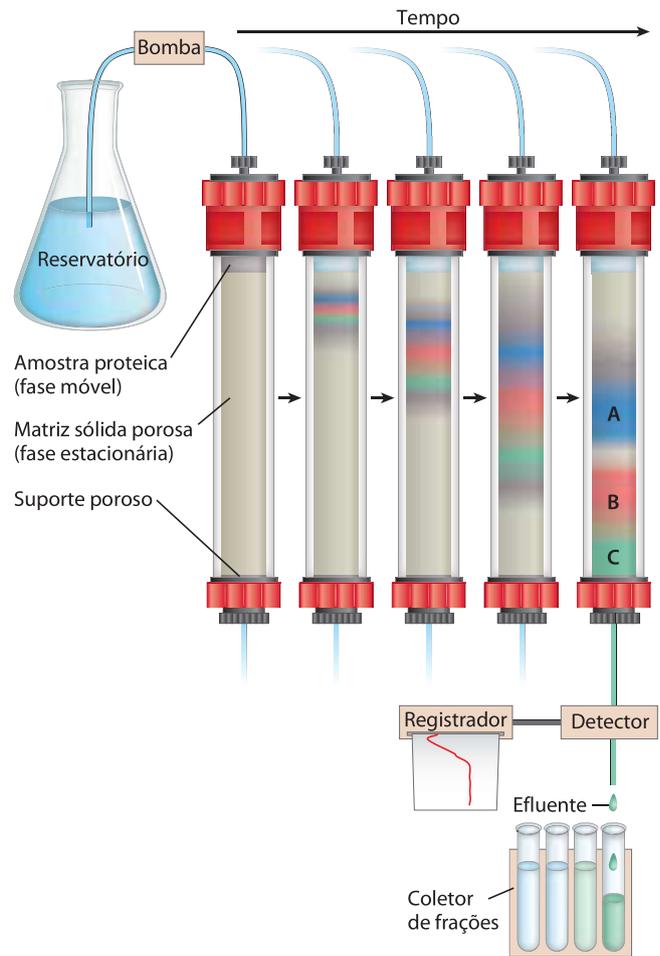
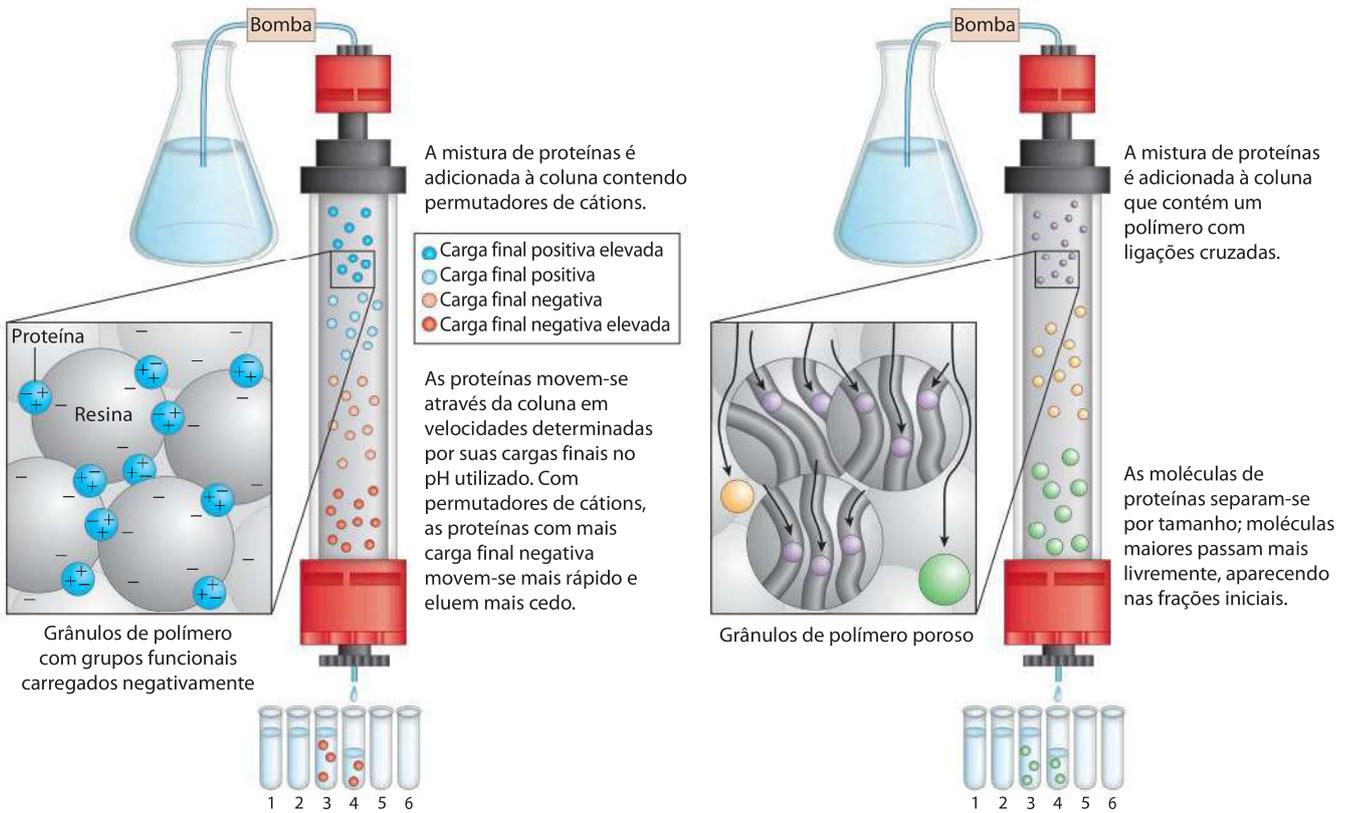


