

11

INTRODUÇÃO À BIOLOGIA ESTRUTURAL DE PROTEÍNAS

Autores 11.1

Rafael Pereira Lemos , Paulo Henrique dos Santos , Aline Rocha 

Revisão: Bibiana Fam  Filipe Teixeira  Carlos Capelini 

Cite este artigo 11.1

Lemos, R; Santos, PH; Rocha, A. **Introdução à Biologia Estrutural de Proteínas**. BIOINFO.
ISSN: 2764-8273. Vol. 3. p.11 (2023). doi: 10.51780/bioinfo-03-11

Resumo 11.1

As **proteínas** são as macromoléculas mais abundantes nos sistemas biológicos, estando presentes em todas as células e tecidos. Apresentam uma diversidade de funções biológicas em nosso organismo, como: estruturação de células e tecidos, transporte e armazenamento de outras moléculas, receptores, hormônios, anticorpos, fatores de transcrição e enzimas. Esta gama de funções é possível devido à diversidade estrutural das proteínas. As proteínas são polímeros cujos monômeros compreendem resíduos dos **20 aminoácidos** que estão naturalmente presentes nestas moléculas. Estes aminoácidos são combinados em sequências diferentes para dar origem à diversidade de proteínas existentes nos diferentes organismos. As proteínas são produzidas no processo de tradução do RNA mensageiro (RNAm), a partir da união de unidades de aminoácidos carregados por RNAs transportadores (RNAt). Este carregamento é feito com base na sequência do molde de RNAm que por sua vez é produzido a partir do processo de transcrição da sequência codificante de DNA. Desta forma, as proteínas correspondem às moléculas pelas quais a informação genética é expressa [1].

11.1 Aminoácidos

Cada **aminoácido** apresenta um grupo amino terminal básico (-NH₂), e um grupo carboxílico terminal ácido (-COOH), além de uma cadeia lateral variável (-R) que determina o tipo de aminoácido, e um hidrogênio que completa as quatro ligações ao carbono central, também chamado carbono alfa (C- α) (Figura 11.1). Para a maioria dos aminoácidos, com exceção da glicina, o carbono alfa é quiral. Dessa forma, esses aminoácidos podem existir na forma de estereoisômeros (D- ou L-), sendo que nas proteínas são encontrados quase exclusivamente os estereoisômeros L- [1,2].

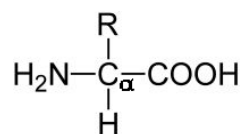


Figura 11.1: Estrutura geral de um aminoácido. Fonte: próprio autor.

Os aminoácidos podem ser representados em códigos de uma ou três letras (Tabela 11.1) [3].

Tabela 11.1: Classificação, códigos e propriedades dos 20 aminoácidos que constituem as proteínas. Fonte: adaptado de [1,3].

Grupo	Aminoácido	Código (3 letras)	Código (1 letra)	Massa Molecular (Da)	pKa (COOH)	pKa (NH3+)	pKa (R)	pI
Apolares	Glicina	Gly	G	75	2,34	9,60	-	5,97
	Alanina	Ala	A	89	2,34	9,69	-	6,01
	Prolina	Pro	P	115	1,99	10,96	-	6,48
	Valina	Val	V	117	2,32	9,62	-	5,97
	Leucina	Leu	L	131	2,36	9,60	-	5,98
	Isoleucina	Ile	I	131	2,36	9,68	-	6,02
	Metionina	Met	M	149	2,28	9,21	-	5,74
Aromáticos	Fenilalanina	Phe	F	165	1,83	9,13	-	5,48
	Tirosina	Tyr	Y	181	2,20	9,11	10,07	5,66
	Triptofano	Trp	W	204	2,38	9,39	-	5,89
Polares não carregados	Serina	Ser	S	105	2,21	9,15	-	5,68
	Treonina	Thr	T	119	2,11	9,62	-	5,87
	Cisteína	Cys	C	121	1,96	10,28	8,18	5,07
	Asparagina	Asn	N	132	2,02	8,80	-	5,41
	Glutamina	Gln	Q	146	2,17	9,13	-	5,65
Básicos	Lisina	Lys	K	146	2,18	8,95	10,53	9,74
	Histidina	His	H	155	1,82	9,17	6,00	7,59
	Arginina	Arg	R	174	2,17	9,04	12,48	10,76
Ácidos	Aspartato	Asp	D	133	1,88	9,60	3,65	2,77
	Glutamato	Glu	E	147	2,19	9,67	4,25	3,22

Eles podem ser classificados em essenciais, ou seja, aqueles que não são produzidos pelo nosso corpo e devem ser obtidos da dieta (histidina, isoleucina, leucina, lisina, metionina, fenilalanina, treonina e triptofano), ou não essenciais, que podem ser sintetizados pelo nosso organismo [4]. Porém, mais importante para a biologia estrutural é a classificação dos aminoácidos de acordo com as propriedades da cadeia lateral, sendo divididos em: alifáticos ou apolares, aromáticos, polares não carregados, carregados positivamente ou básicos e carregados negativamente ou ácidos (Figura 11.2).