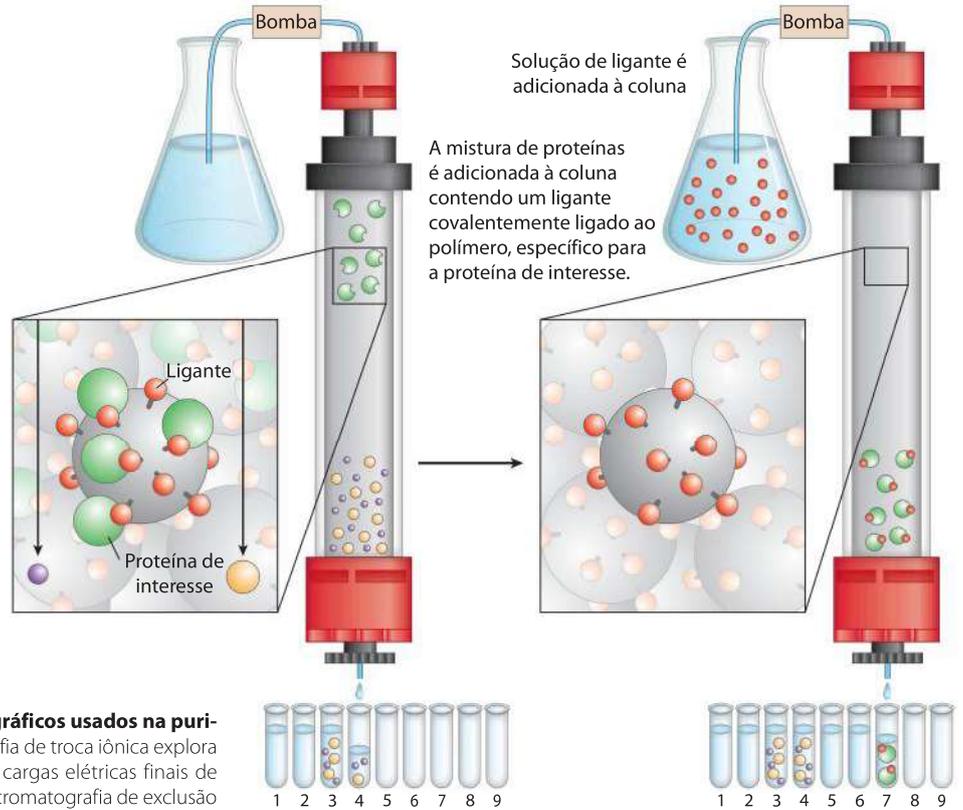
FIGURA 3-16 Cromatografia em coluna. Os elementos padrão de uma coluna cromatográfica incluem um material poroso (matriz) sólido apoiado no interior de uma coluna, geralmente feita de plástico ou vidro. Uma solução, a fase móvel, flui através da matriz, a fase estacionária. A solução que sai da coluna (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



(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 no 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, μ , de uma molécula é a razão de sua velocidade V , em relação ao seu potencial elétrico. A mobilidade eletroforética é também igual à carga

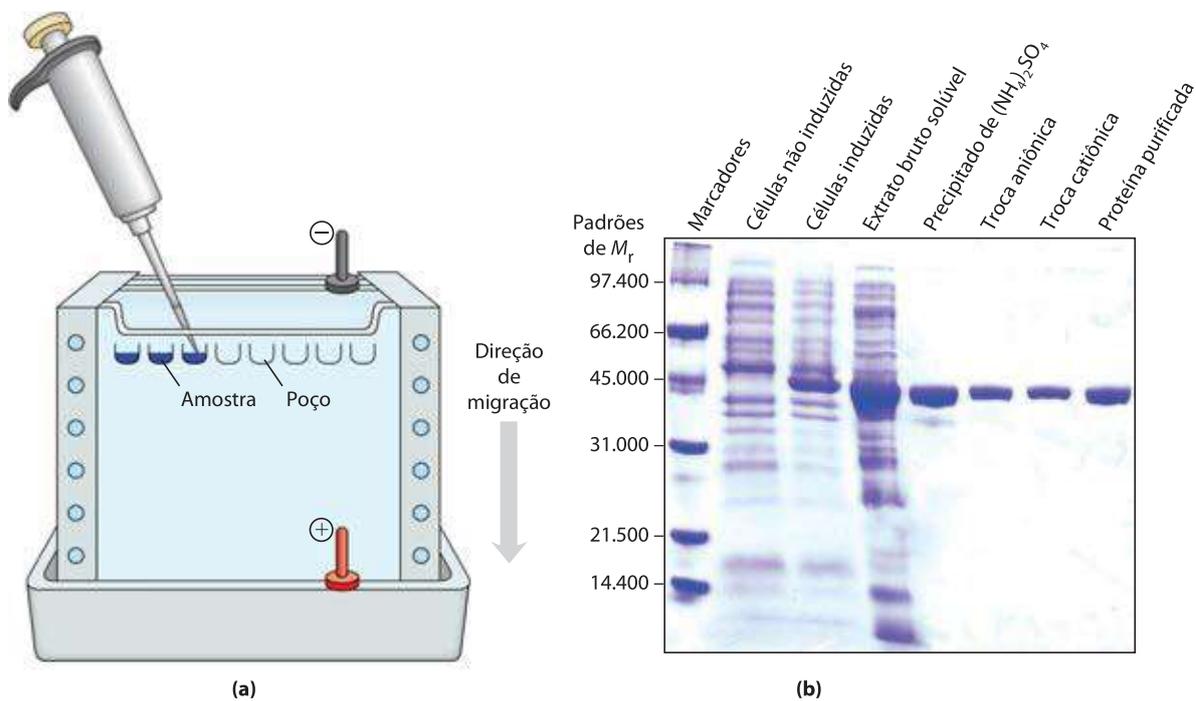


FIGURA 3-18 Eletroforese. (a) Diferentes amostras são colocadas em poços ou depressões no topo do gel de SDS-poli(acrilamida). As proteínas se movem para o gel quando um campo elétrico é aplicado. O gel minimiza as correntes de convecção causadas pelos pequenos gradientes de temperatura, bem como movimentos 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.

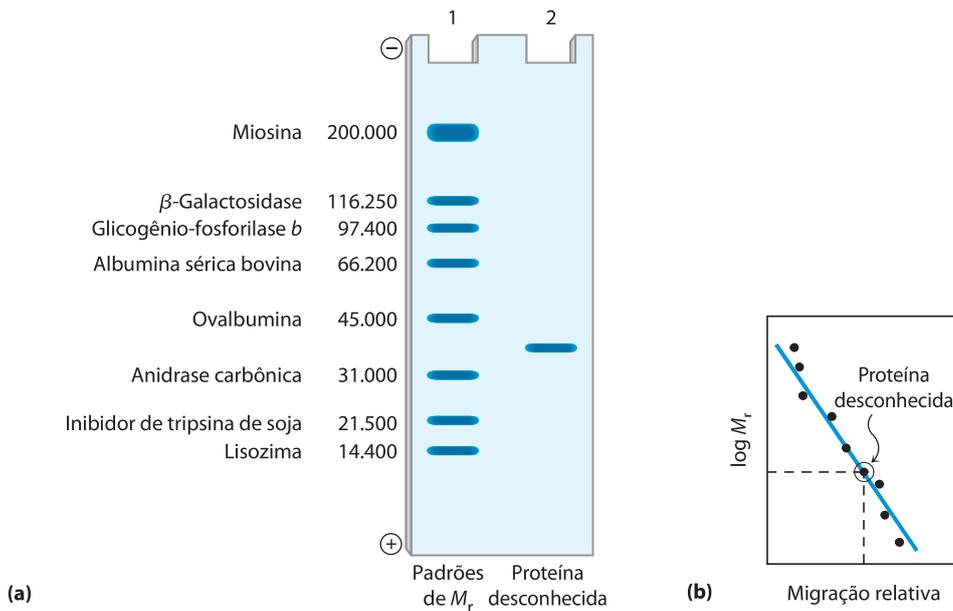


FIGURA 3-19 Estimando a massa molecular de uma proteína. A mobilidade eletroforética de uma proteína em gel de SDS-poliacrilamida está relacionada à sua massa molecular, M_r . **(a)** Proteínas-padrão de massa molecular conhecida são sujeitas à eletroforese (calha 1). Estas proteínas marcadoras podem 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

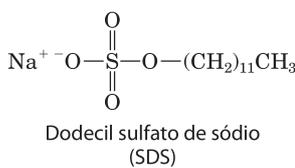
a massa molecular da proteína desconhecida seja lido a partir do gráfico. (De maneira semelhante, um conjunto de proteínas padrão com tempos de retenção reproduzíveis em uma coluna de exclusão por tamanho pode ser usado para criar uma curva padrão de tempo de retenção versus $\log M_r$. O tempo de retenção 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 conhecida 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 isoelétrica** é um procedimento utilizado para determinar o ponto isoelétrico (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 isoelétricos diferentes são, assim, distribuídas de modo diferente ao longo do gel.

A combinação da focalização isoelétrica 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.

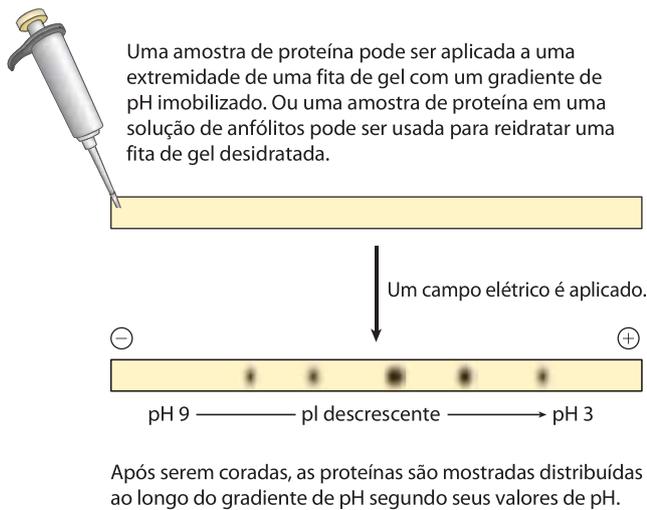


FIGURA 3-20 Focalização isoeétrica. Essa técnica separa proteínas de acordo com seus pontos isoeletricos. Uma mistura de proteínas é colocada em uma fita de gel contendo um gradiente de pH imobilizado. Com a aplicação de um campo elétrico, as proteínas entram no gel e migram até que cada uma atinja um 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-

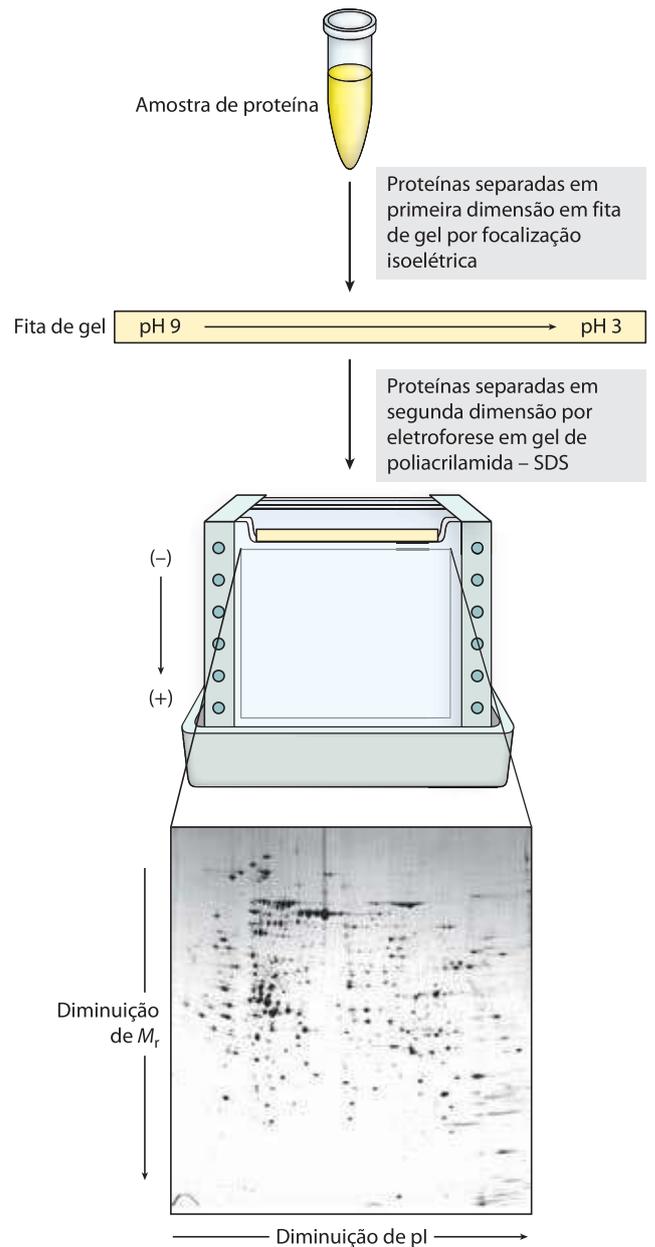


FIGURA 3-21 Eletroforese bidimensional. Em primeiro lugar, as proteínas são separadas por focalização isoeétrica em uma fita de gel fina. O gel é colocado então horizontalmente em um segundo gel em forma de placa, e as proteínas são separadas por eletroforese em gel de poliacrilamida-SDS. A separação 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-