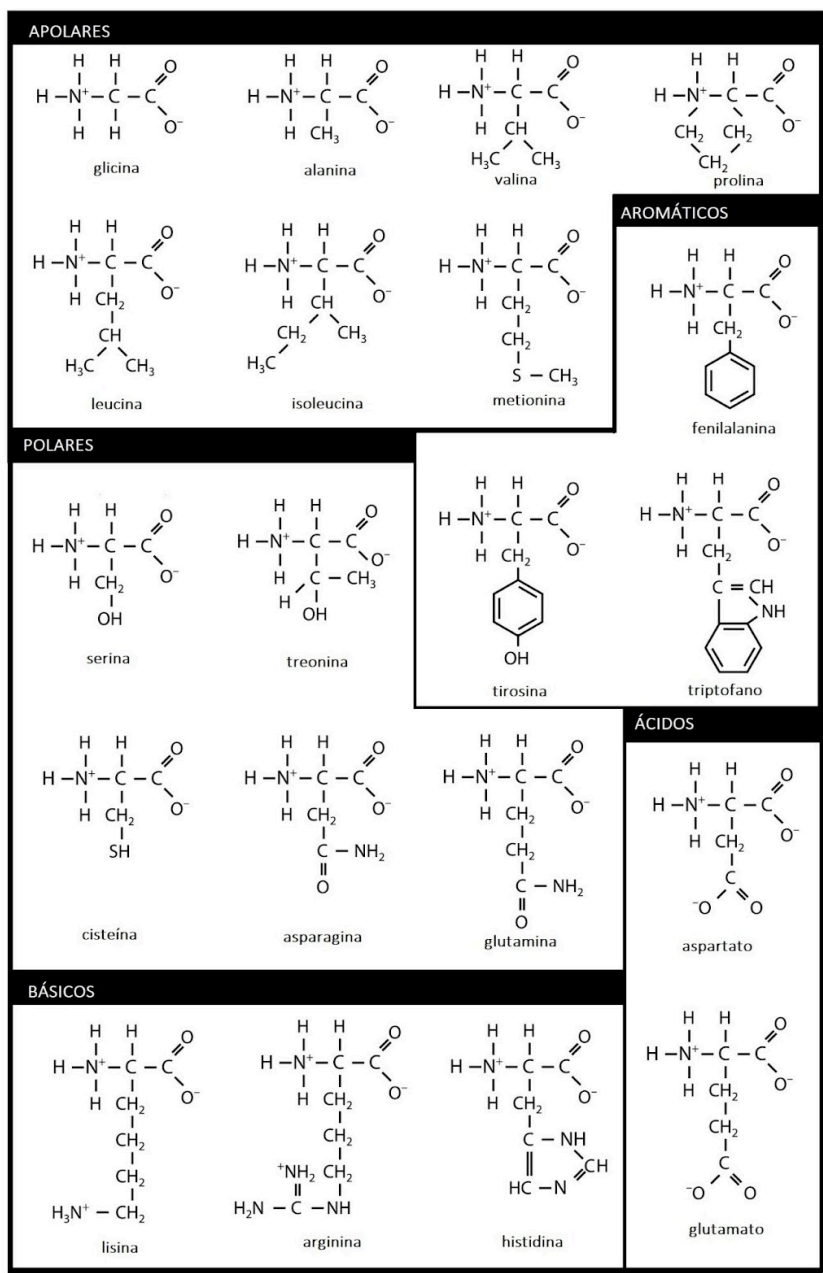


Figura 11.2: Estrutura dos aminoácidos. Fonte: próprio autor.

11.2 Estrutura proteica

Os peptídeos e proteínas são formados por meio de ligações peptídicas entre o grupo carboxílico de um aminoácido e o grupo amino do aminoácido seguinte,

com liberação de uma molécula de água. A ligação peptídica (OC–NH) (Figura 11.3) tem caráter intermediário entre uma ligação simples e uma dupla, sendo uma ligação planar e rígida. Já as ligações entre C_{α} e NH ou entre C_{α} e CO são ligações simples e podem fazer rotações. Os ângulos de rotação dessas ligações são denominados phi (ϕ) para a ligação C_{α} , e psi (ψ) para a ligação C_{α} -CO [5]. Nas proteínas, os valores de ϕ e ψ são limitados pela repulsão estérica entre as cadeias laterais (R) dos resíduos de aminoácidos.