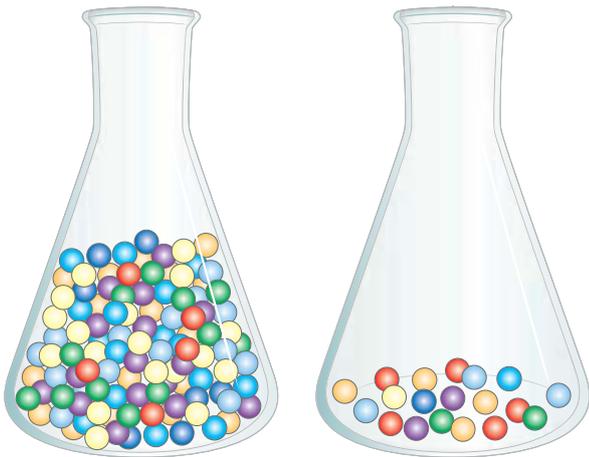


FIGURA 3-22 Atividade versus atividade específica. A diferença entre esses termos pode ser ilustrada considerando dois béqueres contendo esferas. Os béqueres contêm o mesmo número de esferas vermelhas, mas números diferentes de esferas de outras cores. Se as esferas representam proteínas, ambos os béqueres 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.

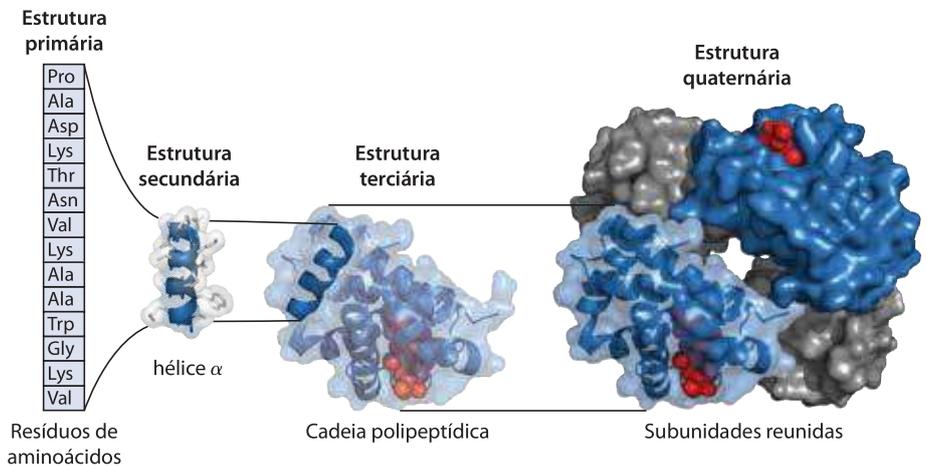
3.4 A estrutura de proteínas: estrutura primária

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

FIGURA 3-23 Níveis de estrutura nas proteínas.

A *estrutura primária* consiste em uma sequência de aminoácidos unidos por ligações peptídicas e inclui quaisquer pontes dissulfeto. O polipeptídeo resultante pode ser disposto em unidades de *estrutura secundária*, como em uma hélice α . A hélice é uma parte da *estrutura terciária* do polipeptídeo dobrado, que é ele mesmo uma das subunidades que compõem a *estrutura quaternária* da proteína multissubunidade, nesse caso a hemoglobina.



(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

Dois 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.

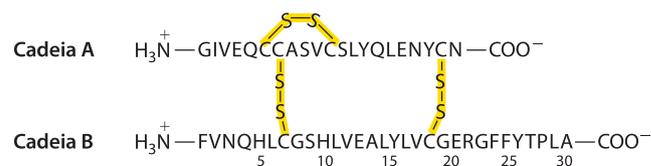
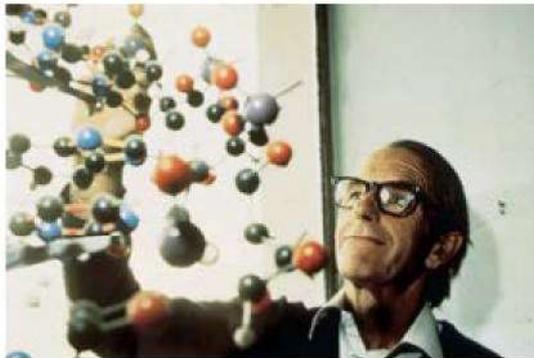


FIGURA 3-24 Sequência de aminoácidos da insulina bovina. As duas cadeias de polipeptídeos estão unidas por ligações cruzadas de dissulfeto (amarelo). A cadeia A da insulina é idêntica em humanos, porcos, cães, coelhos e cachalotes. As cadeias B de vacas, porcos, cães, bodes e cavalos são idênticas.

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 α -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 fenilsotiocianato 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 feniltioindantoína 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-

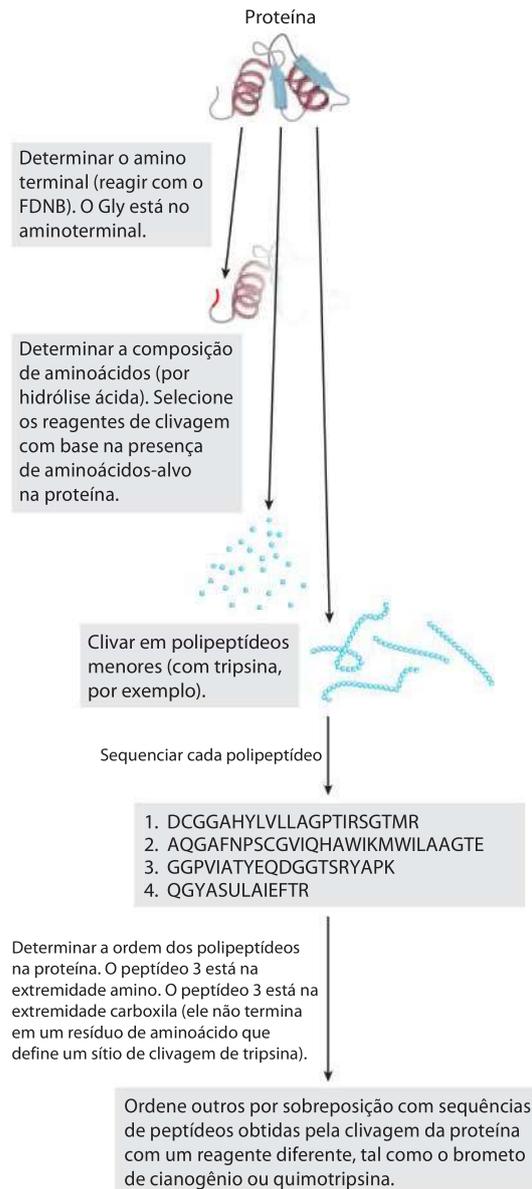


FIGURA 3-25 Sequenciamento direto de proteínas. Os procedimentos aqui apresentados foram aqueles desenvolvidos por Fred Sanger para sequenciar a insulina e têm sido usados posteriormente para várias proteínas adicionais. FDNB é 1-fluoro-2,4-dinitrobenzeno (ver texto e Figura 3-26).