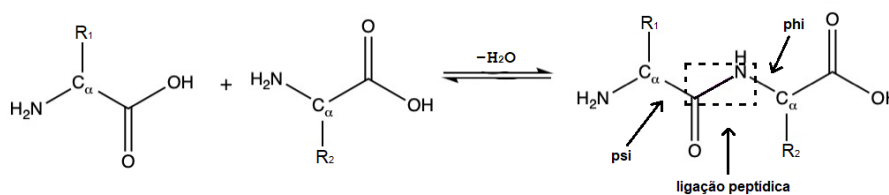


Figura 11.3: A ligação peptídica e os ângulos diedros phi (ϕ) e psi (ψ). Fonte: Próprio Autor.

A estrutura proteica é de fundamental importância para a sua função, e é classificada de forma hierárquica em quatro níveis: estrutura primária, secundária, terciária e quaternária [6]. A **estrutura primária** corresponde à **sequência** dos resíduos de aminoácidos da proteína, unidos pela ligação peptídica, além das pontes de dissulfeto que podem ser formadas pelas cisteínas. A **estrutura secundária** corresponde ao arranjo espacial dos átomos adjacentes na **cadeia principal** da proteína, e é definida pela ligação peptídica e pelos ângulos diedros ϕ e ψ . Certos padrões regulares de ângulos são observados nas estruturas de proteínas, gerando estruturas secundárias comuns, como a **alfa hélice**, as **folhas beta** e as **voltas beta**. Já a **estrutura terciária** corresponde ao arranjo tridimensional dos átomos da proteína devido ao seu enovelamento ou dobramento, incluindo também as interações das cadeias laterais, mesmo entre resíduos de aminoácidos que estão distantes na sequência proteica, mas próximos no espaço. Quando a proteína apresenta duas ou mais cadeias enoveladas distintas, que interagem entre si formando complexos, a **estrutura** é denominada **quaternária**.

Conforme a sua estrutura, as proteínas podem ser classificadas ainda em [7]: **proteínas globulares**, quando possuem cadeias bem enoveladas, contendo diversos tipos de estruturas secundárias, sendo assim adaptadas a diversas

funções, incluindo proteínas solúveis; **proteínas fibrosas**, quando possuem cadeias arranjadas em filamentos ou folhas, contendo geralmente um único tipo de estrutura secundária, estando adaptadas a funções de estruturação de células e tecidos; **proteínas intrinsecamente desordenadas**, que não possuem estrutura tridimensional definida; e as **proteínas de membrana**, que contém estruturas específicas com aminoácidos hidrofóbicos expostos para interação com as membranas celulares.

As proteínas enoveladas geralmente apresentam padrões de enovelamento identificáveis [6]. Quando duas ou mais estruturas secundárias são conectadas na forma de um padrão estrutural identificável, forma-se uma unidade estrutural denominada **motivo** (ex.: alça β - α - β e barril β). Já um **domínio** é uma região da proteína independentemente estável, ou seja, que conserva sua estrutura terciária, formada pela combinação de estruturas secundárias ou motivos, e geralmente tem uma função específica. Proteínas que estão relacionadas evolutivamente e pertencem a uma mesma **família** apresentam domínios semelhantes.

Além da cadeia polipeptídica, outros elementos podem estar presentes na estrutura proteica, como **modificações pós-traducionais**, que são adições de grupos químicos ou moléculas em resíduos de aminoácidos específicos após a tradução proteica (ex.: sítios de glicosilação, fosforilação, etc.) [8]. Adicionalmente, a ação de **cofatores** como **grupos prostéticos e coenzimas** [9], que são componentes não proteicos ligados à cadeia polipeptídica, podem ser necessários para a função de proteínas (ex.: heme, íons, NADH). As proteínas exercem sua função via alterações em sua dinâmica conformacional, provocada pela interação com outras moléculas, sejam elas receptores, ligantes, substratos, etc. A região de interação da proteína com seu ligante é denominada **sítio de ligação**, que em enzimas também é denominado **sítio ativo ou catalítico**. Algumas proteínas podem ter sua função modulada por moléculas que se ligam em regiões diferentes do sítio ativo, denominadas **sítios alostéricos**.

11.3 Bases de dados usadas em biologia estrutural

Por fim, diversas bases de dados e ferramentas de bioinformática são frutos da contribuição de estudos na área de biologia estrutural e têm favorecido o avanço de estudos no entendimento da estrutura e função de proteínas. A identificação da sequência e identidade de proteínas, além da obtenção de informações relevantes a respeito de sua estrutura e função, podem ser obtidas com ferramentas como o BLAST [10] e bases de dados como o Uniprot [11] e o PDB [12]. O portal ExPasy [13] reúne um conjunto de recursos para a predição de características físico-químicas e análises de sequências. A predição de estruturas secundárias também pode ser feita com ferramentas como o PSIPRED [14] e o DeepTMHMM [15]. Nos últimos anos, a predição de estruturas terciárias e quaternárias através da inteligência artificial (AlphaFold) [16] tem sido amplamente aplicada, e a visualização destas estruturas pode ser realizada por meio de programas como o PyMOL (Schrödinger, LLC) [17] e ChimeraX [18].