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.

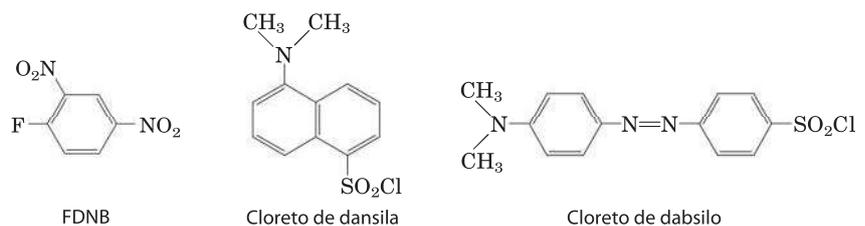


FIGURA 3-26 Reagentes usados para modificar o grupo α -amino do aminoácido terminal.

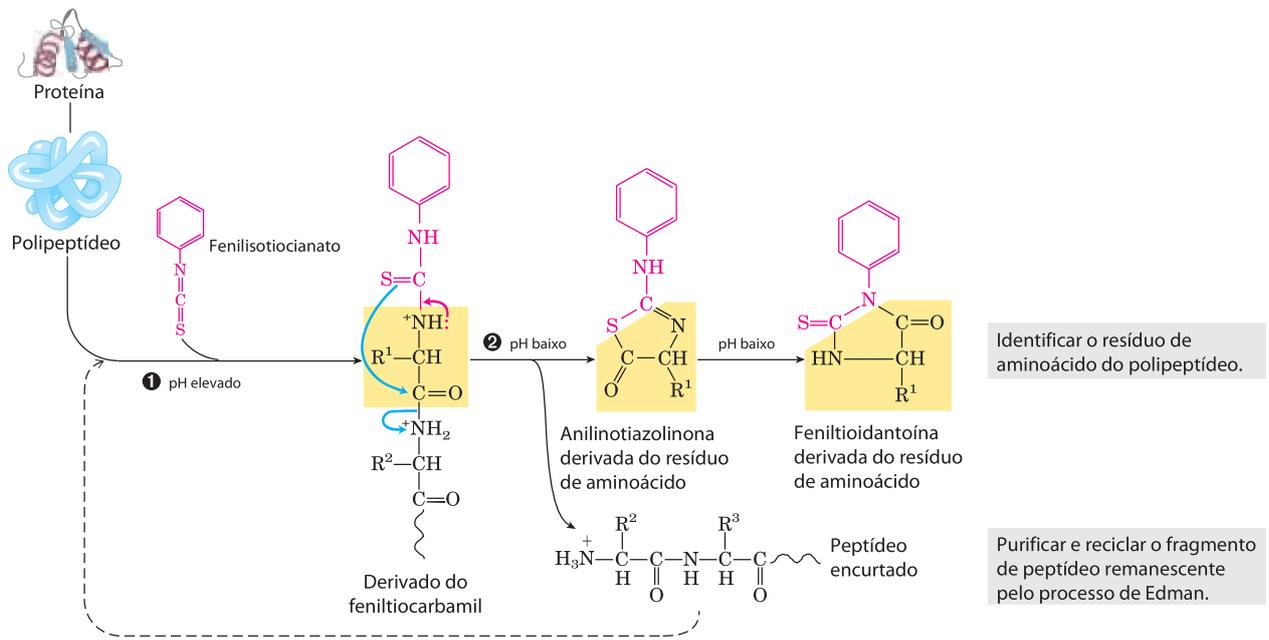


FIGURA 3-27 A química do sequenciamento de proteínas desenvolvida por Pehr Edman. A ligação peptídica mais próxima do aminoterminal da proteína ou polipeptídeo é clivada em duas etapas. As duas etapas são