O objetivo deste artigo foi apresentar uma breve introdução à biologia estrutural de proteínas e outras moléculas. Nos próximos artigos, apresentaremos um roteiro para a utilização dessas bases e ferramentas bioinformáticas para a extração de informações a partir de sequências ou estruturas proteicas. Recomenda-se o senso crítico no uso das ferramentas, e sua combinação com a validação experimental sempre que possível.

Nota de transparência 11.1

Este material foi originalmente produzido para um minicurso ministrado durante o Curso de Inverno em Bioinformática da UFMG, realizado em 4 de Julho de 2023, na Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

Saiba mais 11.1

Este artigo está disponível em <https://bioinfo.com.br/introducao-a-biologia-estrutural-de-proteinas/>

11.4 Referências

- [1] Nelson, D. L.; Cox, M. M. *Princípios de bioquímica de Lehninger*. 6. Ed. Porto Alegre: Artmed, 2014.
- [2] Fujii, N.; Takata, T.; Fujii, N.; Aki, K.; Sakaue, H. D-Amino Acids in Protein: The Mirror of Life as a Molecular Index of Aging. *Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics*, v. 1866, n. 7, p. 840–847, jul. 2018.
- [3] IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature and Symbolism for Amino Acids and Peptides. Recommendations 1983. *European Journal of Biochemistry*, v. 138, n. 1, p. 9–37, jan. 1984.
- [4] Wu, G. Amino Acids: Metabolism, Functions, and Nutrition. *Amino Acids*, v. 37, n. 1, p. 1–17, maio 2009.
- [5] Ramachandran, G. N.; Venkatachalam, C. M.; Krimm, S. Stereochemical Criteria for Polypeptide and Protein Chain Conformations. *Biophysical Journal*, v. 6, n. 6, p. 849–872, nov. 1966.
- [6] Sun, P. D.; Foster, C. E.; Boyington, J. C. Overview of Protein Structural and Functional Folds. *Current Protocols in Protein Science*, v. 35, n. 1, fev. 2004.
- [7] Andreeva, A.; Kulesha, E.; Gough, J.; Murzin, A. G. The SCOP Database in 2020: Expanded Classification of Representative Family and Superfamily Domains of Known Protein Structures. *Nucleic Acids Research*, v. 48, n. D1, p. D376–D382, 8 jan. 2020.
- [8] Ramazi, S.; Zahiri, J. Post-Translational Modifications in Proteins: Resources, Tools and Prediction Methods. *Database*, v. 2021, p. baab012, 7 abr. 2021.
- [9] De Bolster, M. W. G. Glossary of Terms Used in Bioinorganic Chemistry (IUPAC Recommendations 1997). *Pure and Applied Chemistry*, v. 69, n. 6, p. 1251–1304, 1 jan. 1997.
- [10] Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *Journal of Molecular Biology*, v. 215, n. 3, p. 403–410, out. 1990.
- [11] The Uniprot Consortium; Bateman, A.; Martin, M.-J.; Orchard, S.; Magrane, M.; Ahmad, S. *et al.* UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, v. 51, n. D1, p. D523–D531, 6 jan. 2023.
- [12] Burley, S. K.; Berman, H. M.; Bhikadiya, C.; Bi, C.; Chen, L.; Di Costanzo, L. *et al.* RCSB Protein Data Bank: Biological Macromolecular Structures Enabling Research and Education in Fundamental Biology, Biomedicine, Biotechnology and Energy. *Nucleic Acids Research*, v. 47, n. D1, p. D464–D474, 8 jan. 2019.

- [13] Duvaud, S.; Gabella, C.; Lisacek, F.; Stockinger, H.; Ioannidis, V.; Durinx, C. Expasy, the Swiss Bioinformatics Resource Portal, as Designed by Its Users. *Nucleic Acids Research*, v. 49, n. W1, p. W216–W227, 2 jul. 2021.
- [14] Buchan, D. W. A.; Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Research*, v. 47, n. W1, p. W402–W407, 2 jul. 2019.
- [15] Hallgren, J.; Tsirigos, K. D.; Pedersen, M. D.; Armenteros, J. J. A.; Marcatili, P.; Nielsen, H.; Krogh, A.; Winther, O. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *Bioinformatics*, 10 abr. 2022.
- [16] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O. *et al.* Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, v. 596, n. 7873, p. 583–589, ago. 2021.
- [17] Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System.
- [18] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers. *Protein Science*, v. 30, n. 1, p. 70–82, jan. 2021.