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-

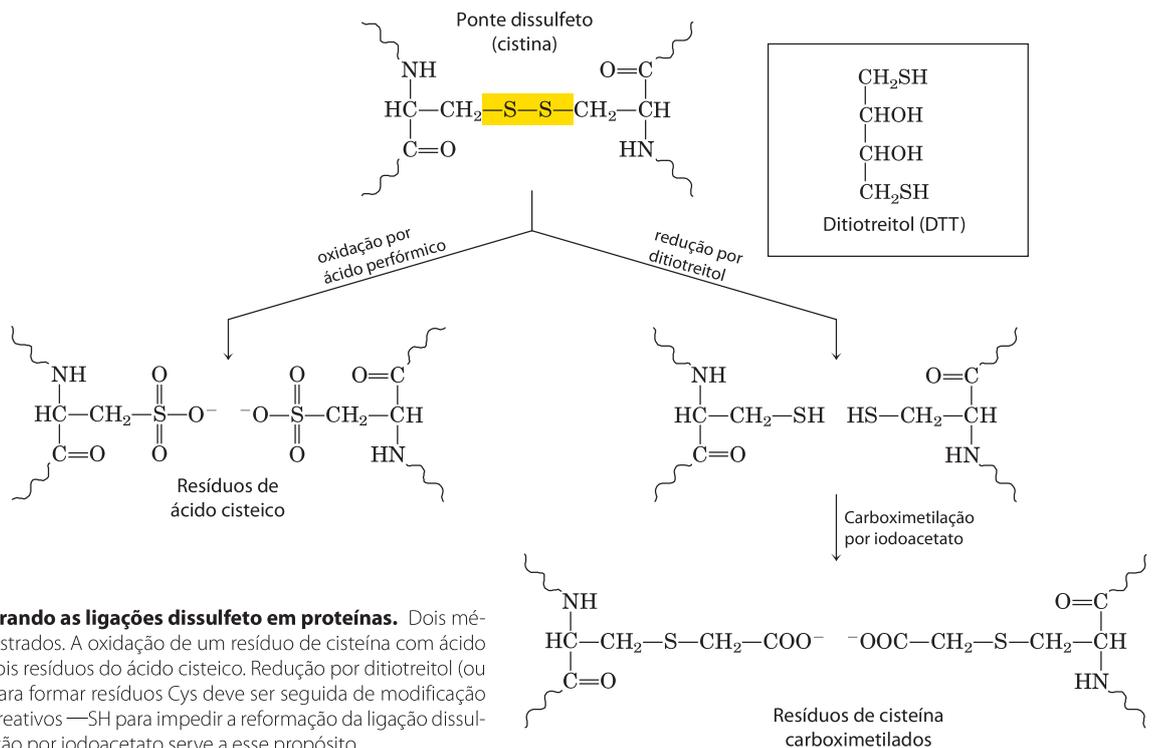


FIGURA 3-28 Quebrando as ligações dissulfeto em proteínas. Dois métodos comuns são ilustrados. A oxidação de um resíduo de cisteína com ácido perfórmico produz dois resíduos do ácido cisteico. Redução por ditioneitol (ou β -mercaptoetanol) para formar resíduos Cys deve ser seguida de modificação adicional 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 α -amino aminoterminal podem ser empregados para marcar as aminas primárias dos resíduos Lys (Figura 3-26). O grupo sulfidríla nos resíduos Cys pode ser

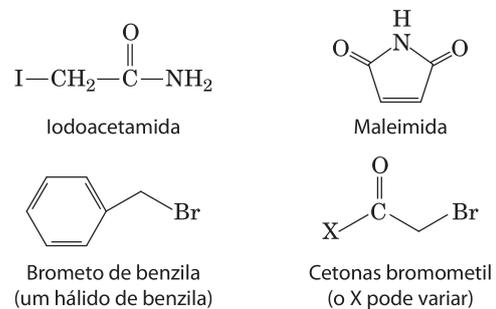


FIGURA 3-29 Reagentes usados para modificar os grupos sulfidríla dos resíduos Cys. (Ver também Figura 3-28.)

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

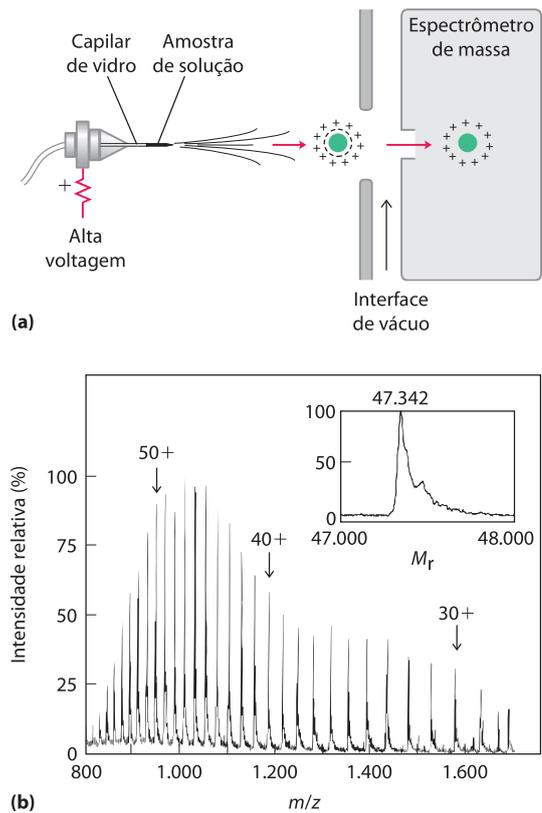


FIGURA 3-30 Espectrometria de massa com ionização por eletroaspersão de uma proteína. (a) Uma solução de proteína é dispersa em gotículas altamente carregadas pela passagem através de uma agulha sob a influência de um campo elétrico de alta voltagem. As gotículas evaporam e os íons (com prótons adicionados, 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 aumentadas 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

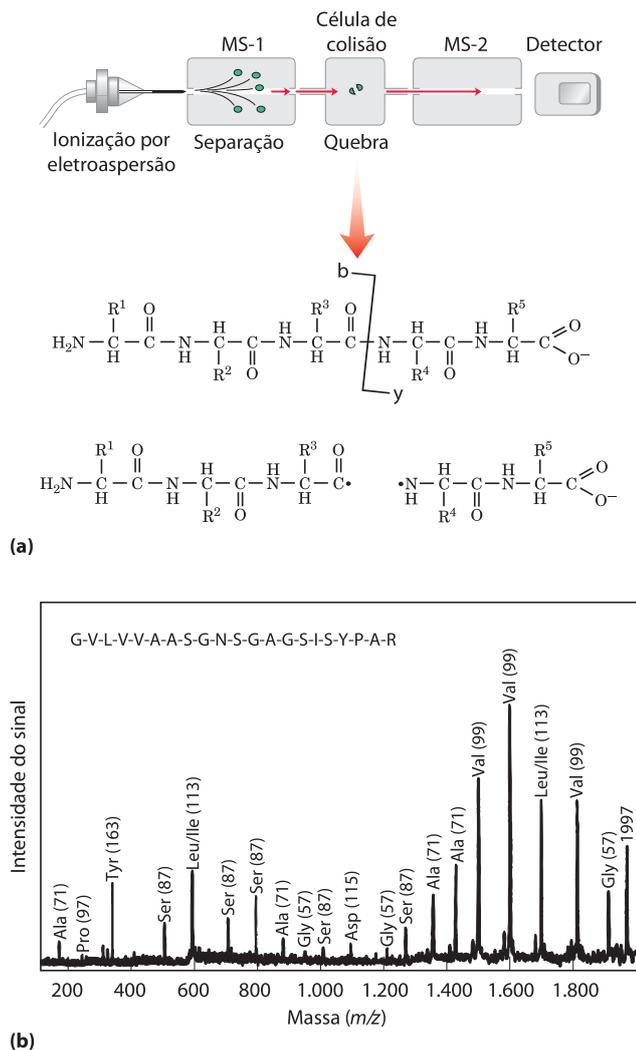


FIGURA 3-31 Obtendo informação da sequência proteica com MS em tandem. (a) Após a hidrólise proteolítica, uma solução proteica é injetada em um espectrômetro de massa (MS-1). Os diferentes peptídeos são dispostos de modo que apenas um tipo é selecionado para análise adicional. O peptídeo selecionado 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 poliacrilamida. 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

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 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.

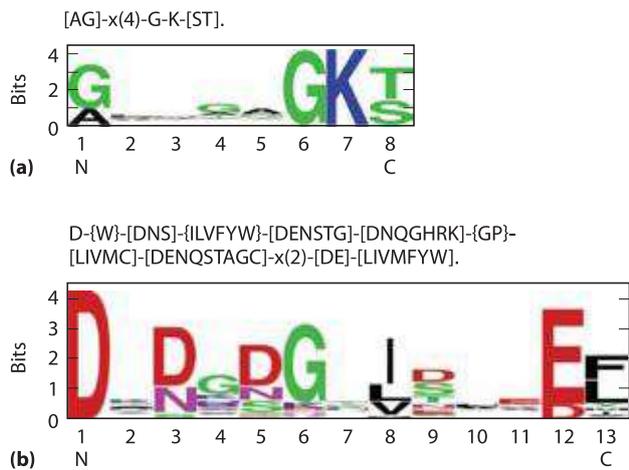


FIGURA Q-1 Representações de duas sequências de consenso. (a) Alça P, estrutura ligadora de ATP; (b) mão EF, estrutura ligadora de Ca^{2+} .

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 predomínio 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 α (fator de alongamento 1 α) 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 α 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 secção 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

```

Escherichia coli  TGNRTIAVYDLGGGTFDISIIIEIDEVDGKTFEVLATNGDTHLGGEDFDSRLIHYL
Bacillus subtilis DEDQTIILLYDLGGGTFDVSILELGDGTFEVRSTAGDNRLGGDDFDQVIIDHL

```

Intervalo

FIGURA 3-33 Alinhando sequências de proteínas com o uso de intervalos. Aqui é mostrada a sequência de alinhamento de uma curta secção 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 α /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 um 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> <i>Saccharomyces cerevisiae</i>	IGHVDHGKSTMVGR LLYETGSVPEHV IEQH
Eucariontes		IGHVDHGKSTLVGR LLMDRGFIDEKT IVKEA
Bactérias gram-positivas	{ <i>Homo sapiens</i> <i>Bacillus subtilis</i> <i>Escherichia coli</i>	IGHVDSGKSTTTGHL LIYKCGGIDKRT IEKF
Bactérias gram-negativas		IGHVDSGKSTTTGHL LIYKCGGIDKRT IEKF
		IGHVDHGKSTMVGR I T T V
		IGHVDHGKSTLTAA I T T V

FIGURA 3-34 Uma sequência-assinatura na família de proteínas EF-1 α /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.

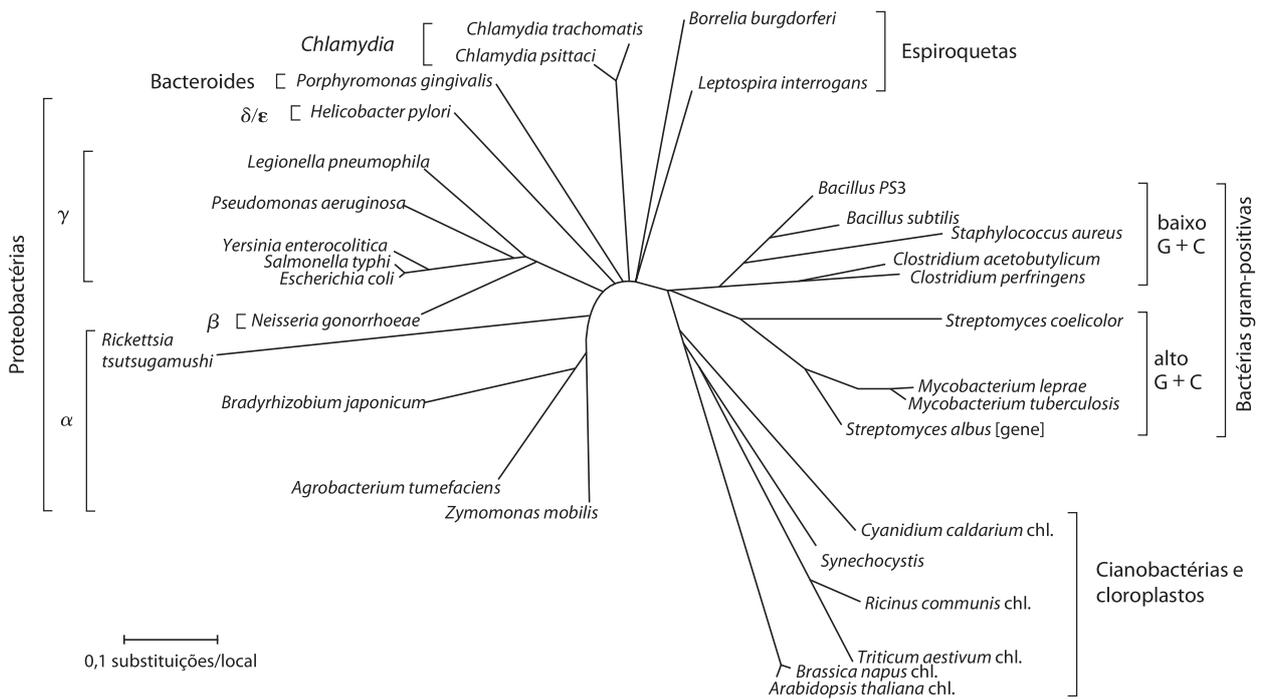


FIGURA 3-35 **Árvore evolutiva derivada de comparações entre as sequências de aminoácidos.** Árvore evolutiva bacteriana, com base na divergência de sequências observada na família de proteínas GroEL. Tam-

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-

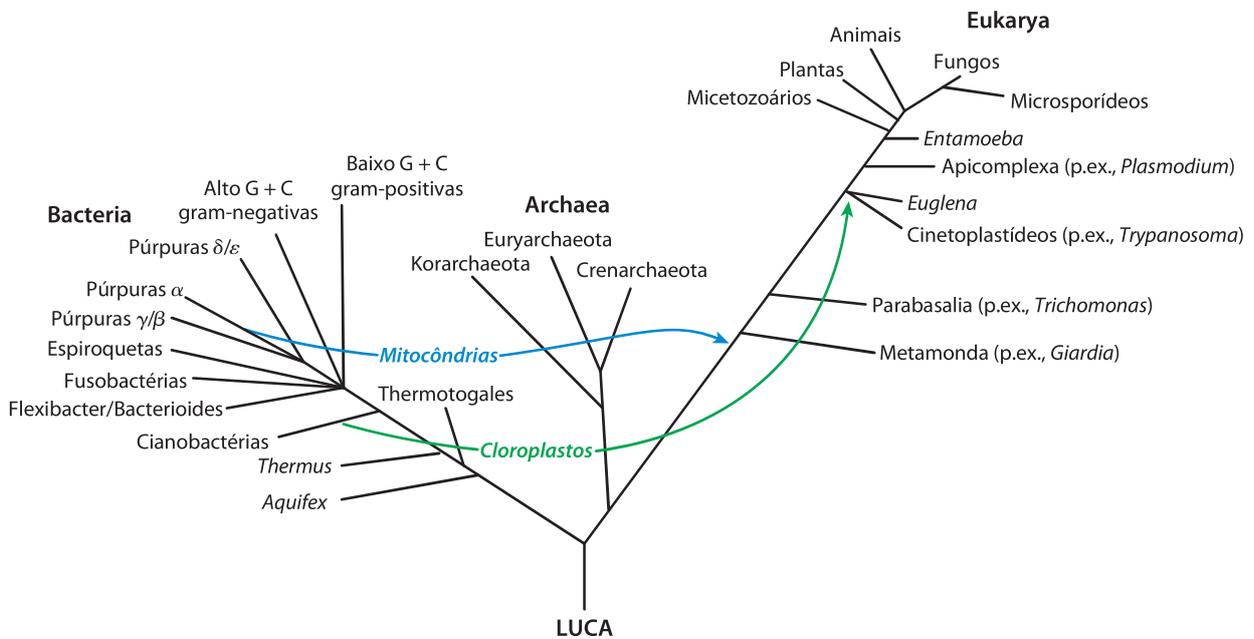


FIGURA 3-36 **Árvore de consenso da vida.** A árvore mostrada aqui baseia-se em análises de muitas sequências de proteínas e características genômicas adicionais. A árvore apresenta apenas uma fração da informação disponível, bem como apenas uma fração dos temas que ainda precisam ser solucionados. Cada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

Termos-chave

Termos em *negrito* são definidos no glossário.

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

proteases 99	horizontal 106
MALDI MS 101	proteínas homólogas 106
ESI MS 101	homólogos 106
sequência consenso 104	parálogos 106
bioinformática 104	ortólogos 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 e 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.

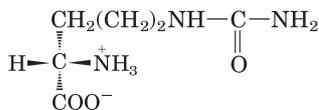
Steen, H. & Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5**, 699–711.

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 pK_a do grupo carboxila.

(e) O pH é igual ao 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 $^+\text{H}_3\text{N}-\text{CH}_2-\text{COO}^-$.

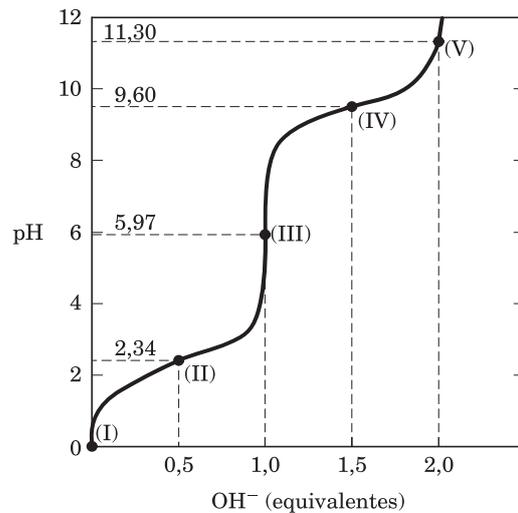
(k) A carga final *média* da glicina é -1 .

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

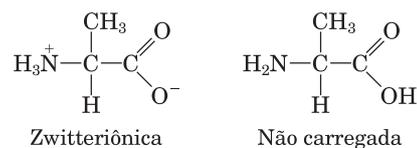
(m) Este é o ponto isoeletrico.

(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 isoeletrico 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 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 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

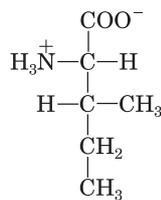
ionização pode ser aproximado tratando-se cada grupo ionizável independentemente.

(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 ditiotreitol, 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⁺, Cl⁻ 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.

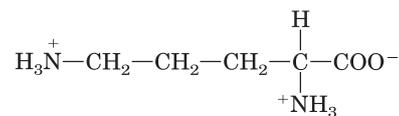
(a) A hidrólise completa por 6 M de HCl a 110°C, seguida pela análise de aminoácidos, indicou a presença de Gly, Leu, Phe e Tyr em uma razão molar de 2:1:1:1.

(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:

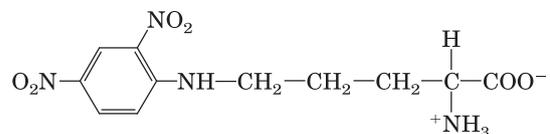
(a) A hidrólise ácida completa do peptídeo seguida de análise de aminoácidos produziu quantidades equimolares de Leu, Orn, Phe, Pro e Val. Orn é ornitina, um aminoácido que não está presente em proteínas, mas aparece em alguns peptídeos. Ela tem a seguinte estrutura:



(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 α -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):

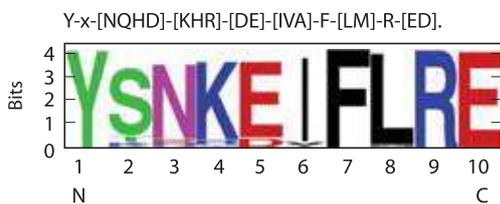
Leu-Phe Phe-Pro Orn-Leu Val-Orn

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. ATKNRASCLVPKHLGALMFWRHKQLVSDPILQKRQHILVCRNAAG
2. GPYFGDEPLDVHDEPEEG
3. PHLLSAWKGMGVGKSQSFAALIVILA

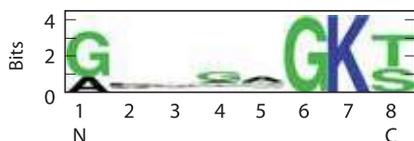
Qual deles migraria mais lentamente durante a cromatografia através de:

(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?

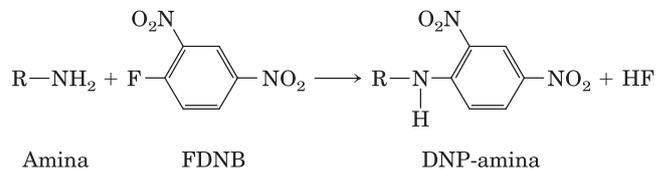


Problemas de análise de dados

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 per fórmico 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 α -DNP-fenilalanina

B2: α -DNP-fenilalanina; valina

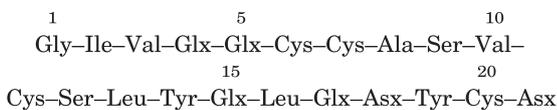
B3: ácido aspártico; α -DNP-fenilalanina; valina

B4: ácido aspártico; ácido glutâmico; α -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 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.